The effects of cross-generational and cross-dialectal variation on vowel identification and classification

Ewa Jacewicz[a] and Robert Allen Fox
Department of Speech and Hearing Science, The Ohio State University, 1070 Carmack Road, Columbus, Ohio 43210

(Received 25 April 2011; revised 16 December 2011; accepted 20 December 2011)

Cross-generational and cross-dialectal variation in vowels among speakers of American English was examined in terms of vowel identification by listeners and vowel classification using pattern recognition. Listeners from Western North Carolina and Southeastern Wisconsin identified 12 vowel categories produced by 120 speakers stratified by age (old adults, young adults, and children), gender, and dialect. The vowels /ɝ, o, ʊ, u/ were well identified by both groups of listeners. The majority of confusions were for the front /i, ɪ, e, ɛ, æ/, the low back /ɑ, ɔ/ and the monophthongal North Carolina /aɪ/. For selected vowels, generational differences in acoustic vowel characteristics were perceptually salient, suggesting listeners' responsiveness to sound change. Female exemplars and native-dialect variants produced higher identification rates. Linear discriminant analyses which examined dialect and generational classification accuracy showed that sampling the formant pattern at vowel midpoint only is insufficient to separate the vowels. Two sample points near onset and offset provided enough information for successful classification. The models trained on one dialect classified the vowels from the other dialect with much lower accuracy. The results strongly support the importance of dynamic information in accurate classification of cross-generational and cross-dialectal variations. © *2012 Acoustical Society of America*. [DOI: 10.1121/1.3676603]

PACS number(s): 43.71.Es, 43.71.Bp, 43.70.Mn, 43.70.Fq [MAH] Pages: 1413–1433

I. INTRODUCTION

There is considerable variability among speakers in the acoustic structure of their vowels. Some of it is related to differences in vocal tract length, but much of it is influenced by external factors underlying the phonetics of language usage. For example, the internal structure of a given vowel category may be differentially affected by the regional dialect of the speakers or their socioeconomic status. The role of such external factors has been increasingly recognized in the experimental approach to study the acoustics of spoken language, as evident in the rapid growth of the emerging field of sociophonetics (e.g., Foulkes and Docherty, 2006; Labov, 2010; Thomas, 2011). The present study addresses the question of how sociophonetic variation affects the recognition of vowel identity, given the fact that vowels of North American English are split into regional subsystems (e.g., Clopper *et al.*, 2005; Labov *et al.*, 2006) which also undergo dialect-specific internal sound changes (e.g., Labov, 1994; Labov *et al.*, 2006; Jacewicz *et al.*, 2011a; 2011c). The emphasis of the study is on dynamic rather than static vowel characteristics.

The importance of dynamic cues in the recognition of vowel identity has been established on the basis of combined evidence from acoustic measurements, statistical pattern recognition and listeners' identification responses. The time-varying dynamic vowel structure, also known as vowel inherent spectral change (VISC) (Nearey and Assmann, 1986), was found to improve vowel intelligibility in experimental listening tasks compared to formant frequency information contained in a steady portion, typically close to the vowel's center. As shown in a number of acoustic studies pertaining primarily to North American and Australian English (Andruski and Nearey 1992, Hillenbrand *et al.* 1995, Watson and Harrington, 1999), VISC is an inherent property of most English vowels, which invited questions about its perceptual relevance to listeners.

Experiments which used synthetic signals with edited static targets widely known as "silent-center vowels" brought to light that static vowel targets are not necessary and may not be sufficient for vowel identification (e.g., Jenkins *et al.*, 1983; Strange *et al.*, 1983; Strange, 1989; Fox, 1989; Andruski and Nearey, 1992). Rather, listeners make use of information in formant transitions close to vowel's onset and offset as well as in preserved formant contours. In particular, Hillenbrand and Nearey (1999) presented listeners with three types of signals: (1) naturally produced vowels in an /hVd/-context, (2) synthesized versions of these signals with formant contours generated from the original measurements, and (3) synthesized versions with flat formants generated from values at the steady portion of the natural vowels. While the natural tokens yielded the highest intelligibility scores, there was a large drop in identification accuracy from the signals containing formant contours to the flat-formant vowels. These results provided essential evidence that VISC is a source of important cues to vowel identity and listeners make use of them.

Statistical pattern recognition studies also showed that vowels are classified with greater accuracy when the

[a]Author to whom correspondence should be addressed. Electronic mail: jacewicz.1@osu.edu

recognition model incorporates spectral changes compared to single measurement of formant pattern taken at the vowel's center. The recognition models based on dynamic rather than static spectral features were also in better agreement with the accuracy and confusions obtained from human listeners (e.g., Zahorian and Jagharghi, 1993; Hillenbrand *et al.,* 1995; Hillenbrand and Nearey, 1999; Hillenbrand *et al.,* 2001). The general conclusion from this modeling work is that two samples of the formant pattern along with vowel duration and information in fundamental frequency yield relatively high classification accuracy. Feature combinations based on a single sample at steady-state yield lower accuracy compared to the two-sample points. Adding a third sample does not result in further significant improvement.

The rate and pattern of listeners' identification responses in two widely known studies by Peterson and Barney (1952) and its replication and extension by Hillenbrand *et al.* (1995) deserve a closer examination with respect to the dynamic vowel structure. In both studies, the overall identification rates for vowels in /hVd/-context were high (94.4 and 95.4 %, respectively) and only several vowels were susceptible to confusions. The generally high intelligibility of all vowels was maintained despite acoustic overlap of neighboring vowels. While duration differences can reasonably explain high identification rates for selected vowels, the differences in the amount of VISC and direction of formant movement may have also contributed to a greater perceptual separation of the overlapping spectral neighbors.

A subsequent detailed analysis in Neel (2008) inquired into the relationship between the acoustic characteristics of vowels in Hillenbrand *et al.* (1995) and the identification scores from their listeners. The effectiveness of several global and fine-grained measures in predicting the proportion of correct identification responses across listeners was evaluated. However, none of the measures (singly or in combination) accounted for more than 18% of the variance in identification scores and, even then, only larger vowel spaces, differences in formant movement and greater duration contrast between long and short vowels were found to be the best predictors of listeners' identification scores. A subsequent analysis of errors for vowels which were most often confused for one another revealed that dynamic vowel cues may have contributed to intelligibility of individual speakers. While more work needs to be done to fully understand the basis for listeners' identification decisions, the Neel study makes it clear that systematic differences in formant movement can, indeed, affect vowel intelligibility.

However, while results suggesting the effectiveness of dynamic rather than static vowel measures are convincing, there is a need for a careful control of dialect of both the speaker and the listener in vowel recognition experiments. As both Hillenbrand *et al.* (1995) and Neel (2008) point out, lower identification accuracy and increased confusions among vowels may be in part due to dialect variation. As already pointed out, American English exhibits considerable amount of regional variation (e.g., Labov *et al.,* 2006). Moreover, recent work in this area has found that regional vowel variants differ not only in their relative positions in the acoustic space but also in the nature and the amount of formant movement (Fox and Jacewicz, 2009). In particular, variation in VISC contributes to the differentiation of both nominal monophthongs and diphthongs in three distinct dialects spoken in Ohio, North Carolina, and Wisconsin.

Another factor that interacts with the dialect-specific variation in VISC is the diachronic sound change which is evident in the productions of native residents in a given dialect area. For example, the vowel system of the American South shows a chain-like vowel rotation termed the Southern Shift (e.g., Thomas, 2001; Labov *et al.,* 2006). In this dialect, the relative positions of selected vowels are "reversed" in the acoustic space so that the /ɛ/ in *debt* produced by older talkers tends to be more fronted than the /e/ in *date*. This ɛ/e reversal is gradually receding in younger talkers and /e/ is undergoing fronting while /ɛ/ is backing. What is intriguing here is the fact that both vowels differ significantly in the nature and the amount of VISC. This feature may be especially important in maintaining the vowel contrast when both vowels overlap acoustically in the process of generational sound change while reversing the relative positions in the acoustic space. Further discussion and extensive evidence for the cross-dialectal and cross-generational variation in the amount of VISC can be found in Jacewicz *et al.* (2011a; 2011c).

The concern for addressing the effects of dialect and generational variation in vowel recognition experiments is also substantiated by the results of a replication of Peterson and Barney (1952) reported in Labov (2010), which tightly controlled for dialect variation. In that study, listeners representing three distinct varieties spoken in Chicago, Birmingham (Alabama), and Philadelphia responded to stimuli produced by speakers of these dialects. Both within- and cross-dialect conditions were included. The overall identification accuracy reached only 77%, being much lower than in both Peterson and Barney (1952) and Hillenbrand *et al.* (1995). For each dialect, local listeners (i.e., participants who spoke that particular regional variety) had higher identification rates compared to non-local listeners. These results along with confusion patterns indicate that dialect misalignment does affect the recognition of vowel identity and dialectal sound change also reduces intelligibility.

The present study explicitly examines the effects of dialect and generational variation (sound change) on the recognition of vowel identity. Given that these two sociolinguistic variables also contribute to the variations in VISC as discussed above, the study has three aims. First, it explicates the nature of the variation in formant dynamics on the basis of acoustic measurements. Second, it assesses within- and cross-dialect vowel identification as a function of generational sound change. Finally, it examines statistical pattern recognition in relation to the variation in VISC.

II. METHODS

A. Listeners

Listeners were 30 middle-aged adults (43–58 years old) who were born, raised and spent most of their lives in either Western North Carolina (15 listeners, 8 males, 7 females, the Jackson County area) or Southeastern Wisconsin (15 listeners,

7 males, 8 females, Madison and its suburbs). They were mostly professionals recruited through printed study advertisements, local radio announcements, and word of mouth. Each listener spoke a local regional variety which was verified by the research staff on the basis of an informal interview. In addition, each listener completed a background questionnaire which also inquired into the residential history, education, and frequency of travel out of state. All reported normal hearing. None of them had participated in a perception experiment before and none was phonetically trained. The majority had a college-level education except for two in North Carolina and one in Wisconsin, who completed high school. One female listener from North Carolina and one male from Wisconsin were excluded from the study during data analysis because their responses to both the practice and experimental trials indicated that they were unable to do the task.

B. Stimuli

The stimuli for perceptual testing were naturally produced utterances *heed, hid, hayed, head, had, hod, who'd, heard, hide, hoed, hood,* and *hawed* containing 12 vowels: /i, ɪ, e, ɛ, æ, ɑ, u, ɝ, aɪ, o, ʊ, ɔ/. While the vowel /ʌ/ in *hud* could also be of interest, it was not included in the set due to an oversight. The tokens were selected from a large corpus of recordings completed for a production study of regional variation in American English vowels (see Jacewicz *et al.,* 2011a; 2011b for further details including a description of the recording procedure). In this corpus, speech samples were collected from nearly 400 participants ranging in age from 8 to 93 years old, which represented several stages of dialect-specific cross-generational sound (vowel) change. For the present perception study, a subset of these recordings was utilized, in which each speaker produced citation form *hVd*-utterances in isolation. Both the North Carolina (NC) and Wisconsin (WI) speakers came from the same geographic locations as the listeners and also spoke a regional variety typical of the area.

For the purposes of the perception study, the selected speakers represented three generations of long-time residents in each geographic area, which were divided into three age groups (in years): children (8–12), young adults (35–50), and old adults (66–91). This division reflects the condition for transmission of the diachronic sound change, namely, that sound changes extend across several generations of speakers in a given speech community. Labov (2001, p. 416) formulated a general condition for sound change: "Children must learn to talk differently from their mothers, and these differences must be in the same direction in each succeeding generation." We therefore expect that children in the present study will produce different vowel variants than the generation of what could be their parents; in turn, the productions of the parents' generation should differ from those of an earlier generation. What is important here is the continuation of sound change within a given speech community which can be observed and assessed in generational increments. To emphasize this generational aspect, we adopted in this study the naming convention children (C), parents (P), and grandparents (GP). Based on the study records, the present selected speakers were not biologically related to one another but were born and raised in the same dialect area.

A total of 120 speakers produced the vowel tokens, 20 (10 males, 10 females) for each dialect and age group (10 speakers $\times$ 2 genders $\times$ 3 ages $\times$ 2 dialects). Each speaker produced six different vowel categories (half of the set of 12) and provided only one repetition of each. Utilizing only six out of 12 vowels from each speaker increased the number of different speakers in the stimulus set and thus elevated stimulus uncertainty. In this way, the stimulus material consisted of 720 unique vowel exemplars (6 vowel categories $\times$ 10 speakers $\times$ 2 genders $\times$ 3 ages $\times$ 2 dialects). The selection was completed by two experienced researchers and trained phoneticians (one of whom was very familiar with the NC dialect and the other with the WI dialect) based on the following auditory criteria: each utterance had a falling pitch, there were no obvious idiosyncrasies related to final stop release or voice characteristics, and the token was a representative example of dialect-specific pronunciation. Prior to presentation, all tokens were equalized for mean intensity.

C. Procedure

Each listener was tested individually in a quiet room at either Western Carolina University in Cullowhee, NC, or University of Wisconsin–Madison, WI. Two different female research assistants, one at each location, administered the experiment. Signals were delivered binaurally over Sennheiser HD600 headphones using similar experimental set-up at both locations. All 720 unique signals were presented in one session lasting approximately one hour, in three separate blocks of 240 tokens each. All conditions (vowels, speakers, genders, ages, and dialects) were randomized in each block. Prior to the experimental blocks, a 20-item practice was administered to each listener for familiarization with the experimental protocol. This was done to ensure that listeners were able to match the orthographic form with the sound (e.g., choosing "heed" upon hearing *who'd* was taken as evidence that the subject was unable to do the task). The goal was to make them comfortable with the task and make certain they understood the instructions. The tokens in the practice were different from those used in the actual testing. However, exemplars of all 12 vowels were included in the practice as were speakers from both dialects and three generations (children, men, and women). Because of the great variety of voices in the stimulus material (including children, women, and men of different ages) and the fact that the listeners were novices in terms of perceptual testing, they were allowed to adjust the intensity level on an individual basis during the practice run to ensure the most comfortable listening conditions. That is, after playing several trials, the experimenter asked if the words sounded "just right" (not too loud and not too soft) or whether they needed adjustment. As a default, the level was set at 70 dB SPL.

Both stimulus presentation and response collection were controlled by a custom MATLAB program. The same instructions were read for each listener at each testing site. The listeners were told that they would hear one word at a time and

were to indicate which word was played by selecting an appropriate response box on the monitor. They were further instructed that they would hear the word once and they should respond immediately upon hearing it. However, if they missed the word because of distraction or stimulus uncertainty, they were allowed to listen to the same signal one additional time and then guess if they were not certain which response to choose. No information was given about the dialect, age and gender of the speakers but listeners were told that they would hear many different voices in the experiment. Listeners responded by clicking with the mouse on one of 12 boxes (one for each token type) which displayed a given word (e.g., "heed," "hid," etc.) on the computer screen. The experiment was self-timed, i.e., the listener heard the next stimulus only after having responded to the current stimulus. They were allowed to take short breaks between experimental blocks.

III. RESULTS

A. Acoustic characteristics of stimulus vowels

We begin with description of the basic acoustic characteristics of the vowel tokens to explicate the types of variation in formant frequencies and duration with which the listeners were presented. To estimate the dynamic changes in the formant pattern, the first three formants were sampled at five equidistant temporal points (20-35-50-65-80 %) in a vowel. The five-point measurement proved useful in characterizing more complex spectral variations which are difficult to capture by sampling formant frequencies only twice such as close to vowel onset and offset (see Fox and Jacewicz, 2009, for further details and discussion). Formant values were based on 14-pole linear predictive coding analysis and were extracted automatically using MATLAB by centering a 25-ms Hanning window at each temporal point. Formant estimations were then hand corrected as needed during two reliability checks using smoothed FFT spectra and wideband spectrograms with formant tracking (TF32, Milenkovic, 2003).

To compare the measurements across men, women and children, the formant values in Hz for each individual speaker were converted to *z*-scores (Lobanov, 1971). This normalization procedure minimizes—although does not eliminate—variation caused by differences in vocal tract length while preserving dialectal and cross-generational differences. The Lobanov method was chosen as Adank *et al.* (2004) found it most effective in a comparison of 12 different normalization procedures. Another widely used procedure, that of Nearey (1978), placed second in this comparison. As a reference, the unnormalized mean Hz values for vowel midpoints are reported in Tables I and II.

Figure 1 shows means for NC speakers split by generation and gender. The GP productions indicate the presence of the Southern Shift, which is a salient regional feature of Southern English in the United States (see Labov *et al.*, 2006). It is manifested primarily in front vowels /i, ɪ, e, ɛ, æ/, which are heavily diphthongized and either overlap as in males (upper left) or have reversed locations of /e, ɛ/ as in females (upper right). In the Southern Shift, /æ/ is raised approximating the position of /ɛ/ and the diphthong /aɪ/ is produced as a monophthong. The fronted /u, ʊ/ and the

TABLE I. Average durations (ms) and formant frequencies (Hz) measured at midpoints (50%) of vowels produced by North Carolina speakers representative of each generation: Grandparents (GP), parents (P), and children (C). The average values are for five speakers within each generation and gender group.

Generation and gender		/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/ɝ/	/aɪ/	/o/	/ʊ/	/u/
GP male	Dur	308	279	325	293	351	289	332	279	323	339	346	312
	F1	318	407	480	428	568	699	816	391	430	669	441	370
	F2	2256	2060	2013	2056	1896	1029	1526	1345	878	1080	1320	1459
	F3	2833	2665	2590	2551	2463	2609	2420	2390	2478	2558	1604	2340
GP female	Dur	321	306	355	282	329	334	341	312	337	365	304	305
	F1	372	383	506	488	713	936	1005	439	530	821	580	452
	F2	2749	2730	2439	2594	2394	1249	1851	1578	1328	1223	1565	1752
	F3	3662	3490	3155	3286	3092	3068	3173	2976	2808	3131	1948	2909
P male	Dur	282	259	314	268	347	275	338	264	297	325	324	296
	F1	267	387	413	441	579	732	809	419	454	639	456	342
	F2	2394	2099	1998	2123	1905	1085	1399	1365	1218	902	1337	1520
	F3	3068	2644	2633	2689	2502	2644	2558	2486	2450	2616	1595	2351
P female	Dur	277	300	302	290	284	329	307	235	318	356	297	282
	F1	350	498	519	550	816	870	994	411	579	869	540	411
	F2	2857	2396	2360	2491	2183	1264	1567	1681	1520	1136	1496	1948
	F3	3486	3051	3005	3133	2915	2982	2840	2855	2788	3005	1884	2678
C male	Dur	254	243	289	249	316	286	318	240	281	304	269	320
	F1	417	484	512	611	910	923	1020	512	462	833	434	430
	F2	2930	2486	2612	2280	2210	1274	1623	1783	1513	1212	1698	2021
	F3	3574	3344	3402	3163	3156	3107	3087	3083	2786	3156	1970	2846
C female	Dur	300	260	274	254	299	307	297	291	361	326	287	327
	F1	424	549	630	732	1143	966	1104	557	552	941	696	465
	F2	3221	2708	2846	2534	2241	1419	1879	1866	1554	1244	1847	2131
	F3	3673	3346	3404	3313	3203	3057	3186	3130	3081	3163	2192	3036

TABLE II. Average durations (ms) and formant frequencies (Hz) measured at midpoints (50%) of vowels produced by Wisconsin speakers representative of each generation: Grandparents (GP), parents (P), and children (C). The average values are for five speakers within each generation and gender group.

Generation and gender		/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/ɝ/	/aɪ/	/o/	/ʊ/	/u/
GP male	Dur	240	167	255	186	247	293	290	169	295	292	277	253
	F1	264	455	373	585	628	788	592	510	435	661	391	329
	F2	2304	1772	2196	1701	1828	1300	1772	1162	787	1108	1266	856
	F3	2863	2357	2774	2506	2422	2578	2554	2332	2435	2673	1587	2295
GP female	Dur	278	171	299	209	321	293	303	187	319	346	313	291
	F1	303	508	420	672	760	953	834	522	406	777	480	404
	F2	2818	2234	2760	2151	2379	1506	2080	1255	785	1195	1518	949
	F3	3469	3044	3214	2963	3138	3006	2928	2941	3016	2840	1810	2877
P male	Dur	261	173	265	184	287	229	281	181	235	253	232	246
	F1	292	478	385	643	691	828	654	484	471	727	467	320
	F2	2334	1804	2166	1690	1757	1406	1644	1184	841	1091	1320	902
	F3	2946	2566	2665	2534	2452	2450	2431	2383	2575	2629	1621	2288
P female	Dur	252	186	255	206	287	273	268	197	294	281	263	222
	F1	313	545	372	775	843	997	749	562	506	872	527	363
	F2	2728	2217	2650	2002	1970	1569	2011	1461	955	1259	1526	956
	F3	3438	2972	3109	2866	2814	2788	2956	2683	2795	2760	1748	2655
C male	Dur	222	184	256	165	250	316	283	180	275	313	272	222
	F1	330	499	534	860	1005	1100	889	608	545	997	533	374
	F2	3240	2596	2906	2426	2213	1589	2206	1776	1091	1438	1733	1186
	F3	3876	3681	3497	3502	2986	3221	3041	3161	3077	3122	2103	3473
C female	Dur	303	184	283	208	286	289	292	246	307	320	257	266
	F1	381	586	536	889	992	1124	925	715	589	1042	584	466
	F2	3186	2667	3090	2426	2347	1849	2262	1871	1121	1442	1853	1203
	F3	3632	3600	3611	3458	3428	3143	3322	3352	3236	3253	2183	3320

upgliding diphthongal /ɔ/ (known as the back upglide) are further characteristics of the Southern English. All these features are present in GP. They still can be found in the younger P-generation but their vowel system shows signs of a restructuring which becomes evident in C. In particular, children's vowels in the lower panels have a reduced formant movement in /i, ɪ, e, ɛ, æ/, their variant of /æ/ is well separated from /ɛ/, the /ɔ/ is not produced as an upglide anymore (particularly in girls) and the /aɪ/ has a formant movement more typical of a diphthong, again greater in girls than in boys. These changes in the vowel system indicate that the Southern features are receding in children.

The mean formant values for WI speakers are displayed in Fig. 2. The plots show some features of the Northern Cities Shift (NCS), a vowel chain shift affecting a relatively broad area in the northern United States (e.g., Gordon, 2004; Labov *et al.*, 2006). For example, the WI /æ/ in Fig. 2 is in a close proximity to /ɛ/ (including children) and has a different direction of formant movement compared to the NC /æ/ known as Northern breaking (Labov *et al.*, 2006). The low and fronted position of WI /ɑ/—another feature of the NCS—is in contrast with the NC variant in Fig. 1 which is positioned comparatively higher and more back in the vowel space. The final set of WI features which contrast with those in NC include the far back position of /u/, monophthongal versions of /e, o/, the full diphthong /aɪ/ and a different pattern of formant dynamics in /ɔ/.

As reported elsewhere (Jacewicz *et al.*, 2011a; 2011c), WI monophthongs have on average reduced formant movement relative to the NC monophthongs and, in both vowel systems, children's productions show further reduction in formant dynamics compared to adults. It is therefore of interest to the present study whether and how such cross-dialectal and generational variation in spectral dynamics along with positional differences within the two vowel systems affect the recognition of vowel identity.

Vowel duration varied as a function of dialect, age group, gender, and vowel category. NC vowels were on average longer ($M = 302$ ms) than WI vowels ($M = 254$ ms), the oldest adults produced longer vowels ($M = 291$ ms) than young adults ($M = 270$ ms) and children ($M = 274$ ms); female vowels were longer ($M = 286$ ms) than male ($M = 270$ ms). In terms of individual vowel categories, most average durations were between 275–304 ms. The shortest durations were for /ɪ, ʊ, ɛ/ ($M = 226$, 232, and 232 ms, respectively) and the /ɔ/ was the longest vowel in the set ($M = 318$ ms). More details about durations of individual vowels can be found in Tables I and II.

B. Vowel identification results

Table III summarizes average identification rates (IDRs) for NC and WI listeners in response to vowels of their own (native) dialect and of the other (non-native) dialect. The IDRs were slightly higher overall for the native dialect although this general trend was not true for all individual vowel categories. The overall IDR of 80% approximated the overall IDR of 77% in a dialect-controlled experiment reported in Labov (2010). However, the present rates were about 15% lower than those in Hillenbrand *et al.* (1995)—also included in the table—although similar responses were

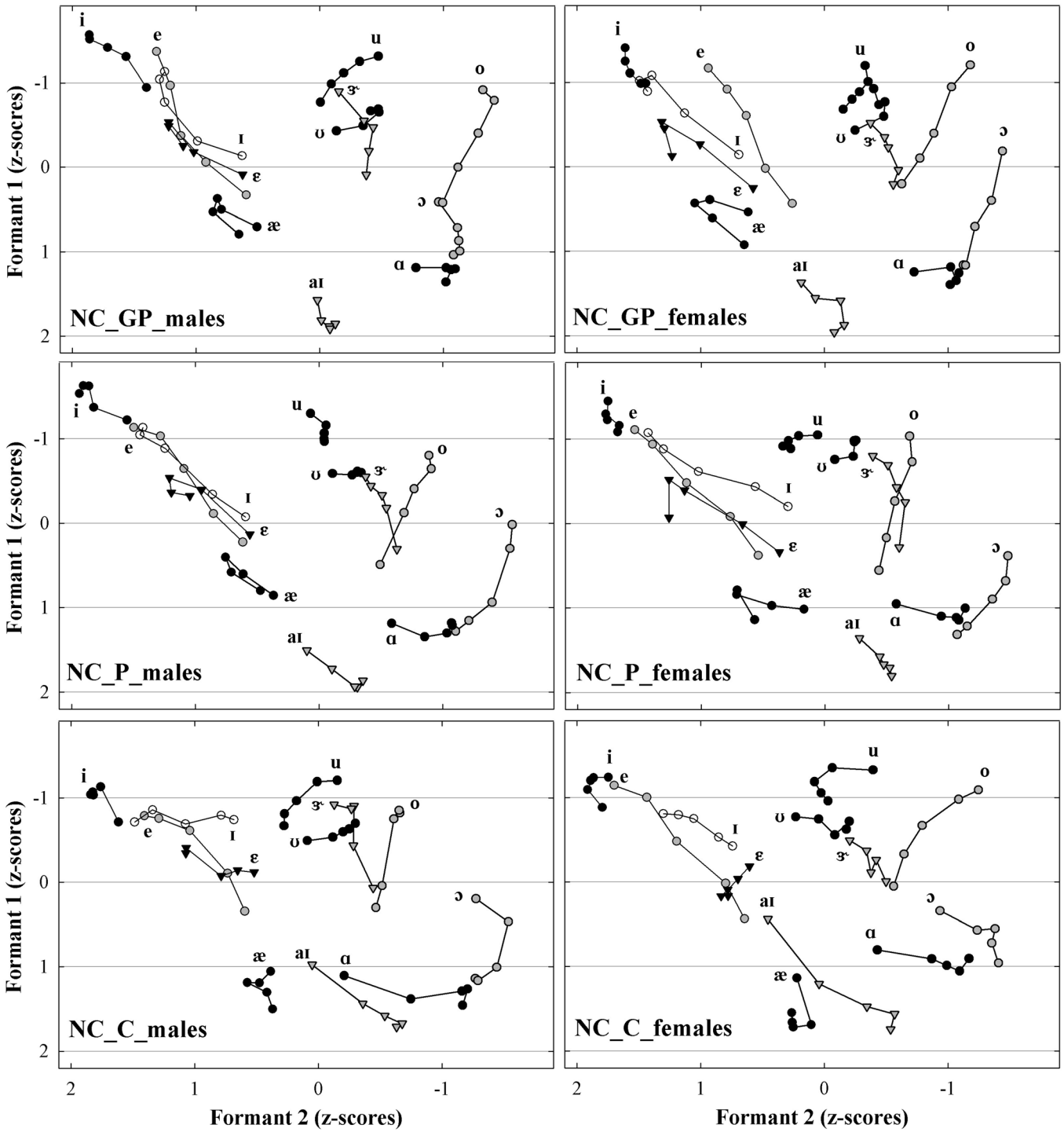

FIG. 1. Normalized formant frequencies for North Carolina stimulus tokens produced by three generations of speakers: Grandparents (GP), parents (P), and children (C). Each data point represents a mean for five speakers. Each vowel symbol is placed near the 80%-measurement point.

obtained for /ɝ, o, ʊ/. While it may be expected that the rhotacized vowel in *heard* will be well identified due to its *r*-coloring and thus unaffected by dialectal and generational variations, the higher IDRs for /o/ and /ʊ/ compared to the remaining vowels in the set are not easily interpretable.

Compared to Hillenbrand *et al*. (1995), the IDRs for /ɔ/ and /ɑ/ seem particularly low. However, much lower accuracy for /ɔ/ was also reported by Labov (2010, pp. 52–58). In particular, Chicago listeners identified correctly their local Chicago variant of /ɔ/ 40% of the time and their IDRs for the non-local Philadelphia and Birmingham variants were even lower, reaching only 11 and 8 %, respectively. Overwhelmingly, the confusions were with /ɑ/ . In response to the same stimuli, the IDRs percentages for Birmingham listeners were higher (48, 37, and 74) and they were still higher for Philadelphia listeners (82, 89, and 76). Thus, only the results for the Philadelphia listeners approximated those in Hillenbrand *et al*. (1995). Labov's study makes it clear that dialect variation does contribute to differential recognition of vowel identity, which is also the case in the present study.

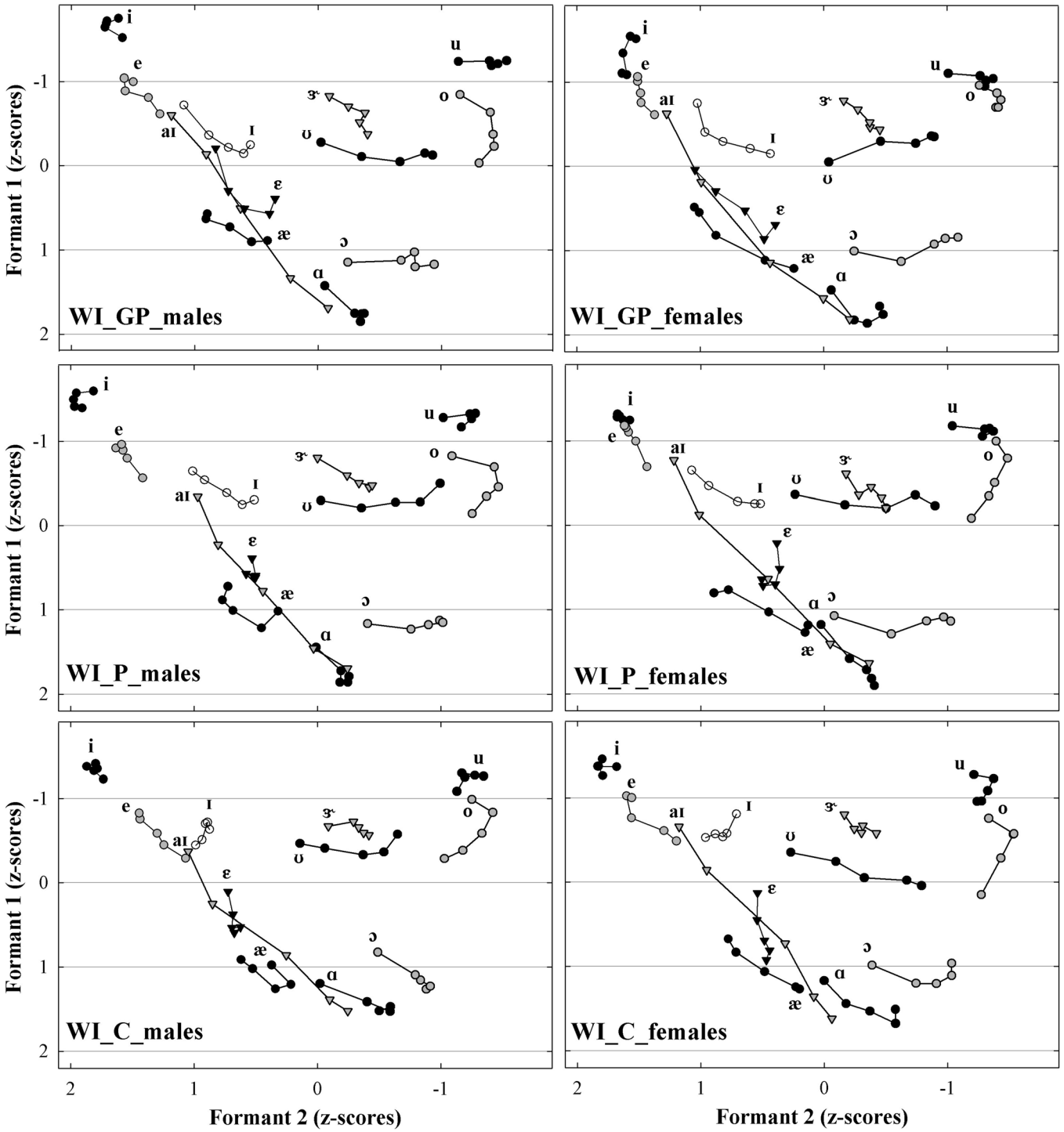

FIG. 2. Normalized formant frequencies for Wisconsin stimulus tokens produced by three generations of speakers: Grandparents (GP), parents (P), and children (C). Each data point represents a mean for five speakers. Each vowel symbol is placed near the 80%-measurement point.

The lower accuracy for the remaining vowels in Table III, particularly /aɪ, e, ɪ/, is also related to dialect variation. These results will be discussed in more detail below.

Overall IDRs were assessed by a repeated-measures analysis of variance (after arcsine transformation, Studebaker, 1985). Of interest is the significant listener dialect by speaker dialect interaction [$F(1,\ 28) = 20.14,\ p < 0.001,\ \eta^2 = 418$] which showed native dialect advantage in that IDRs were higher for NC listeners responding to NC speakers and WI listeners responding to WI speakers. Speaker generation was a significant factor showing the highest IDRs for P-groups which were of the same age as the listeners [$F(1.4,\ 38) = 6.6,\ p = 0.008,\ \eta^2 = 191$]. Speaker gender was also significant and higher IDRs were obtained for females [$F(1,\ 28) = 57.95,\ p < 0.001,\ \eta^2 = 674$].

1. Identification results for North Carolina listeners

The vowel IDRs and confusion matrices for NC listeners broken down by speaker dialect, age group and gender

TABLE III. Overall identification rates (in % correct) by vowel category for each listener group responding to both native and non-native dialect vowels. Results by Hillenbrand *et al*. (1995) are included for a comparison.

Vowel intended by speaker	NC listeners		WI listeners		Total	Hillenbrand *et al*. (1995)
	Native dialect	Non-native dialect	Native dialect	Non-native dialect		
/i/	82.0	86.7	95.1	83.8	86.9	99.6
/ɪ/	70.9	72.9	77.8	73.6	73.8	98.8
/e/	86.9	45.6	61.6	85.1	69.8	98.3
/ɛ/	90.2	82.7	88.7	77.8	84.8	95.1
/æ/	85.3	68.4	89.3	93.1	84.1	94.1
/ɑ/	55.8	33.8	74.9	67.3	57.9	92.3
/ɔ/	74.7	45.1	36.4	60.9	54.3	82.0
/ɝ/	99.1	98.7	98.9	98.0	98.7	99.5
/aɪ/	52.9	88.2	95.1	40.7	69.2	—
/o/	97.6	95.8	97.1	95.6	96.5	99.2
/ʊ/	97.8	96.9	96.9	96.0	96.9	97.5
/u/	91.8	94.2	90.2	83.3	89.9	97.2
Total	82.1	75.8	83.5	79.6	80.2	95.8

are shown in Table IV. We are particularly interested in vowels which were identified poorly as well as in their confusion patterns. We will first focus on the back vowels /ɔ, ɑ, aɪ/. The IDRs for /ɔ/ can be linked to the acoustic pattern in Fig. 1. Namely, when the NC /ɔ/ was produced as a back upglide in P and GP-females, it yielded the highest IDRs. The accuracy then dropped for GP-males and C, most likely in response to the reduced amount of VISC combined with a different direction of formant movement in their variants of /ɔ/. This indicates that NC listeners were familiar with the upgliding /ɔ/ which was confused less often with other vowels. However, increased confusions with /ɑ/, the spectral neighbor, arose most likely due to insufficient spectral contrast between /ɔ/ and /ɑ/ when the former was produced as a more typical variety by GP-males and C. The low IDRs in response to the WI variants of /ɔ/ would support this interpretation. The further drop in accuracy for WI C suggests that the spectral contrast between the two vowels was even more reduced when the children's /ɑ/ was produced in a manner similar to /ɔ/ (see Fig. 2).

The low IDRs for NC /ɑ/ and its predominant confusions with /ɔ/ can be similarly explained on the basis of spectral proximity of both vowels. However, the low IDRs for WI /ɑ/ bring yet another pattern of confusions, particularly in response to WI P and C, which were identified as /aɪ/ more often than /ɔ/. These confusions can be explained on the basis of the spectral overlap of the monophthongal NC /aɪ/ with the WI /ɑ/ (compare Figs. 1 and 2).

The sensitivity of NC listeners to the amount of formant movement in /aɪ/ is further manifested in their differential response to this vowel as a function of speaker generation and dialect. While the IDRs were low in response to its monophthongal variants in NC GP, P, and C-males (because it was confused with either /æ/ or /ɑ/), the accuracy improved dramatically in response to NC C-females who produced it as a more diphthongal variant (see Fig. 1). Consequently, the accuracy was high in response to WI /aɪ/ which was produced as a full diphthong.

As for the front vowels, we find a striking dialect-related discrepancy in IDRs for /e/. While the accuracy was generally high for the NC variants, it was much lower, mostly below 50% correct, for WI. Differences in VISC seem to provide the most plausible explanation of this pattern. That is, the heavily diphthongized NC variants (see Fig. 1) were rarely confused with their spectral neighbors whereas the monophthongal WI /e/ (Fig. 2) was more often misidentified as /i/ (when produced by females) and /ɛ/ (when produced by males). While the source of this gender-related variation needs to be determined, the amount of spectral change in a vowel is most likely a salient cue in guiding listeners' identification choices.

Lower IDRs were also apparent for /ɪ/, both for NC and WI variants, and much variation was associated with speaker generation and gender. In general, the acoustic proximity of the front vowels /i, ɪ, e, ɛ/ and their positional variations as a function of dialect and sound change created an especially crowded spectral neighborhood. It could be the case that NC /ɪ/ was confused with /ɛ/ because of the nature of its spectral change, in which the initial portion of the vowel started unambiguously as /ɪ/ and transitioned to /ɛ/ over the time course of the vowel [this dialect feature is described in the literature as Southern breaking, in which the vowel "breaks" into two parts (Sledd, 1966)]. Apparently, for some listeners, information in the latter portion of the vowel was the dominant cue in making their labeling decisions. Dialectal differences in the direction of formant movement could have also contributed to the lower IDRs for the WI variants of /æ/. That is, the unfamiliarity of NC listeners with the Northern breaking in WI /æ/ (manifested as initially starting with /ɛ/ and transitioning to /æ/, see Fig. 2) could have led to mislabeling of the vowel as /ɛ/. These possibilities await further experimental exploration.

Assessing the differences in IDRs for NC listeners, we took into consideration the above confusion patterns and analyzed three subsets of vowels (rather than the entire set) in order to better understand the main effects and interactions

TABLE IV. 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/	/ɝ/	/aɪ/	/o/	/ʊ/	/u/
(a) North Carolina listeners: Responses to North Carolina GP speakers													
Male	/i/	**49.3**	1.3	44.0	4.0	—	—	—	—	—	1.3	—	—
	/ɪ/	1.3	**60.0**	—	37.3	—	—	—	—	1.3	—	—	—
	/e/	2.7	—	**89.3**	5.3	1.3	—	—	—	1.3	—	—	—
	/ɛ/	—	—	1.3	**96.0**	1.3	1.3	—	—	—	—	—	—
	/æ/	—	—	1.3	20.0	**74.7**	4.0	—	—	—	—	—	—
	/ɑ/	—	—	1.3	—	1.3	**50.7**	45.3	—	—	—	—	1.3
	/ɔ/	—	—	2.7	—	1.3	25.3	**62.7**	—	4.0	2.7	—	1.3
	/ɝ/	—	—	—	1.3	—	—	—	**98.7**	—	—	—	—
	/aɪ/	—	—	—	—	37.3	6.7	—	—	**56.0**	—	—	—
	/o/	1.3	—	—	—	—	—	—	—	—	**98.7**	—	—
	/ʊ/	—	—	—	—	—	—	—	—	—	—	**100.0**	—
	/u/	—	—	—	1.3	—	—	—	—	—	6.7	2.7	**89.3**
Female	/i/	**86.7**	—	12.0	1.3	—	—	—	—	—	—	—	—
	/ɪ/	1.3	**54.7**	2.7	41.3	—	—	—	—	—	—	—	—
	/e/	2.7	—	**85.3**	1.3	1.3	—	—	2.7	1.3	2.7	2.7	—
	/ɛ/	—	10.7	2.7	**82.7**	4.0	—	—	—	—	—	—	—
	/æ/	—	—	—	17.3	**82.7**	—	—	—	—	—	—	—
	/ɑ/	—	—	1.3	—	2.7	**65.3**	26.7	—	4.0	—	—	—
	/ɔ/	—	—	1.3	—	—	8.0	**88.0**	—	—	2.7	—	—
	/ɝ/	—	—	—	—	—	—	—	**100.0**	—	—	—	—
	/aɪ/	—	1.3	1.3	—	40.0	4.0	4.0	—	**49.3**	—	—	—
	/o/	—	—	—	—	—	—	—	—	—	**97.3**	2.7	—
	/ʊ/	—	—	—	—	—	—	—	—	—	—	**100.0**	—
	/u/	—	—	—	—	—	—	—	—	—	16.0	6.7	**77.3**
(b) North Carolina listeners: Responses to North Carolina P speakers													
Male	/i/	**74.7**	1.3	21.3	2.7	—	—	—	—	—	—	—	—
	/ɪ/	—	**85.3**	—	12.0	—	—	—	—	2.7	—	—	—
	/e/	4.0	1.3	**81.3**	4.0	1.3	—	2.7	—	4.0	—	1.3	—
	/ɛ/	—	4.0	—	**93.3**	2.7	—	—	—	—	—	—	—
	/æ/	—	—	—	13.3	**84.0**	2.7	—	—	—	—	—	—
	/ɑ/	—	—	—	—	—	**60.0**	38.7	—	—	—	—	1.3
	/ɔ/	—	—	—	—	—	16.0	**78.7**	—	—	6.7	—	—
	/ɝ/	—	—	—	—	—	—	—	**100.0**	—	—	—	—
	/aɪ/	—	1.3	2.7	—	13.3	18.7	9.3	—	**53.3**	—	—	1.3
	/o/	—	—	—	—	—	1.3	—	—	—	**94.7**	1.3	2.7
	/ʊ/	—	—	—	—	—	—	—	—	—	—	**98.7**	1.3
	/u/	—	—	—	—	—	—	1.3	—	—	6.7	—	**92.0**
Female	/i/	**97.3**	1.3	—	—	—	1.3	—	—	—	—	—	—
	/ɪ/	—	**80.0**	1.3	16.0	—	—	—	—	1.3	—	1.3	—
	/e/	1.3	—	**93.3**	2.7	—	—	—	—	2.7	—	—	—
	/ɛ/	—	—	—	**92.0**	8.0	—	—	—	—	—	—	—
	/æ/	—	—	—	5.3	**93.3**	1.3	—	—	—	—	—	—
	/ɑ/	—	—	1.3	—	2.7	**49.3**	36.0	—	10.7	—	—	—
	/ɔ/	—	—	4.0	—	—	5.3	**89.3**	—	—	1.3	—	—
	/ɝ/	—	—	—	—	—	—	—	**100.0**	—	—	—	—
	/aɪ/	—	4.0	1.3	—	30.7	17.3	4.0	—	**42.7**	—	—	—
	/o/	—	—	—	—	—	—	—	—	—	**97.3**	—	2.7
	/ʊ/	—	—	—	—	—	2.7	—	—	—	—	**97.3**	—
	/u/	—	—	—	—	—	—	—	—	—	1.3	1.3	**97.3**
(c) North Carolina listeners: Responses to North Carolina C speakers													
Male	/i/	**93.3**	1.3	4.0	1.3	—	—	—	—	—	—	—	—
	/ɪ/	—	**78.7**	—	16.0	—	—	—	—	1.3	—	4.0	—
	/e/	1.3	—	**78.7**	8.0	1.3	—	—	4.0	1.3	1.3	1.3	2.7
	/ɛ/	—	6.7	—	**82.7**	—	—	—	—	—	—	6.7	4.0
	/æ/	—	—	—	13.3	**84.0**	2.7	—	—	—	—	—	—
	/ɑ/	—	—	—	—	—	**61.3**	34.7	—	2.7	1.3	—	—
	/ɔ/	—	—	2.7	—	—	22.7	**65.3**	—	—	9.3	—	—

TABLE IV. (*Continued*)

Speaker gender	Vowel intended by speaker	Vowel identified by listeners											
		/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/ɝ/	/aɪ/	/o/	/ʊ/	/u/
	/ɝ/	—	—	—	1.3	—	—	—	**98.7**	—	—	—	—
	/aɪ/	—	2.7	2.7	—	14.7	29.3	14.7	1.3	**33.3**	—	1.3	—
	/o/	—	—	—	—	—	1.3	—	—	—	**97.3**	—	1.3
	/ʊ/	—	—	—	1.3	—	—	—	—	—	—	**94.7**	4.0
	/u/	—	—	—	—	—	—	—	1.3	—	1.3	—	**97.3**
Female	/i/	**90.7**	5.3	2.7	1.3	—	—	—	—	—	—	—	—
	/ɪ/	—	**66.7**	—	32.0	—	—	—	—	1.3	—	—	—
	/e/	2.7	—	**93.3**	4.0	—	—	—	—	—	—	—	—
	/ɛ/	—	1.3	—	**94.7**	2.7	1.3	—	—	—	—	—	—
	/æ/	—	—	—	4.0	**93.3**	1.3	—	—	1.3	—	—	—
	/ɑ/	—	—	—	—	—	**48.0**	33.3	—	18.7	—	—	—
	/ɔ/	—	—	—	—	—	26.7	**64.0**	—	8.0	—	—	1.3
	/ɝ/	—	—	—	1.3	1.3	—	—	**97.3**	—	—	—	—
	/aɪ/	—	—	1.3	—	13.3	1.3	—	1.3	**82.7**	—	—	—
	/o/	—	—	—	—	—	—	—	—	—	**100.0**	—	—
	/ʊ/	—	—	—	—	—	—	—	1.3	—	—	**96.0**	2.7
	/u/	—	—	—	—	—	—	—	1.3	—	1.3	—	**97.3**
(d) North Carolina listeners: Responses to Wisconsin GP speakers													
Male	/i/	**89.3**	6.7	—	4.0	—	—	—	—	—	—	—	—
	/ɪ/	—	**72.0**	—	26.7	—	—	—	—	—	—	1.3	—
	/e/	6.7	12.0	**49.3**	30.7	—	—	—	—	—	—	1.3	—
	/ɛ/	—	1.3	—	**94.7**	1.3	—	—	—	—	—	2.7	—
	/æ/	—	1.3	—	40.0	**58.7**	—	—	—	—	—	—	—
	/ɑ/	—	1.3	2.7	—	9.3	**53.3**	14.7	1.3	17.3	—	—	—
	/ɔ/	—	—	—	—	—	41.3	**53.3**	—	4.0	—	—	1.3
	/ɝ/	—	—	—	—	—	—	—	**100.0**	—	—	—	—
	/aɪ/	1.3	4.0	10.7	—	1.3	—	—	—	**82.7**	—	—	—
	/o/	—	—	—	—	—	1.3	—	—	—	**96.0**	1.3	1.3
	/ʊ/	—	—	—	—	—	—	—	—	—	—	**100.0**	—
	/u/	—	—	—	—	—	—	—	—	—	1.3	17.3	**81.3**
Female	/i/	**74.7**	25.3	—	—	—	—	—	—	—	—	—	—
	/ɪ/	—	**98.7**	—	1.3	—	—	—	—	—	—	—	—
	/e/	38.7	4.0	**37.3**	20.0	—	—	—	—	—	—	—	—
	/ɛ/	—	14.7	—	**84.0**	1.3	—	—	—	—	—	—	—
	/æ/	—	—	—	56.0	**44.0**	—	—	—	—	—	—	—
	/ɑ/	—	—	1.3	—	9.3	**46.7**	25.3	—	17.3	—	—	—
	/ɔ/	—	—	—	—	—	37.3	**61.3**	—	1.3	—	—	—
	/ɝ/	—	—	—	—	—	—	—	**100.0**	—	—	—	—
	/aɪ/	1.3	5.3	4.0	—	—	—	—	—	**89.3**	—	—	—
	/o/	—	—	1.3	—	—	1.3	—	—	—	**92.0**	1.3	4.0
	/ʊ/	—	—	—	—	—	—	—	—	—	—	**100.0**	—
	/u/	—	—	—	—	—	—	—	—	—	—	2.7	**97.3**
(e) North Carolina listeners: Responses to Wisconsin P speakers													
Male	/i/	**93.3**	5.3	1.3	—	—	—	—	—	—	—	—	—
	/ɪ/	—	**60.0**	—	40.0	—	—	—	—	—	—	—	—
	/e/	8.0	4.0	**36.0**	52.0	—	—	—	—	—	—	—	—
	/ɛ/	—	—	—	**68.0**	29.3	1.3	—	1.3	—	—	—	—
	/æ/	—	—	—	24.0	**76.0**	—	—	—	—	—	—	—
	/ɑ/	—	1.3	1.3	—	28.0	**24.0**	6.7	—	37.3	—	—	1.3
	/ɔ/	—	—	1.3	—	—	52.0	**46.7**	—	—	—	—	—
	/ɝ/	—	—	—	—	—	—	—	**98.7**	—	—	—	1.3
	/aɪ/	—	1.3	2.7	—	1.3	—	—	—	**94.7**	—	—	—
	/o/	—	—	—	—	—	—	—	—	—	**97.3**	—	2.7
	/ʊ/	—	—	—	—	—	—	—	—	—	—	**100.0**	—
	/u/	—	—	—	—	—	—	—	—	—	—	4.0	**96.0**
Female	/i/	**84.0**	13.3	1.3	1.3	—	—	—	—	—	—	—	—
	/ɪ/	—	**94.7**	—	5.3	—	—	—	—	—	—	—	—
	/e/	30.7	5.3	**44.0**	20.0	—	—	—	—	—	—	—	—

TABLE IV. (*Continued*)

Speaker gender	Vowel intended by speaker	Vowel identified by listeners											
		/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/ɝ/	/aɪ/	/o/	/ʊ/	/u/
	/ɛ/	—	—	—	**90.7**	9.3	—	—	—	—	—	—	—
	/æ/	—	—	—	28.0	**72.0**	—	—	—	—	—	—	—
	/ɑ/	—	1.3	1.3	4.0	13.3	**29.3**	12.0	—	38.7	—	—	—
	/ɔ/	—	—	—	—	—	50.7	**46.7**	—	2.7	—	—	—
	/ɝ/	—	—	—	—	—	—	—	**100.0**	—	—	—	—
	/aɪ/	—	2.7	4.0	2.7	—	—	1.3	—	**89.3**	—	—	—
	/o/	—	—	—	—	—	—	—	—	—	**100.0**	—	—
	/ʊ/	—	—	—	—	—	1.3	—	—	—	—	**96.0**	2.7
	/u/	—	—	—	—	—	—	—	—	—	2.7	—	**97.3**
(f) North Carolina listeners: Responses to Wisconsin C speakers													
Male	/i/	**90.7**	9.3	—	—	—	—	—	—	—	—	—	—
	/ɪ/	—	**58.7**	—	33.3	—	—	—	—	1.3	—	5.3	1.3
	/e/	10.7	6.7	**58.7**	21.3	—	1.3	1.3	—	—	—	—	—
	/ɛ/	—	—	—	**80.0**	14.7	2.7	—	—	1.3	—	1.3	—
	/æ/	—	—	—	20.0	**76.0**	4.0	—	—	—	—	—	—
	/ɑ/	—	—	2.7	—	9.3	**33.3**	20.0	—	34.7	—	—	—
	/ɔ/	—	1.3	1.3	—	1.3	46.7	**26.7**	—	18.7	—	4.0	—
	/ɝ/	—	1.3	—	—	—	—	—	**98.7**	—	—	—	—
	/aɪ/	1.3	1.3	4.0	—	6.7	—	—	—	**86.7**	—	—	—
	/o/	—	—	—	—	—	—	—	—	—	**97.3**	—	2.7
	/ʊ/	—	—	—	—	1.3	1.3	—	4.0	—	—	**92.0**	1.3
	/u/	—	—	—	—	—	—	—	1.3	—	2.7	1.3	**94.7**
Female	/i/	**88.0**	12.0	—	—	—	—	—	—	—	—	—	—
	/ɪ/	—	**53.3**	—	46.7	—	—	—	—	—	—	—	—
	/e/	36.0	6.7	**48.0**	6.7	1.3	—	—	—	—	1.3	—	—
	/ɛ/	—	—	—	**78.7**	21.3	—	—	—	—	—	—	—
	/æ/	—	—	—	14.7	**84.0**	1.3	—	—	—	—	—	—
	/ɑ/	—	5.3	1.3	—	17.3	**16.0**	13.3	—	46.7	—	—	—
	/ɔ/	—	—	2.7	—	1.3	60.0	**36.0**	—	—	—	1.3	—
	/ɝ/	—	—	—	—	—	—	—	**94.7**	—	—	5.3	—
	/aɪ/	—	1.3	6.7	—	5.3	—	—	—	**86.7**	—	—	—
	/o/	—	—	—	—	—	1.3	4.0	—	—	**92.0**	—	2.7
	/ʊ/	—	—	—	1.3	—	2.7	—	1.3	—	—	**93.3**	1.3
	/u/	—	—	—	—	—	—	—	—	—	—	1.3	**98.7**

of several variables. We considered separately: (1) the four high and mid front vowels /i, ɪ, e, ɛ/ which were confused most often with one another; (2) the vowel /æ/ which was confused either with /ɛ/ or the low vowels /ɑ, aɪ/ to the exclusion of high front vowels; and (3) the low vowels /ɑ, ɔ, aɪ/ which were mutually confused. The remaining vowels /ɝ, o, ʊ, u/ were not assessed because of their high IDRs and rather obvious confusion patterns. The overall arcsine transformed IDRs for each vowel subset for 15 NC listeners were analyzed using repeated-measures analysis of variance (ANOVA) with the within-subject factors speaker dialect, generation and gender. Because of the multiple analyses, we adopted a more stringent alpha level of 0.01 to indicate statistical significance (to protect against type I error rate inflation). The analyses were followed by *post hoc t*-tests. A complete summary of significant main effects and interactions is provided in Table V.

Speaker dialect was significant for all five front vowels, uniformly showing that NC variants were identified significantly better than WI variants. There was no significant native dialect advantage for the back vowels, however. The significant main effect of generation was manifested differently for different vowel combinations. For the four front vowels, it is noteworthy that the worse IDRs were for GP, which was due not only to the low IDRS for the vowels discussed above but also to the lower IDRs for NC /i/ (both genders) and for WI /i/ in females. The lower IDRs for GP speakers may be again related to the amount of VISC because these variants, being more diphthongized, caused substantial confusions with the spectral neighbors (compare Figs. 1 and 2 and Table IV).

For the vowel /æ/, the IDRs were significantly lower for GP and higher for P- and C-groups, which did not differ significantly one from another. This pattern can be explained on the basis of an ongoing sound change which can be detected in P. Namely, the plots show a lowering and backing of /æ/ in the acoustic space compared to GP. NC listeners appeared to be sensitive to this sound change since they confused /æ/ less often with /ɛ/ in P and C. It could also be the case that the cross-generational change in VISC contributed to these results (compare the smaller spectral change in WI P /ɛ/ and NC C /æ/ which would enhance the /æ/-/ɛ/ contrast in both groups).

TABLE V. Summary of significant main effects and interactions from repeated-measures ANOVAs for overall identification rates. Shown are partial eta-squared values (η^2). — = not significant, NC = North Carolina, WI = Wisconsin, GP = grandparents, P = parents, C = children, F = females, M = males.

Main effects and interactions	NC listeners			WI listeners		
	/i, ɪ, e, ɛ/	/æ/	/ɑ, ɔ, aɪ/	/i, ɪ, e, ɛ/	/æ/	/ɑ, ɔ, aɪ/
Dialect	0.528[a] NC > WI	0.617[a] NC > WI	—	—	—	0.501[b] WI > NC
Generation	0.299[b] P > C > GP	0.542[a] C > P > GP	0.471[a] GP > P > C	—	0.444[a] P > GP > C	—
Gender	—	—	—	—	—	0.650[a] F > M
Dialect × Generation	0.658[a]	—	0.332[b]	0.538[a]	0.378[b]	—
Dialect × Gender	—	—	—	—	0.551[a]	—
Generation × Gender	0.418[a]	—	—	—	—	—
Dialect × Gen × Gender	—	—	—	0.369[b]		—

[a] $p < 0.001$.
[b] $p < 0.010$.

For the three back vowels, all three age groups differed significantly one from another. The higher IDRs for GP were no doubt related to the upgliding /ɔ/ in NC and a relatively good spectral separation between /ɔ/ and /ɑ/ in WI. The lower IDRs for P- and C-groups may also be related to an ongoing sound change in the back vowel corner, which increased confusions of WI /ɑ/ with the monophthongal /aɪ/ as already discussed. The remaining significant interactions listed in Table V do not advance our understanding of the above patterns and therefore are not discussed at present.

2. Identification results for WI listeners

Identification results for WI listeners are shown in Table VI. Several patterns indicate sensitivity of WI listeners to formant dynamics. First, the IDRs for /i/ in NC GP speakers were low and the mislabeling with /e/ suggests that the greater spectral change in GP variants of /i/ was a likely cause of the confusions. In contrast, their responses to the monophthongal WI /i/ were near ceiling. Second, WI listeners confused the monophthongal NC /aɪ/ with /ɑ/ (and sometimes with /æ/) and did so more often than NC listeners. The IDRs then increased for NC girls in response to their more diphthongal /aɪ/. The full WI diphthong yielded mostly ceiling effects. This pattern of responses to the variation in the formant movement in /aɪ/ is thus consistent with NC listeners.

Third, the monophthongal WI /e/ caused both lower IDRs and confusions mostly with /ɛ/ (when produced by males) and /i/ (when produced by females), which is again consistent with NC listeners. The responses to WI C may provide some hints as to the source of this gender-related variation. In particular, listeners may have been guided by higher fundamental frequency in females by choosing the /i/-response more often. The fact that the confusions for WI C were mostly with /i/ and /ɪ/ would suggest that higher fundamental frequencies in children (both in boys and girls) reduced the likelihood of confusing /e/ with /ɛ/. This possibility needs to be verified in separate experiments, however. In terms of the sensitivity to formant movement, the responses of WI listeners to NC /e/ indicate that the greater amount of spectral change in the NC variant produced higher IDRs, as it was also the case for NC listeners.

Finally, the lower IDRs for NC (and to some extent, WI) /ɪ, ɛ/ and their mutual confusions are difficult to explain with reference to formant dynamics only. As already pointed out, the variable IDRs as a function of both generation and dialect most likely reflect the complexity of the spectral neighborhood in this vowel region. This will include at least partial acoustic overlap of several front vowels as well as spectro-temporal variations such as spectral rate of change in individual vowel categories. A more focused investigation is needed to better understand the relative salience of acoustic cues in this spectral region.

The results of the analogous ANOVAs for WI listeners showed that speaker dialect was a significant factor only for /ɑ, ɔ, aɪ/ and, for these vowels, there was a native dialect advantage (see Table V). There was also a significant main effect of gender for these vowels, revealing higher IDRs for female variants. The main effect of generation was significant only for /æ/, revealing significantly higher IDRs for P compared to either C or GP, which also differed significantly one from another. These pattern is not easily understood given that the identification of this vowel was generally high across all age groups (above 80%), suggesting that WI listeners had no major difficulties identifying the vowel in both dialects. However, a significant dialect by generation interaction arose due to the fact that the IDRs were lowest for NC C and for WI GP. While no obvious spectral cues can be detected in the plots for the later group, the reduction of formant movement in NC C could be a reason for greater confusions of /æ/ with /ɛ/ and /ɑ/. It is also noteworthy that IDRs were higher for NC female /æ/ but no gender-related differences were found for WI, which resulted in a significant dialect by gender interaction.

Although the main effect of dialect was not significant for /i, ɪ, e, ɛ/, a significant dialect by generation interaction revealed that for the children's productions, the IDRs were lower for NC C and higher for WI C, whereas there were no differences in the adult groups. These results illustrate that native dialect advantage may not be manifested uniformly across even selected vowel categories, but may show up under specific conditions such as generational differences in vowel production.

C. Discriminant analyses

The third type of analysis was linear discriminant analysis (DA) of the acoustic measurements. The purpose of using

TABLE VI. Identification rates and confusion matrix for Wisconsin listeners.

Speaker gender	Vowel intended by speaker	Vowel identified by listeners											
		/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/ɝ/	/aɪ/	/o/	/ʊ/	/u/
(a) Wisconsin listeners: Responses to Wisconsin GP speakers													
Male	/i/	**92.0**	4.0	—	4.0	—	—	—	—	—	—	—	—
	/ɪ/	1.3	**73.3**	—	20.0	—	—	—	1.3	1.3	—	2.7	—
	/e/	—	13.3	**54.7**	32.0	—	—	—	—	—	—	—	—
	/ɛ/	—	2.7	1.3	**93.3**	—	—	—	—	—	—	2.7	—
	/æ/	—	—	—	17.3	**82.7**	—	—	—	—	—	—	—
	/ɑ/	—	—	—	1.3	4.0	**82.7**	5.3	—	6.7	—	—	—
	/ɔ/	—	—	—	—	1.3	69.3	**29.3**	—	—	—	—	—
	/ɝ/	—	—	—	—	—	—	—	**100.0**	—	—	—	—
	/aɪ/	—	1.3	12.0	1.3	—	—	—	—	**85.3**	—	—	—
	/o/	—	—	—	—	—	1.3	1.3	—	—	**97.3**	—	—
	/ʊ/	—	—	—	—	1.3	—	—	—	—	—	**98.7**	—
	/u/	1.3	—	—	—	—	—	—	—	—	2.7	21.3	**74.7**
Female	/i/	**92.0**	2.7	—	1.3	—	—	—	—	—	1.3	—	1.3
	/ɪ/	—	**97.3**	—	1.3	—	—	—	—	—	—	1.3	—
	/e/	17.3	8.0	**58.7**	13.3	—	—	—	—	1.3	1.3	—	—
	/ɛ/	—	2.7	—	**94.7**	1.3	1.3	—	—	—	—	—	—
	/æ/	1.3	—	—	5.3	**90.7**	1.3	1.3	—	—	—	—	—
	/ɑ/	—	—	1.3	2.7	1.3	**78.7**	5.3	—	10.7	—	—	—
	/ɔ/	—	—	—	—	—	49.3	**48.0**	—	—	—	2.7	—
	/ɝ/	—	—	—	—	—	—	—	**100.0**	—	—	—	—
	/aɪ/	—	—	2.7	—	1.3	—	—	—	**96.0**	—	—	—
	/o/	—	—	—	—	—	—	—	—	—	**97.3**	—	2.7
	/ʊ/	1.3	—	—	—	1.3	—	—	—	—	—	**97.3**	—
	/u/	—	1.3	—	—	—	—	—	1.3	—	—	4.0	**93.3**
(b) Wisconsin listeners: Responses to Wisconsin P speakers													
Male	/i/	**94.7**	1.3	—	—	—	—	1.3	—	1.3	—	—	1.3
	/ɪ/	—	**70.7**	—	29.3	—	—	—	—	—	—	—	—
	/e/	6.7	5.3	**50.7**	33.3	—	—	1.3	1.3	—	—	—	1.3
	/ɛ/	—	—	—	**81.3**	14.7	1.3	—	1.3	—	1.3	—	—
	/æ/	—	—	1.3	4.0	**93.3**	1.3	—	—	—	—	—	—
	/ɑ/	—	1.3	—	1.3	8.0	**73.3**	1.3	—	14.7	—	—	—
	/ɔ/	—	—	—	—	1.3	76.0	**22.7**	—	—	—	—	—
	/ɝ/	—	—	1.3	—	—	1.3	—	**97.3**	—	—	—	—
	/aɪ/	—	—	2.7	—	—	—	—	—	**96.0**	1.3	—	—
	/o/	1.3	—	—	—	—	1.3	—	—	—	**96.0**	—	1.3
	/ʊ/	—	1.3	—	—	—	—	—	—	—	1.3	**97.3**	—
	/u/	1.3	—	—	—	1.3	—	—	—	—	2.7	10.7	**85.3**
Female	/i/	**96.0**	4.0	—	—	—	—	—	—	—	—	—	—
	/ɪ/	1.3	**90.7**	—	2.7	—	—	—	1.3	2.7	—	1.3	—
	/e/	8.0	8.0	**66.7**	14.7	—	—	—	—	—	1.3	—	1.3
	/ɛ/	—	—	—	**84.0**	14.7	—	—	1.3	—	—	—	—
	/æ/	—	—	—	2.7	**93.3**	2.7	—	—	—	1.3	—	—
	/ɑ/	—	—	—	—	5.3	**64.0**	4.0	—	26.7	—	—	—
	/ɔ/	—	—	1.3	—	—	46.7	**50.7**	—	—	—	—	1.3
	/ɝ/	—	—	—	1.3	—	—	—	**98.7**	—	—	—	—
	/aɪ/	—	—	—	—	—	1.3	1.3	—	**97.3**	—	—	—
	/o/	1.3	—	—	1.3	1.3	—	—	—	—	**94.7**	—	1.3
	/ʊ/	—	1.3	—	—	—	4.0	—	1.3	—	1.3	**92.0**	—
	/u/	—	—	—	1.3	—	2.7	—	—	—	1.3	1.3	**93.3**
(c) Wisconsin listeners: Responses to Wisconsin C speakers													
Male	/i/	**98.7**	1.3	—	—	—	—	—	—	—	—	—	—
	/ɪ/	—	**64.0**	—	25.3	—	—	—	—	1.3	—	9.3	—
	/e/	8.0	12.0	**68.0**	9.3	—	—	—	—	1.3	—	—	1.3
	/ɛ/	—	4.0	—	**85.3**	8.0	—	—	—	—	1.3	1.3	—
	/æ/	1.3	—	—	13.3	**80.0**	4.0	—	—	—	—	1.3	—
	/ɑ/	1.3	—	1.3	1.3	—	**82.7**	4.0	—	9.3	—	—	—
	/ɔ/	—	—	2.7	—	1.3	69.3	**24.0**	—	2.7	—	—	—

TABLE VI. (*Continued*)

Speaker gender	Vowel intended by speaker	Vowel identified by listeners											
		/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/ɝ/	/aɪ/	/o/	/ʊ/	/u/
	/ɝ/	—	—	—	—	—	1.3	—	**97.3**	—	—	—	1.3
	/aɪ/	—	—	2.7	—	1.3	—	—	—	**96.0**	—	—	—
	/o/	—	—	—	—	—	1.3	—	—	—	**98.7**	—	—
	/ʊ/	—	—	—	—	—	—	—	—	—	—	**98.7**	1.3
	/u/	—	—	—	—	—	—	—	1.3	—	—	1.3	**97.3**
Female	/i/	**97.3**	—	—	—	—	—	1.3	—	—	—	—	1.3
	/ɪ/	—	**70.7**	—	29.3	—	—	—	—	—	—	—	—
	/e/	12.0	4.0	**70.7**	9.3	—	—	1.3	—	1.3	—	1.3	—
	/ɛ/	—	—	—	**93.3**	6.7	—	—	—	—	—	—	—
	/æ/	—	—	—	2.7	**96.0**	—	—	—	1.3	—	—	—
	/ɑ/	—	—	—	—	8.0	**68.0**	—	—	24.0	—	—	—
	/ɔ/	—	—	—	1.3	—	50.7	**44.0**	2.7	—	1.3	—	—
	/ɝ/	—	—	—	—	—	—	—	**100.0**	—	—	—	—
	/aɪ/	—	—	—	—	—	—	—	—	**100.0**	—	—	—
	/o/	—	—	—	—	—	—	—	—	—	**98.7**	—	1.3
	/ʊ/	—	—	—	—	—	2.7	—	—	—	—	**97.3**	—
	/u/	—	—	—	—	—	—	—	—	—	1.3	1.3	**97.3**
(d) Wisconsin listeners: Responses to North Carolina GP speakers													
Male	/i/	**61.3**	—	33.3	4.0	—	—	1.3	—	—	—	—	—
	/ɪ/	—	**68.0**	1.3	28.0	—	—	—	—	1.3	1.3	—	—
	/e/	1.3	—	**90.7**	5.3	—	—	—	—	1.3	—	1.3	—
	/ɛ/	—	6.7	6.7	**73.3**	10.7	1.3	—	—	—	—	1.3	—
	/æ/	—	—	—	4.0	**94.7**	1.3	—	—	—	—	—	—
	/ɑ/	—	—	—	—	1.3	**73.3**	25.3	—	—	—	—	—
	/ɔ/	—	—	—	1.3	2.7	42.7	**48.0**	—	—	4.0	—	**1.3**
	/ɝ/	—	1.3	—	—	—	—	—	**97.3**	—	1.3	—	—
	/aɪ/	—	2.7	—	1.3	25.3	28.0	2.7	1.3	**38.7**	—	—	—
	/o/	—	—	—	—	—	1.3	—	—	—	**96.0**	—	2.7
	/ʊ/	—	—	—	—	—	—	—	—	—	2.7	**96.0**	1.3
	/u/	—	—	—	2.7	—	—	—	—	—	16.0	6.7	**74.7**
Female	/i/	**84.0**	6.7	2.7	2.7	—	—	—	—	2.7	1.3	—	—
	/ɪ/	—	**60.0**	1.3	34.7	—	—	—	—	4.0	—	—	—
	/e/	5.3	1.3	**78.7**	2.7	—	—	1.3	1.3	5.3	—	2.7	1.3
	/ɛ/	—	4.0	5.3	**84.0**	4.0	—	—	1.3	—	1.3	—	—
	/æ/	—	—	—	—	**97.3**	1.3	—	—	—	—	1.3	—
	/ɑ/	—	—	—	—	2.7	**81.3**	13.3	—	1.3	—	1.3	—
	/ɔ/	—	—	—	—	1.3	20.0	**73.3**	—	—	5.3	—	—
	/ɝ/	—	—	—	—	—	1.3	—	**96.0**	—	2.7	—	—
	/aɪ/	—	—	—	1.3	36.0	22.7	—	—	**40.0**	—	—	—
	/o/	—	—	—	1.3	—	1.3	—	—	—	**94.7**	—	2.7
	/ʊ/	—	—	—	—	—	—	—	—	—	—	**98.7**	1.3
	/u/	1.3	—	—	—	—	—	—	—	—	28.0	8.0	**62.7**
(e) Wisconsin listeners: Responses to North Carolina P speakers													
Male	/i/	**90.7**	—	6.7	1.3	—	—	—	—	1.3	—	—	—
	/ɪ/	—	**92.0**	—	6.7	—	—	—	—	—	—	—	1.3
	/e/	—	—	**88.0**	5.3	—	—	—	—	6.7	—	—	—
	/ɛ/	—	13.3	2.7	**76.0**	5.3	1.3	—	1.3	—	—	—	—
	/æ/	—	—	—	4.0	**96.0**	—	—	—	—	—	—	—
	/ɑ/	1.3	—	—	—	4.0	**64.0**	29.3	—	—	—	—	1.3
	/ɔ/	—	—	—	1.3	—	21.3	**69.3**	—	—	5.3	—	2.7
	/ɝ/	—	—	—	—	—	—	—	**100.0**	—	—	—	—
	/aɪ/	—	2.7	—	—	—	56.0	4.0	—	**37.3**	—	—	—
	/o/	—	—	—	1.3	2.7	—	—	—	—	**94.7**	—	1.3
	/ʊ/	1.3	—	—	—	—	—	—	—	—	1.3	**96.0**	1.3
	/u/	—	—	—	1.3	—	—	—	—	—	4.0	8.0	**85.3**
Female	/i/	**96.0**	1.3	—	1.3	—	—	1.3	—	—	—	—	—
	/ɪ/	1.3	**62.7**	4.0	28.0	1.3	—	—	—	1.3	—	1.3	—
	/e/	1.3	—	**88.0**	4.0	1.3	1.3	—	—	1.3	1.3	—	1.3

TABLE VI. (*Continued*)

Speaker gender	Vowel intended by speaker	Vowel identified by listeners											
		/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/ɝ/	/aɪ/	/o/	/ʊ/	/u/
	/ɛ/	—	4.0	5.3	**73.3**	16.0	—	1.3	—	—	—	—	—
	/æ/	—	—	—	1.3	**98.7**	—	—	—	—	—	—	—
	/ɑ/	—	—	2.7	—	1.3	**73.3**	21.3	—	1.3	—	—	—
	/ɔ/	—	—	—	—	1.3	16.0	**77.3**	1.3	—	2.7	—	1.3
	/ɝ/	—	—	—	—	—	—	—	**100.0**	—	—	—	—
	/aɪ/	—	—	—	1.3	6.7	76.0	1.3	—	**13.3**	1.3	—	—
	/o/	—	1.3	—	1.3	—	1.3	—	—	—	**94.7**	—	1.3
	/ʊ/	—	—	—	—	—	—	—	—	—	—	**96.0**	4.0
	/u/	—	—	—	—	—	1.3	—	1.3	—	4.0	—	**93.3**
(f) Wisconsin listeners: Responses to North Carolina C speakers													
Male	/i/	**84.0**	1.3	8.0	2.7	1.3	1.3	—	1.3	—	—	—	—
	/ɪ/	—	**88.0**	2.7	—	—	—	—	1.3	2.7	1.3	4.0	—
	/e/	1.3	—	**74.7**	6.7	—	—	1.3	1.3	1.3	10.7	2.7	—
	/ɛ/	—	20.0	—	**70.7**	1.3	—	—	—	—	—	8.0	—
	/æ/	—	—	1.3	6.7	**89.3**	1.3	—	—	1.3	—	—	—
	/ɑ/	—	—	1.3	—	—	**57.3**	40.0	—	1.3	—	—	—
	/ɔ/	—	—	—	—	—	22.7	**50.7**	1.3	—	24.0	—	1.3
	/ɝ/	—	—	—	—	—	—	—	**97.3**	—	1.3	1.3	—
	/aɪ/	—	1.3	—	1.3	2.7	48.0	5.3	2.7	**38.7**	—	—	—
	/o/	—	—	—	—	—	—	—	2.7	—	**96.0**	—	1.3
	/ʊ/	—	—	—	—	—	—	—	—	—	—	**93.3**	6.7
	/u/	—	—	—	—	—	—	—	1.3	—	6.7	—	**92.0**
Female	/i/	**86.7**	—	8.0	4.0	—	—	—	—	—	1.3	—	—
	/ɪ/	—	**70.7**	1.3	25.3	1.3	—	—	—	1.3	—	—	—
	/e/	—	1.3	**90.7**	1.3	—	—	—	—	5.3	—	—	1.3
	/ɛ/	1.3	—	5.3	**89.3**	4.0	—	—	—	—	—	—	—
	/æ/	—	—	—	9.3	**82.7**	5.3	—	—	2.7	—	—	—
	/ɑ/	—	—	1.3	1.3	1.3	**54.7**	34.7	—	6.7	—	—	—
	/ɔ/	—	—	1.3	—	1.3	46.7	**46.7**	1.3	—	1.3	—	1.3
	/ɝ/	—	—	1.3	—	—	—	—	**97.3**	—	—	1.3	—
	/aɪ/	—	—	—	—	13.3	8.0	—	1.3	**76.0**	1.3	—	—
	/o/	—	—	—	1.3	—	—	—	—	—	**97.3**	—	1.3
	/ʊ/	—	—	—	—	—	1.3	—	2.7	—	—	**96.0**	—
	/u/	—	—	—	—	—	—	—	1.3	—	4.0	2.7	**92.0**

DA in the present study was to determine how well the combinations of acoustic measurements would separate vowels in the stimulus tokens and how effective were the "static" versus "dynamic" cues in separating dialects and generations. Given the five-point sampling used here, we analyzed the statistical variation in acoustic measurements to gain insights into the optimal number of sample points to effectively classify the vowels.

The identification results informed us that listeners were sensitive to dynamic changes, which could have guided at least some of their labeling choices. Given this, we now examine how effective is dynamic information in classification of vowels in the stimulus set. A series of linear DAs were conducted to determine whether there is an improvement in overall accuracy when formant frequencies are sampled at multiple time points as opposed to a single measurement point in a vowel.

1. Dialect-specific overall classification accuracy

In the first analysis, we examined classification accuracy for all twelve vowels as a function of the number and location of sample time points in the vowel: one (the 50% point), two (20 and 80%), three (20, 50, and 80%) and all five points (20, 35, 50, 65, and 80%). The analyses were performed separately for each dialect, generation and gender group using the cross-validation (jack-knife) approach (e.g., Fox and Nissen, 2005; Nissen and Fox, 2005). In this procedure (see Klecka, 1980), the measurements of a given vowel token are excluded from the training data prior to the classification of that token. The parameter set included F1, F2, and F3 measures and vowel duration. The formant frequency values were not normalized because no considerable differences in the vocal tract length were expected within each age and gender group. Table VII shows correct classification rates for formant values at one, two and three time points; the rates for all five points did not show further improvement and are not included in the table.

For NC speakers, the mean accuracy for one sample point was 62.5%. A two-sample model resulted in a dramatic increase of classification rates by 23.4 percentage points but including values from all three sample points did not change the overall classification. Inclusion of duration did not

TABLE VII. Dialect-specific classification accuracy for 12 vowels using the cross-validation approach. The one-sample results are based on F1, F2, and F3 measurements (using values in Hz) at 50% point in a vowel (with and without duration measure), two-sample results on formant pattern sampled at 20 and 80 % points and three-sample results on formant pattern sampled at 20, 50, and 80 % points. Correct classification results (%) are shown for each generation of North Carolina (NC) and Wisconsin (WI) speakers: Grandparents (GP), parents (P), and children (C).

Speaker group	One sample		Two samples		Three samples	
	No Dur	Dur	No Dur	Dur	No Dur	Dur
NC GP males	70.0	68.3	91.7	90.0	88.3	88.3
NC GP females	56.7	46.7	85.0	80.0	80.0	76.7
NC P males	70.0	70.0	90.0	88.3	93.3	93.3
NC P females	63.3	65.0	86.7	85.0	86.7	88.3
NC C males	53.3	55.0	81.7	80.0	85.0	83.3
NC C females	61.7	53.3	80.0	80.0	83.3	83.3
Mean NC	62.5	59.7	85.9	83.9	86.1	85.5
WI GP males	70.0	70.0	81.7	86.7	88.3	86.7
WI GP females	66.7	76.7	86.7	93.3	86.7	88.3
WI P males	78.3	81.7	93.3	95.0	96.7	96.7
WI P females	76.7	80.0	91.7	93.3	91.7	90.0
WI C males	58.3	66.7	71.7	73.3	78.3	76.7
WI C females	60.0	68.3	86.7	88.3	85.5	90.0
Mean WI	68.3	73.9	85.3	88.3	87.9	88.1

improve classification accuracy and even worsened it slightly. This was particularly the case for GP females whose classification rates were the lowest in all NC groups (averaging 70.9% for all measures). While the classification accuracy for NC GP females increased by 28.3 percentage points with two samples, adding duration to one sample decreased it. This indicates that the long durations of NC GP female vowels (which were the longest in the present set, averaging 322 ms) did not benefit the classification rate and spectral cues alone contributed to the improvement.

In contrast, including duration did improve classification accuracy for most of WI speaker groups, particularly for one- and two-sample models. The largest improvement was for WI GP females. As for NC speakers, two samples resulted in a large increase in overall classification accuracy, averaging 17 percentage points. Addition of the third sample point benefitted only selected speaker groups and the improvement was comparatively smaller.

In summary, the first analysis established that considering each dialect separately, a model including two sample points can effectively predict overall accuracy in statistical classification of vowels. However, the results also revealed dialect-specific use of duration cues so that a considerable improvement (such as in WI) or a considerable decrease in accuracy (such as in NC) can be obtained by adding duration to a single measurement point.

2. Cross-generational classification

The second set of DA examined the degree to which the vowels can be correctly classified on the basis of acoustic characteristics of a specific generation of speakers. Utilizing the specific group approach as in Nissen and Fox (2005), the overall data set was split into subsets. Because the perception data gave us some indication that listeners tend to rate higher the exemplars from speakers of the same generation, we selected the P group as the subset employed to train the discriminant functions. Specifically, discriminant functions were trained on only NC P speakers and WI P speakers and were then applied to the remaining generation groups (ignoring speaker gender): NC GP, WI GP, NC children, and WI children. Predictor variables in both types of analysis included F1, F2, F3 for one, two, and three sample points and vowel duration. To reduce the variation across age groups and gender due to the differences in vocal tract length, we used normalized formant frequency values (Lobanov, 1971). This analysis should give us indication of any changes in classification accuracy as a function of speaker generation within and across each dialect.

Classification results are shown in Table VIII. The discriminant model developed on NC P adult exemplars resulted in a mean classification rate of 88.9% for the data in the training set. The accuracy for one sample was 76.7 and 96.7 % for the two-sample model, an increase of 20 percentage points. Addition of a third sample or duration to either model minimally decreased the accuracy. When applied to productions by NC GP, the mean correct classification rates were slightly lower (86%) although the general pattern of improvement as well as a decrease in accuracy as a function of the number of samples and duration was maintained. When applied to NC children, the mean classification accuracy was lower (77.1%), showing again a two-sample model advantage and a slight improvement when duration was included in the one-sample model.

Altogether, these classification results show only small differences between the two NC adult groups. However, the lower classification accuracy for children suggests a significant change in their vowel system (rather than potential differences related to vocal tract length). We will return to this below. When applied to productions by WI speakers, the model trained on NC P adults categorized WI vowels with much lower accuracy, indicating large dialectal differences in vowel production. This finding corresponds to listeners' responses in the present study which, on several occasions, reflected native dialect advantage. In general, there was no evidence of cross-generational variation in classification accuracy for WI speakers. Within each WI group, the rate of successful classification improved most when two samples were included in the model relative to the one-sample model. However, adding duration to one sample also resulted in a considerable improvement, as shown in Table VIII.

The second classification model was developed on WI P adults. For the data in this training set, the mean correct classification rate was 93.9%, slightly higher compared to NC P adults' model. Consistent with previous findings, there was a large increase in accuracy for the two-sample model relative to the one-sample model and neither the three-sample model nor addition of the duration to any of the models resulted in major improvement. Applying the WI P model to WI GP adults, the classification accuracy worsened, indicating differences in vowel characteristics between the two adult groups. When applied to WI children, the mean accuracy was slightly higher than for WI GP adults. There was again a

TABLE VIII. Results of discriminant classification using both cross-validation and "specific group" procedures. Original training sets (in boldface) included 12 vowels of parents (P) and children (C) generations in North Carolina (NC) and Wisconsin (WI) dialect. Each training set was applied to a selected generation group as listed in the table. All values represent % correct classification. The results are shown for one, two and three sample points in a vowel using normalized formant frequency values.

Training set (cross-validation)	Data set evaluated (selected group)	One sample		Two samples		Three samples		Cross-analysis mean
		No Dur	Dur	No Dur	Dur	No Dur	Dur	
NC P adults		**76.7**	**73.3**	**96.7**	**95.8**	**95.8**	**95.0**	**88.9**
	NC GP adults	74.2	73.3	91.7	90.8	93.3	92.5	86.0
	NC children	63.3	66.7	84.2	85.0	82.5	80.8	77.1
	WI GP adults	35.8	43.3	48.3	50.0	50.8	51.7	46.7
	WI P adults	36.7	41.7	45.0	46.7	49.2	49.2	44.8
	WI children	37.5	44.2	47.5	50.0	47.5	49.2	46.0
WI P adults		**83.3**	**86.7**	**97.5**	**98.3**	**98.3**	**99.2**	**93.9**
	WI GP adults	75.0	76.7	86.7	87.5	86.7	88.3	83.5
	WI children	75.0	77.5	91.7	90.8	92.5	91.7	86.5
	NC GP adults	41.7	45.8	50.8	50.8	51.7	52.5	48.9
	NC P adults	41.7	46.7	47.5	50.0	52.5	52.5	48.5
	NC children	44.2	49.2	55.8	53.3	55.8	55.0	52.2
NC children		**70.8**	**71.7**	**90.0**	**89.2**	**90.0**	**89.2**	**83.5**
	WI children	55.0	59.2	55.8	59.2	59.2	60.8	58.2
WI children		**83.3**	**85.0**	**91.7**	**92.5**	**94.2**	**93.3**	**90.0**
	NC children	57.5	56.7	65.0	64.2	62.5	63.3	61.5

large two-sample model advantage and only negligible changes for all three sample points or inclusion of the duration. When applied to NC productions, the WI P model classified the vowels with much lower accuracy, which was consistent with the results of the model trained on NC P speakers and applied to WI data. Again, these results underscore large dialectal differences in vowel production. Another consistent finding was that the rate of successful classification improved most when two sample points were included in the analysis and, to a lesser extent, when duration was added to one sample.

Two more models were developed for the specific group approach, one for NC children and one for WI children. These two models were then applied interchangeably to the productions of both children groups. Because the model trained on NC P adults resulted in much lower classification rates when applied to NC children, we interpreted this outcome as an indication of a change in the productions of NC children related to a partial loss of Southern features. We therefore reasoned that a model trained on NC children should produce comparatively higher classification rates when applied to WI children. Similarly, a model trained on WI children should produce higher rates when applied to NC children because of the smaller differences between the two child systems compared with adults in each dialect.

The results supported these predictions. When NC children model was applied to WI children, the mean classification accuracy increased by 12.2 percentage points (58.2%) compared to the results from the model trained on NC P adults (46.0%). Applying WI children model to NC children's productions, mean accuracy was again higher (61.5%) compared to the model trained on WI P adults (52.2%). These higher classification rates indicate that the dialect differences between the spectral characteristics of vowels in the two child systems were reduced. It was also the case that the two-sample model resulted in a large improvement in classification accuracy in the NC children training set but the corresponding improvement in the WI set was smaller. A likely explanation of this discrepancy is that the classification rate was already higher for one sample point in the WI set (83.3%) compared to the NC set (70.8%) and the two-sample model provided comparatively more information about spectral characteristics of NC vowels.

3. Dialect classification

The third set of analyses examined dialect classification. The accuracy of the classification was considered as the percentage of cases that were correctly classified into NC and WI groups. Of interest was whether dialect classification accuracy could be improved with addition of more information about vowel characteristics (i.e., in two- and three-sample models and duration). The cross-validation procedure was used utilizing the normalized formant frequency values (Lobanov, 1971).

The results for the entire set of 12 vowels shown in Table IX indicate that one sample point provides very limited dialect information in order to correctly classify vowels

TABLE IX. Classification accuracy for dialect using the cross-validation approach. One, two, and three sample-point analyses were performed on normalized formant frequency values for the entire set of 12 vowels and selected subsets. All values represent % correct dialect classification.

	One sample		Two samples		Three samples	
Vowels	No Dur	Dur	No Dur	Dur	No Dur	Dur
Entire set	53.6	62.6	59.0	65.0	64.4	66.8
/i, ɪ, e, ɛ, æ/	62.6	66.7	52.3	68.0	70.0	73.0
/ɑ, ɔ, aɪ/	51.1	66.7	65.5	67.8	62.2	70.0

(53.6%). However, adding vowel duration improves the classification (62.6%), as does the utilization of two sample points (59%). Adding duration to two samples results in further improvement (65%) but using all three points without duration does not (64.4%). A modest further improvement was found for three sample points with duration (66.8%). These results indicate that, for the entire set of vowels, the model including two sample points and duration provides reasonable classification accuracy and the addition of further measurement points does not improve it considerably.

We also examined dialect classification rates for the subsets of vowels which were most often confused with one another in the listening task, i.e., the front and low back vowels. For the front vowel set /i, ɪ, e, ɛ, æ/, one sample point provided comparatively more information about dialect than was found for the entire set and adding duration resulted in further modest improvement (see Table IX). This indicates that information about "static" positional vowel characteristics along with duration differences between NC and WI vowels provide relatively strong cues to dialect classification for the front vowels set. A surprising result was that two sample points lowered accuracy rather than improving it (52.3%) but inclusion of duration increased it (68%). Adding the third sample to two samples without duration resulted in a dramatic improvement by 17.7 percentage points but the improvement was very modest when duration was included in the model. The accuracy was highest for the three-sample model with duration (73%). These results suggest that two sample points most likely provided ambiguous cues because the NC variants of /ɪ, ɛ, æ/ had comparatively more complex formant patterns and addition of the third sample representing the midpoint resolved this ambiguity. Also, addition of duration increased accuracy when information in two samples was insufficient to improve dialect classification.

The classification accuracy for the low back vowels /ɑ, ɔ, aɪ/ was poor using one sample (51.1%) but increased by 15.6 percentage points with the addition of duration. This indicates that the vowel midpoint alone (representing static information) does not provide sufficient spectral cues to separate the dialectal overlap and cues in vowel duration improve dialect classification. Two sample points provided more cues about dialectal differences in diphthongal changes such as in WI exemplars of /aɪ/ and NC /ɔ/, improving classification rates by 14.4 percentage points. Using the three-sample model without duration did not result in further improvement but duration cues again contributed to higher classification rates, reaching 70%.

In summary, for the range of vowels examined here, the models using two and (or) three sample points improved dialect classification accuracy compared to a one-sample model. Duration cues also improved dialect classification, especially when included with one sample for back vowels and with two samples for front vowels.

IV. GENERAL DISCUSSION

As stated at the outset, combined evidence from acoustic measurements, statistical pattern recognition and listeners' identification responses supports the position that dynamic vowel structure enhances information about vowel identity, particularly in relation to nominal monophthongs. The present study draws attention to sociophonetic variation as a source of additional cues to vowel recognition and provides further evidence that the importance of dynamic structure is not evenly distributed across the vowel system and that it interacts with sociolinguistic variables.

The acoustic measurements of the stimulus tokens used in the study revealed substantial differences in the dynamic formant patterns of the vowels as a function of speaker dialect and generation. We then examined listeners' sensitivity to this variation by presenting them with these exemplars, which were produced by speakers of two dialects, three generations and both genders. Listeners showed a native dialect advantage identifying vowels of their own dialect with higher accuracy than the vowels from another dialect. Overall accuracy was also significantly higher when they responded to vowels of their own generation (compared to the older and younger age groups) and to vowels produced by females. Finally, a series of discriminant analyses explored the accuracy of dialect and generation classification and showed that, overall, the models which included two sample points and duration outperformed the models that sampled formant frequency only once, at the vowel midpoint. Our discussion will now consider the importance of these findings to research on vowel recognition.

The listening results brought to light great variability in identification of individual vowel categories with most of the low identification rates due to dialectal differences. As pointed out earlier in the paper, such low identifications of individual vowels were not surprising because highly variable results were already reported in a dialect-controlled experiment (Labov, 2010), emphasizing that dialect can be a strong source of confusions and mislabeling of vowels. The present results further demonstrate that dialect differences are also manifested in the dynamic vowel structure and listeners are sensitive to this variation.

Formant dynamics can provide identification cues which are particularly important to accurate perceptual categorization of vowels positioned in a close spectral proximity. To illustrate this point, Fig. 3 displays normalized formant midpoints (redrawn from Figs. 1 and 2) for NC and WI variants of /ɑ, ɔ, aɪ/ produced in the low back corner of the vowel space. The means are displayed for three age groups which results in three data points for each vowel category. Presented with a mixture of spectrally overlapping exemplars from two dialects, both NC and WI listeners mostly confused the vowels one for another. However, those exemplars that displayed a greater amount of spectral change likely provided less ambiguous cues and their identification rates were comparatively higher. This was true for NC /ɔ/ produced with an upgliding diphthongal change and a diphthongized variant of /aɪ/ in NC girls (a vowel which was produced by other NC speakers as a relative monophthong). Given that the fully diphthongal variant of WI /aɪ/ (not shown here) caused almost no confusions for the two listener groups, we understand these results as evidence that listeners utilized dynamic cues to disambiguate vowel signal. Several other

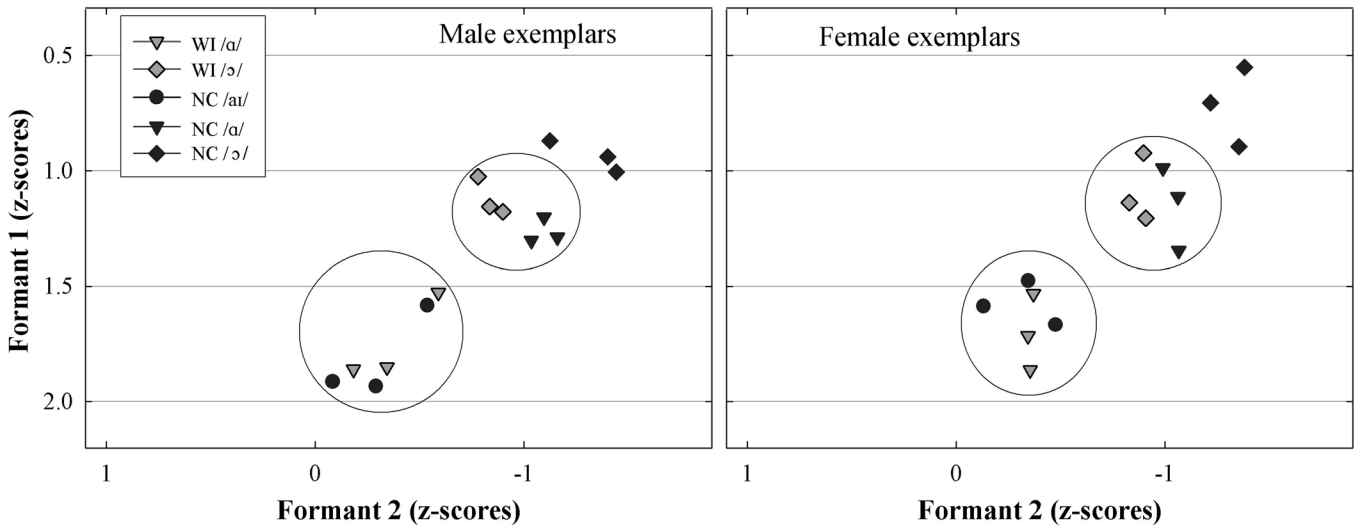

FIG. 3. Mean normalized formant frequencies measured at vowel midpoint only for cross-generational variants of North Carolina /ɑ, ɔ, aɪ/ and Wisconsin /ɑ, ɔ/. Data represent a subset of points shown in Figs. 1 and 2.

examples discussed earlier (see Sec. III B) provide further evidence for listeners' use of variation in VISC in making their identification choices.

Much of the earlier discussions in this paper focused on the front vowels /i, ɪ, e, ɛ, æ/ because their spectral positions and amount of VISC were affected not only by dialect but also by generational differences. Listeners were sensitive to these variations as they created a substantial number of confusions. In some cases, the variation in VISC actually contributed to these confusions (rather than disambiguated the signal as for the back vowels). This was the case for NC GP /i/ which, being diphthongized (i.e., produced with Southern breaking), was more often confused with the spectral neighbors. On the other hand, this example provides further evidence that listeners were indeed sensitive to the spectral changes in front vowels.

The crowded spectral neighborhood of the front vowel group and the considerable spectral overlap caused by dialect and generational differences call for further examination of these variable patterns on vowel identification. More focused perception experiments are needed to better understand the effectiveness of spectral cues in differentiating among spectral neighbors in this region of the vowel space. It could be the case that ambiguities (other than created by dialect and generational variations) can only be resolved at higher organizational levels of linguistic structure or at least by presenting the listener with lexical items other than minimally contrastive /hVd/-tokens. This possibility will also need to be verified.

However, the study also found that differences in the acoustic formant patterns did not always produce perceptual confusions. This was the case for three back vowels which were identified with high accuracy despite great variability in their productions. In particular, the acoustic patterns of /o/ in Figs. 1 and 2 were highly variable both in terms of formant frequency values and in the amount of spectral change. But these dynamic variations had negligible effects on vowel identity and the tokens were labeled as intended by speakers 96.5% of the time. Similar results were for /ʊ/, although the acoustic differences were primarily in greater spectral changes in WI compared to NC variants rather than in positional differences in the vowel space. Again, the vowel was labeled as /ʊ/ 96.9% of the time.

The responses to /u/ provide us with yet another puzzle. Given the great positional differences (very fronted NC /u/ versus far back WI /u/), the pattern of confusions for the NC GP (both genders) and WI GP males runs counter to what could be expected on the basis of the proximity to spectral neighbors. That is, both listener groups consistently labeled some of the NC exemplars as /o/ and the WI exemplars as /ʊ/ although their acoustic proximities would suggest the reverse. The confusions were sparse in younger generations (overall IDR for P and C was 94.1%) but again, neither the positional differences nor differences in the amount and direction of spectral change provide a compelling explanation of the results.

It is important to point out that the present listeners were required to choose a label from the set of 12, which corresponded to vowel categories used in the experiment. However, their responses to low back vowels could potentially differ if the /ʌ/-label were also available. It would also be informative to obtain identification responses in an open set condition and not forcing the listeners to choose among the labels. Also, by including dialect and generational variation in the same task, the present experiment elevated the level of stimulus uncertainty which most likely affected identification responses. While the present design, i.e., full randomization of speakers and vowels, follows naturally the previous experiments by Peterson and Barney (1952) and Hillenbrand *et al.* (1995), their adult speakers were not stratified in terms of dialect and generational differences. It could be the case that blocking the experimental conditions by dialect and age group would result in higher accuracy, at least for selected vowels. All these methodological aspects will need to be addressed in the future.

The results made it clear that listeners' ability to identify the intended vowel was significantly affected by generational differences in the stimulus set. However, the statistically significant generation effects were variable across the

vowel subsets. This variation needs to be explored separately in relation to cross-generational sound change but its important aspect here is that the generational differences in the acoustic vowel characteristics were perceptually salient. It was also the case that female variants produced higher identification rates for the majority of vowel subsets, which is in accord with studies reporting higher intelligibility scores for female speech (e.g., Bradlow *et al.,* 1996; Ferguson, 2004).

The set of discriminant analyses addressed the issue of how many samples of formant frequencies are needed in order to successfully classify the vowels. Considering the overall classification rates and cross-generational classifications, the results support the position dominant in the relevant literature that two sample points considerably improve classification accuracy over a single sample point but adding duration to a single sample (at the vowel midpoint) can also produce modest improvement (e.g., Hillenbrand *et al.,* 1995; Hillenbrand and Nearey, 1999). Inclusion of the third sample point leads to more variable results but generally does not improve the classification accuracy or improves it very little.

While the present results also underscore the importance of dynamic information in formant trajectories in separation of vowel categories, a few discrepancies were found for dialect classification. Considering the entire set of 12 vowels, adding duration to the formant frequencies sampled at the vowel's midpoint resulted in a greater improvement in separating dialects than using formant information from two sample points near vowel onset and offset. For the front vowel set, classification using these two sample points was actually worse than using a single point although adding duration to two endpoint samples caused a dramatic improvement. Ignoring the duration measure, a great improvement in separating dialects was found when adding the midpoint sample to two endpoint samples, which indicates that more complex formant trajectories (such as in some NC variants) require more spectral information in order to be well separated. It is also important to note that vowel duration contributed consistently to improved dialect classification when added to each sample point. Our previous work found that NC vowels are significantly longer than WI vowels, which may be related to significant differences in articulation rate between these two dialects (e.g., Fox and Jacewicz, 2009; Jacewicz *et al.,* 2010). The productions of speakers selected for the present study also demonstrated vowel duration differences and DA indicated that this information is important in separating vowels in the two dialects.

Some of the DA results correspond to listeners' identification responses in the present study. For example, the model trained on the native dialect categorized the vowels of the non-native dialect with much lower accuracy, indicating both large dialectal differences in vowel production and native dialect advantage. The identification results from the present listeners also revealed their generally higher IDRs for vowels in their native dialect. Also, the DA models developed to test generational variation and trained on P-adults showed lower classification accuracy for GP and C groups, which is again consistent with higher overall identification rates for vowels produced by P-adults compared to GP and C.

As a whole, this study demonstrated the effects of dialect and generational variation on formant dynamics and informed us about identification choices made by listeners in response to this type of variation. It also showed that statistical pattern recognition models using time-varying features outperform the simpler models developed on the basis of spectral information at vowel midpoint in dialect and generational classification. Following this study, much work remains to be done to understand listeners' use of acoustic cues in vowel recognition in light of their experience (and lack of it) with dialect-specific production patterns. In turn, these findings will increase our understanding of the progressing sound (vowel) change in English, shedding more light on its possible causes and future directions.

ACKNOWLEDGMENTS

This work was supported by Research Grant No. R01 DC006871 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health. We thank Joseph Salmons for his contributions to this research and Janaye Houghton and Dilara Tepeli for help with data collection. We also thank three anonymous reviewers for their helpful comments on an earlier draft of the paper.

Adank, P., Smits, R., and van Hout, R. (**2004**). "A comparison of vowel normalization procedures for language variation research," J. Acoust. Soc. Am. **116**, 3099–3107.

Andruski, J., and Nearey, T. M. (**1992**). "On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables," J. Acoust. Soc. Am. **91**, 390–410.

Bradlow, A. R., Toretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Clopper, C. G., Pisoni, D. B., and de Jong, K. (**2005**). "Acoustic characteristics of the vowel systems of six regional varieties of American English," J. Acoust. Soc. Am. **118**, 1661–1676.

Ferguson, S. H. (**2004**). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," J. Acoust. Soc. Am. **116**, 2365–2373.

Foulkes, P., and Docherty, G. (**2006**). "The social life of phonetics and phonology," J. Phonetics **34**, 409–438.

Fox, R. A. (**1989**). "Dynamic information in the identification and discrimination of vowels," Phonetica **46**, 97–116.

Fox, R. A., and Jacewicz, E. (**2009**). "Cross-dialectal variation in formant dynamics of American English vowels," J. Acoust. Soc. Am. **126**, 2603–2618.

Fox, R. A., and Nissen, S. (**2005**). "Sex-related acoustic changes in voiceless English fricatives," J. Speech Lang. Hear. Res. **48**, 753–765.

Gordon, M. J. (**2004**). "Investigating chain shifts and mergers," in *The Handbook of Language Variation and Change*, edited by J. K. Chambers, P. Trudgill, and N. Shilling-Estes (Blackwell, Oxford), pp. 244–266.

Hillenbrand, J. M., and Nearey, T. M. (**1999**). "Identification of resynthesized /hVd/ utterances: Effects of formant contour," J. Acoust. Soc. Am. **105**, 3509–3523.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2001**). "Effects of consonant environment on vowel formant pattern," J. Acoust. Soc. Am. **109**, 748–763.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Jacewicz, E., Fox, R. A., and Salmons, J. (**2011a**). "Cross-generational vowel change in American English," Lang. Var. Change **23**, 1–42.

Jacewicz, E., Fox, R. A., and Salmons, J. (**2011b**). "Regional dialect variation in the vowel systems of typically developing children," J. Speech Lang. Hear. Res. **54**, 448–470.

Jacewicz, E., Fox, R. A., and Salmons, J. (**2011c**). "Vowel change across three age groups of speakers in three regional varieties of American English," J. Phonetics **39**, 683–693.

Jacewicz, E., Fox, R. A., and Wei, L. (**2010**). "Between-speaker and within-speaker variation in speech tempo of American English," J. Acoust. Soc. Am. **128**, 839–850.
Jenkins, J. J., Strange, W., and Edman, T. R. (**1983**). "Identification of vowels in 'vowelless' syllables," Percept. Psychophys. **34**, 441–450.
Klecka, W. R. (**1980**). *Discriminant Analysis* (Sage Publications, Newbury Park), pp. 23–51.
Labov, W. (**1994**). *Principles of Linguistic Change. I: Internal Factors.* (Blackwell, Oxford), pp. 115–291.
Labov, W. (**2001**). *Principles of Linguistic Change. II: Social Factors.* (Blackwell, Oxford), pp. 415–465.
Labov, W. (**2010**). *Principles of Linguistic Change. III: Cognitive and Cultural Factors* (Blackwell, Oxford), pp. 1-419.
Labov, W., Ash, S., and Boberg, C. (**2006**). *Atlas of North American English: Phonetics, Phonology, and Sound Change* (Mouton de Gruyter, Berlin), pp. 1–318.
Lobanov, B. (**1971**). "Classification of Russian vowels spoken by different speakers," J. Acoust. Soc. Am. **49**, 606–608.
Milenkovic, P. (**2003**). TF32 software program, University of Wisconsin, Madison, WI.
Nearey, T. M. (**1978**). *Phonetic Feature Systems for Vowels* (Indiana University Linguistic Club, Indiana), pp. 137–188.
Nearey, T. M., and Assmann, P. F. (**1986**). "Modeling the role of inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.
Neel, A. (**2008**). "Vowel space characteristics and vowel identification accuracy," J. Speech Lang. Hear. Res. **51**, 574–585.
Nissen, S., and Fox, R. A. (**2005**). "Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective," J. Acoust. Soc. Am. **118**, 2570–2578.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Sledd, J. (**1966**). "Breaking, umlaut, and the southern drawl," Lang. **42**, 18–41.
Strange, W. (**1989**). "Dynamic specification of coarticulated vowels spoken in sentence context," J. Acoust. Soc. Am. **85**, 2135–2153.
Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. **74**, 695–705.
Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Thomas, E. R. (**2001**). *An Acoustic Analysis of Vowel Variation in New World English* (Duke University Press, Durham, NC), pp. 1–230.
Thomas, E. R. (**2011**). *Sociophonetics: An Introduction* (Palgrave Macmillan, Houndmills, UK), pp. 1–356.
Watson, C. I., and Harrington, J. (**1999**). "Acoustic evidence for dynamic formant trajectories in Australian English vowels," J. Acoust. Soc. Am. **106**, 458–468.
Zahorian, S., and Jagharghi, A. (**1993**). "Spectral shape features versus formants as acoustic correlates for vowels," J. Acoust. Soc. Am. **94**, 1966–1982.