Perceptual learning of systematic variation in Spanish-accented speech

Sabrina K. Sidaras, Jessica E. D. Alexander, and Lynne C. Nygaard
Department of Psychology, Emory University, 532 Kilgo Circle, Atlanta, Georgia 30322

Spoken language is characterized by an enormous amount of variability in how linguistic segments are realized. In order to investigate how speech perceptual processes accommodate to multiple sources of variation, adult native speakers of American English were trained with English words or sentences produced by six Spanish-accented talkers. At test, listeners transcribed utterances produced by six familiar or unfamiliar Spanish-accented talkers. With only brief exposure, listeners perceptually adapted to accent-general regularities in spoken language, generalizing to novel accented words and sentences produced by unfamiliar accented speakers. Acoustic properties of vowel production and their relation to identification performance were assessed to determine if the English listeners were sensitive to systematic variation in the realization of accented vowels. Vowels that showed the most improvement after Spanish-accented training were distinct from nearby vowels in terms of their acoustic characteristics. These findings suggest that the speech perceptual system dynamically adjusts to the acoustic consequences of changes in talker's voice and accent. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3101452]

I. INTRODUCTION

A signature problem in the study of speech perception is how listeners maintain stable linguistic percepts despite the large amount of variability inherent in the acoustic speech signal. Each talker's utterances are uniquely shaped by a host of talker-specific characteristics such as individual identity, emotional state, and region of origin (Frick, 1985; Labov, 1972; Van Lancker *et al.*, 1985). Although these properties are highly informative, differences in the way each talker produces an utterance introduce considerable variability into the speech signal. Listeners must somehow cope with this variability to arrive at the constant linguistic percepts necessary for subsequent stages of linguistic analysis.

Prior research suggests that variability among different talkers may not necessarily be a perceptual problem for listeners but rather a source of lawful variation that is learned, retained, and used during spoken language processing. A number of studies have shown that listeners both attend to variation in talker's voice (Green *et al.*, 1991; Magnuson and Nusbaum, 2007; Mullennix *et al.*, 1989; Mullennix and Pisoni, 1990; Nusbaum and Magnuson, 1997) and retain talker-specific characteristics of speech in memory (Bradlow *et al.*, 1999; McLennan and Luce, 2005; Nygaard *et al.*, 2000; Palmeri *et al.*, 1993). Further, when given experience with particular speakers, listeners appear to engage in perceptual learning of surface characteristics of speech (Allen and Miller, 2004; Ladefoged and Broadbent, 1957; Nygaard and Pisoni, 1998; Nygaard *et al.*, 1994; Yonan and Sommers, 2000), and this learning facilitates the processing of linguistic structure.

Other research has investigated the degree to which listeners can adapt to systematic variation in synthesized, noise-vocoded, and time-compressed speech (Davis *et al.*, 2005; Dupoux and Green, 1997; Greenspan *et al.*, 1988; Schwab *et al.*, 1985). Greenspan *et al.* (1988) exposed listeners to synthetic speech, either word- or sentence-length utterances, over a training period of several days. Listeners who received training showed better transcription accuracy than those listeners who did not receive training. Additional research suggests that listeners can even perceptually accommodate to drastic alterations in the acoustic speech signal, such as time-compressed (Dupoux and Green, 1997) and noise-vocoded speech (Davis *et al.*, 2005).

Although these results demonstrate that listeners perceptually adapt to the unique characteristics of synthetic and altered speech, the variation in these types of signals is highly systematic, altering the speech signal in regularized ways depending on the particular synthesis or resynthesis technique. As a consequence, this type of input is arguably less variable across utterances than are the types of embedded sources of variation found in natural speech. One such source of natural variation that listeners routinely encounter in everyday communication is speech produced by non-native speakers of a particular language or foreign-accented speech. Because utterances produced by non-native speakers are filtered through the articulatory habits and phonological structure of their native language, accentedness systematically affects the linguistic realization of multiple aspects of spoken language (Flege *et al.*, 1997; Flege and Fletcher, 1992; Flege *et al.*, 1999). Systematic variation due to accentedness has been found to influence the intelligibility of non-native speech such that non-native talkers are less intelligible than native talkers, and listening to accented speech requires increased processing effort and time (Goggin *et al.*, 1991; Munro, 1998; Munro and Derwing, 1995; Schmid and Yeni-Komshian, 1999; van Wijngaarden *et al.*, 2002).

One challenge for the listener is that variation due to accent is produced in conjunction with variation due to individual talkers' voices. In order to understand accented speech, listeners must identify the independent contributions of talker-specific variation and the accent-general variation introduced by speakers' non-native articulatory habits and

native phonological structure. Only a handful of studies have begun to examine adaptation to this type of variation (Bradlow and Bent, 2008; Clarke and Garrett, 2004; Weil, 2001). In a recent study, Bradlow and Bent (2008) exposed native English listeners to Chinese-accented English and then tested transcription of English utterances produced by a single novel Chinese- or Slovakian-accented talker. Listeners who received training showed better sentence transcription performance for a novel Chinese- than a novel Slovakian-accented talker at test. Additionally, listeners exposed to multiple accented talkers during training performed better than those trained with a single accented talker. Although this study as well as others suggest that listeners may be sensitive to the lawful variation inherent in accented speech, less clear is the extent to which listeners are learning general systematic attributes of the accent or instead, properties specifically relevant to the particular talker used at test. Studies to date have focused on assessing generalization to just a single novel accented speaker and as such, the extent to which systematic variation is learned during these tasks remains an open question. The current investigation examined the issue of whether listeners learn accent-general or talker-specific properties of variation by determining the extent to which listeners generalize to *multiple* talkers and utterances.

Another question that remains to be addressed concerns *what* properties of foreign-accented speech listeners might be learning with exposure to non-native speech. Previous research has focused almost exclusively on perceptual adaptation to sentence-length utterances (e.g., Bradlow and Bent, 2008) and the extent to which higher-level lexical, semantic, and syntactic constraints might be instrumental in tuning perceptual mechanisms to particular properties of altered or accented speech (Davis *et al.*, 2005; Norris *et al.*, 2003). However, because sentences contain multiple sources of information including prosodic and segmental structure, at issue are what accent-specific properties listeners are learning. When judging degree of accentedness, listeners appear sensitive to both prosodic and segmental aspects of non-native speech (Boula de Mareüil and Vieru-Dimulescu, 2006) and with sentence-length utterances, listeners may be adapting either to global properties such as prosodic and intonational contours or to regularities in the acoustic-phonetic structure of accented speech.

Certainly, previous research suggests that listeners are sensitive to systematic variation due to accent and alter their processing of linguistic structure accordingly (Evans and Iverson, 2004). One example of this perceptual precision comes from several recent studies (Eisner and McQueen, 2005; Kraljic and Samuel, 2006, 2007; Ladefoged and Broadbent, 1957; Norris *et al.*, 2003) demonstrating that listeners are able to use lexical support to shift their phonetic category structure to include unusual pronunciations of particular contrasts. Norris *et al.* (2003) found that when listeners were given experience with ambiguous phonetic segments in lexically constraining contexts, their phonetic category boundaries shifted in keeping with the lexically driven learning. Although these studies suggest that listeners track systematicities in variation at the segmental level and alter their linguistic category structure when relevant to lin-guistic processing, it is unclear to what extent perceptual adjustments occur when listeners are confronted with multiple talkers and items in a high-variability learning and test paradigm.

For the current investigation, a high-variability training paradigm was created in which native English-speaking listeners were exposed to Spanish-accented speech produced by multiple (3 males and 3 females) non-native talkers. At test, listeners were presented either with the same set of six talkers heard during training or with a different set of six Spanish-accented speakers. Assessing generalization to multiple familiar and unfamiliar accented talkers provided a crucial test of the degree to which listeners engage in perceptual learning of the overarching lawful variation found in accented speech. It was predicted that if listeners are simply learning properties that are specific to individual accented talkers encountered during training, then improved transcription performance should be found only for accented talkers that are familiar at test. However, if listeners are perceptually adapting to general, systematic properties of accent, then listeners should generalize both to novel utterances and to multiple unfamiliar speakers.

In addition to assessing generalization of learning, perceptual learning of accented speech was examined using both sentence- and word-length utterances. If listeners are primarily learning the global properties associated with accent, such as prosodic and intonational contours, then accent learning should occur only with sentence-length utterances. However, if listeners are also sensitive to segmental properties of speech that vary with accent, then perceptual learning should be observed with word-length utterances as well.

In addition to general measures of perceptual tuning, we conducted further analyses of the particular types of acoustic-phonetic cues listeners may be using to perceptually adapt to accented speech. Production and identification of accented vowels served as a starting point for the investigation of the fine structure of perceptual learning. Analysis of listeners' identification of a subset of accented vowels that were more or less confusable was performed on the word transcription data from the test phase of the perceptual learning task. If listeners are learning systematic segmental information during training, their identification of certain vowels should be better at test than listeners who were not exposed to the accented speech.

Finally, acoustic analyses were performed to investigate how the systematic variation at the phonetic level may have influenced learning. Both temporal and spectral analyses of the accented Spanish vowels as well as the same vowels produced by native English speakers were compared to determine whether the native Spanish speakers produced systematic cues to particular segments and to what extent those cues were similar to or different from those produced by native English speakers. It was predicted that those vowels that were distinct with respect to temporal or spectral cues would be identified more accurately and learned more readily than those vowels that tended to overlap in acoustic-phonetic space.

II. EXPERIMENT 1

Experiment 1 examined perceptual learning of accented speech using sentence-length utterances. Accented speech differs systematically from native speech not only in segmental characteristics but also in prosodic structure. Experiment 1 examined the extent to which listeners would be able to exploit these multiple sources to perceptually adapt to regularities in accented speech and generalize that learning to both novel utterances and multiple novel talkers.

A. Method

1. Listeners

Listeners were 80 undergraduates who received partial credit in an introductory psychology course. The participants in this and the following experiments were native speakers of English with no reported history of speech or hearing disorders and were not fluent speakers of Spanish.

2. Materials

Twelve native Spanish speakers (6 males and 6 females) from Mexico City were recruited from the Atlanta area. Their mean age in years at the time of recording was 32.75 (range 26–39), on arrival to the U.S. was 26.42 (range 21–34), and when speakers began to learn English was 16.67 (range 2–28). Native English speakers (3 males and 3 females) provided control stimuli.

A set of 100 Harvard sentences (IEEE Subcommittee, 1969) and 144 monosyllabic words was recorded onto digital audiotape and re-digitized at a 22.050 kHz sampling rate, edited into separate files, and amplitude normalized.[1] All sentences were monoclausal and contained five key words (e.g., The *birch canoe slid* on the *smooth planks*). Sentences used at test were mixed with white noise at a +10 signal to noise ratio.

Separate groups of ten listeners transcribed all 100 sentences and 144 words for each of the 12 accented talkers to determine baseline intelligibility. An additional ten native English-speaking listeners rated the accentedness of ten sentence-length utterances from each of the 12 talkers. Listeners rated the accentedness of each utterance on a seven-point Likert-type scale, from 1 = "not accented" to 7 = "very accented". Table I lists mean accent ratings as well as baseline word and sentence intelligibility scores for each talker. Accentedness ratings and baseline intelligibility were correlated, $r = -0.88$, $p < 0.05$, indicating that more intelligible speakers were judged as less accented.

Talkers were divided into two groups for counterbalancing purposes based on mean accentedness (based on sentences) and single word intelligibility score. Each group was made up of three males and three females with approximately equivalent intelligibility and accentedness. Groups did not differ significantly on accentedness, $t(5) = 0.61$, $p = 0.57$ ($M_{\text{group 1}} = 4.12$, $M_{\text{group 2}} = 4.33$) or on intelligibility, $t(10) = 0.23$, $p = 0.75$ ($M_{\text{group 1}} = 46.6\%$, $M_{\text{group 2}} = 49.9\%$).

TABLE I. Accentedness and intelligibility for Spanish-accented speakers.

Speaker group	Gender	Mean accentedness ratings	Mean intelligibility (sentences) (%)	Mean intelligibility (words) (%)
Spanish group 1	Female	5.59	75.6	32.93
	Female	4.43	83.0	39.73
	Female	3.10	89.8	68.80
	Male	4.77	65.9	54.50
	Male	2.83	90.5	58.27
	Male	4.01	82.9	49.20
Spanish group 2	Female	4.31	85.5	48.93
	Female	6.17	74.6	35.20
	Female	4.54	82.6	53.50
	Male	3.55	81.8	42.93
	Male	2.68	90.7	60.27
	Male	4.75	89.0	52.80

3. Procedure

Training varied across conditions but materials and speakers at test remained the same. This design allowed for the comparison of listeners' performance with the exact same items (words and talkers) at test. During training, listeners were exposed to spoken items produced by one of two groups of six Spanish-accented speakers, six native English speakers, or received no training at all. The English training and the no training groups served as controls. Listeners trained with the Spanish-accented speakers heard either the same voices during training and at test or different voices during training and at test. Talker group was counterbalanced such that half the listeners in each condition heard group 1 at test and half heard group 2 at test.

Training phase. Training consisted of four comparison blocks and three variability blocks that were presented in alternation. In each of the comparison blocks, listeners heard each of the six Spanish-accented talkers or native English-speaking controls (3 males and 3 females) produce four different sentences and rated the accentedness of each sentence on a scale of 1–7. In the variability blocks, listeners heard two repetitions of three sentences per speaker presented in random order, with novel sentences in each block. Across repetitions within a block, talker/sentence pairings changed so that listeners never heard the same sentence produced by the same talker more than once. Listeners were asked to type the sentences they heard and were given as much time as needed to transcribe each sentence. After each response, the intended target sentence was presented both on the screen and repeated over the headphones. The training period lasted approximately 40 min. All training sentences were presented in the clear.

Generalization test. Listeners in all conditions heard the same group of six Spanish-accented talkers producing 30 novel sentences at test. Five sentences produced by each talker were presented in random order and listeners performed the transcription task with no feedback. All of the sentences in the test phase were mixed in noise. The listeners trained with Spanish-accented speech all heard a familiar ac-

cent at test. What varied was whether the talkers were familiar (same condition) or unfamiliar (different condition). For the control groups, both accent and talkers' voices were unfamiliar.

Listeners were tested individually in a quiet room. Stimulus presentation and data collection were controlled using PSYSCOPE (Cohen *et al.*, 1993) on a PowerMac G3 computer. The auditory stimuli were presented binaurally over Beyerdynamic DT100 headphones at approximately 75 dB sound pressure level (SPL).

B. Results and discussion

Sentence transcription performance was scored for proportion total words correct in the sentences as well as for proportion key words correct. Proportion total words correct are reported as there were no differences in the effects using total or key words correct.

Training phase. Because performance in the English training group was uniformly high ($M_1=98.9$; $M_2=99.6$; $M_3=99.5$), transcription performance during training was only analyzed for the two Spanish-accented training groups. Participant (F_1) and item (F_2) analyses of variance (ANOVA) were conducted with variability blocks across training (blocks 1–3) and training group (same vs different) as factors. A significant main effect of training block was found for participants, $F_1(2,80)=28.72$, $p<0.001$, *partial* $\eta^2=0.42$ and $F_2(1,52)=1.00$, $p=0.374$, partial $\eta^2=0.037$. The main effect of training group was not significant for participants but was for items, $F_1(1,40)=1.62$, $p=0.21$, partial $\eta^2=0.039$ and $F_2(1,52)=6.39$, $p<0.02$, partial $\eta^2=0.11$. In general, transcription performance improved across blocks for both Spanish-accented training groups: same ($M_1=91.1$, $M_2=93.4$, $M_3=94.4$) and different ($M_1=92.2$, $M_2=94.6$, $M_3=95.3$), with better performance for the different than same group for items. No significant interaction between training group and training blocks was found either for participants or items. Planned comparisons (for participants) showed significant improvement in transcription performance between blocks 1 and 2, $F(1,40)=22.64$, $p<0.001$, partial $\eta^2=0.36$, and between blocks 2 and 3, $F(1,40)=6.00$, $p<0.02$, partial $\eta^2=0.13$.

Generalization test. Figure 1 shows percent correct transcription performance at test for each training group. One-way participant (F_1) and item (F_2) ANOVAs assessing listeners' performance at test revealed a significant main effect of training group, $F_1(3,76)=4.97$, $p<0.004$, partial $\eta^2=0.16$ and $F_2(3,90)=20.13$, $p<0.001$, partial $\eta^2=0.40$. Planned comparisons revealed no significant differences between the Spanish-accented training groups, same ($M=62.0$, SD=7.7) vs different ($M=61.1$, SD=5.5), $p_1=0.75$, $p_2=0.71$, or between the two control groups, English ($M=55.7$, SD=6.3) vs no training ($M=56.7$, SD=6.4), $p_1=0.66$, $p_2=0.34$, at test. However, listeners who received training with Spanish-accented speech ($M=60.1$, SD=6.4) performed better at test than listeners who received English or no training ($M=55.2$, SD=5.9), $F_1(1,76)=13.97$, $p<0.001$, partial $\eta^2=0.16$ and $F_2(1,90)=35.14$, $p<0.001$, partial $\eta^2=0.54$.
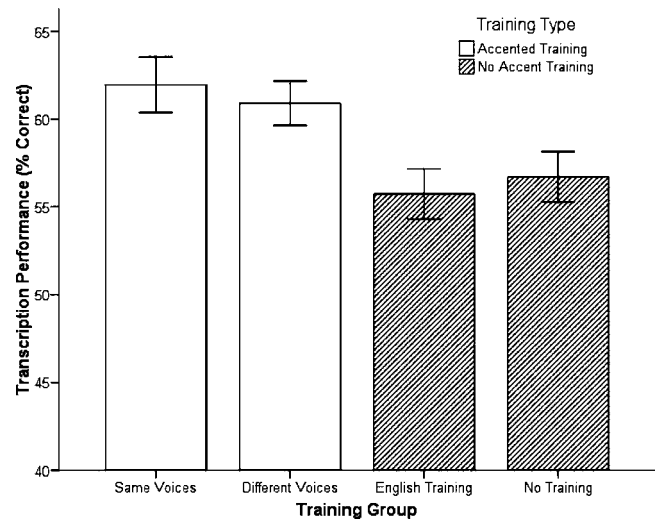

FIG. 1. Mean transcription performance at test for sentences as a function of training group.

These findings indicate that training with Spanish-accented speech resulted in perceptual adaptation to accent-general characteristics of non-native speech. Listeners generalized both to novel utterances and to novel voices within the same accent group suggesting that learning was not tied to particular tokens or talkers. In addition, improvement was observed after relatively brief exposure to accented speech suggesting that listeners adapted quickly to the lawful variation in accented speech.

Little evidence was found for talker-specific learning in addition to accent-general learning in this task. Perhaps because listeners received relatively more experience with the Spanish accent and relatively less experience with any particular talker, this type of training may have encouraged listeners to track commonalities *across* speakers rather than focus on the idiosyncrasies of any particular talker.

III. EXPERIMENT 2

Experiment 2 examined listeners' ability to perceptually adapt to properties of accented speech in single words. The results of experiment 1, along with previous research (e.g., Bradlow and Bent, 2008), suggest that listeners are sensitive to the regularities found in foreign-accented sentences. Using single words at training and test reduced the availability of global properties and allowed us to examine the extent to which listeners can learn systematic variation specific to the acoustic-phonetic structure of accented speech.

A. Method
1. Listeners

Listeners were 98 undergraduate students who received partial course credit in an introductory psychology course for their participation.

2. Materials

The same non-native Spanish and native English speakers that were used in the previous experiment also recorded a list of 144 monosyllabic English words (72 *easy* and 72

J. Acoust. Soc. Am., Vol. 125, No. 5, May 2009

Sidaras *et al.*: Perceptual learning of accented speech   3309

hard).[2] Easy words were high frequency words (M=309.69; Kučera and Francis, 1967) with few (M=38.32) low frequency neighbors (e.g., size, piece; Luce and Pisoni, 1998). Hard words were low frequency words (M=12.21) with many (M=282.22) high frequency neighbors (e.g., sane, lace). Both easy and hard words were rated as being highly familiar (M=6.97; on a scale of 1–7 with 1 being not familiar at all and 7 being highly familiar (Nusbaum *et al.*, 1984).

3. Procedure

Training phase. The same design was used as in experiment 1. Because more words than sentences were available, in each variability block, listeners heard two repetitions of each talker producing four different English words, with novel words in each block. All training words were presented in the clear, and listeners received feedback as in experiment 1 on their transcriptions.

Generalization test. At test, listeners transcribed a total of 48 novel accented words, eight words from each talker. All of the words in the test phase were presented in the clear, and no feedback was given. All other aspects of the procedure were the same as in experiment 1.

B. Results and discussion

Transcription accuracy was averaged across words for each participant. Words were scored as correct if listeners provided either the correct spelling or a homophone equivalent.

Training phase. As in experiment 1, since transcription performance was uniformly high in the English control condition (M_1=93.6, M_2=92.6, M_3=93.5), training performance was only evaluated for the two Spanish-accented conditions. Participant (F_1) and item (F_2) ANOVAs were conducted with training block (blocks 1–3) and training group (same vs different) as factors. A significant main effect of training block was found for participants, $F_1(2,96)=29.00$, $p<0.001$, partial $\eta^2=0.38$ and $F_2(2,70)=0.99$, $p=0.37$, partial $\eta^2=0.01$. The main effect of training group was not significant for participants, but was for items, $F_1(2,96)=1.10$, $p=0.30$, partial $\eta^2=0.02$, and $F_2(2,70)=4.5$, $p<0.05$, partial $\eta^2=0.06$. Transcription performance changed as a function of block for both Spanish-accented training groups: same (M_1=59.3, M_2=58.7, M_3=66.3) and different (M_1=58.5, M_2=56.1, M_3=64.6) with indication of better performance for the same than different group. No significant interaction between training group and training block was found for participants or items. Planned contrasts (for participants) showed significant improvement in transcription performance between blocks 2 and 3, $F(1,48)=50.65$, $p<0.001$, partial $\eta^2=0.51$, but not between blocks 1 and 2, $p=0.16$.

Generalization phase. Figure 2 shows transcription performance during the generalization test for each training group condition. One-way participant (F_1) and item (F_2) ANOVAs revealed a significant main effect of training group, $F_1(3,94)=4.08$, $p<0.01$, partial $\eta^2=0.11$ and $F_2(3,141)=3.06$, $p<0.05$, partial $\eta^2=06$. Planned comparisons showed no significant differences between the Spanish-accented training groups, same (M=48.1, SD=6.2) vs differ-
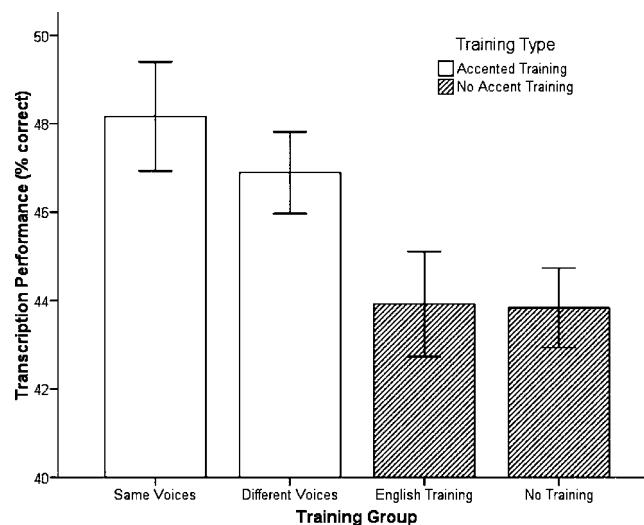

FIG. 2. Mean transcription performance at test for words as a function of training group.

ent (M=46.9, SD=4.6), p_1=0.399, p_2=0.40, or between the two control groups, English (M=43.9, SD=5.8) vs no training (M=43.8, SD=4.4), p_1=0.96, p_2=0.96. However, a significant difference was found between listeners that received Spanish-accented training (M=47.5, SD=5.4) and those that did not (M=43.9, SD=5.1), $F_1(1,96)=11.7$, $p<0.05$, partial $\eta^2=0.06$ and $F_2(1,47)=6.36$, $p<0.05$, partial $\eta^2=0.12$.

These results indicate that a brief training session with isolated accented words produced perceptual adaptation. As in experiment 1, the intelligibility benefits of the training session generalized both to novel utterances and to novel talkers. The finding that perceptual learning occurred with single words suggests that listeners can attend to and learn not only the unique prosodic structure of accented speech but also the fine-grained details of the acoustic-phonetic structure of accented speech.

IV. PERCEPTION AND PRODUCTION OF SPANISH-ACCENTED VOWELS

In order to examine precisely what properties of Spanish-accented speech listeners were learning, we examined the perception of individual accented vowels from experiment 2 to determine which ones showed improvement as a function of training. In addition, we conducted acoustic analyses of the Spanish-accented vowels and the same vowels produced by native English speakers to determine if the native Spanish speakers produced reliable cues to particular segments.

Accented vowel production and perception was deemed a good starting place because the accuracy of both production and perception of vowels in a non-native language varies as a function of native language background (Bohn and Flege, 1992; Flege *et al.*, 1997; Flege *et al.*, 1999; Flege *et al.*, 2003; Munro, 1993). With respect to the present study, the Spanish vowel inventory /i, e, a, o, u/ differs from English both in number (Spanish has 5 vowels and English has approximately 11) and in their realization in spectral and temporal space (Bradlow, 1995). Based on this previous research, the native Spanish speakers in the present study

TABLE II. Confusion matrix for error analyses of word transcriptions.

Intended targets	Listeners' responses						
	/i/	/ɪ/	/e/	/æ/	/ʌ/	/a/	Other
No accented training							
/i/	**45**	46	6				3
/ɪ/	31	**48**	10				11
/e/	1	2	**95**				2
/æ/			1	**71**	15	1	12
/ʌ/		1		10	**37**	25	27
/a/	1				15	**76**	8
Spanish-accented training							
/i/	**51**	41	4	1			3
/ɪ/	28	**48**	13	1			10
/e/	1		**98**	1			
/æ/				**82**	13	1	4
/ʌ/		1		11	**36**	24	28
/a/		1			15	**80**	4

Values represent percent responses to target.

should have difficulty producing vowels that have no counterpart in their native language vowel inventory. In turn, the native English-speaking listeners would be expected to have difficulty identifying those same vowels. To that end, patterns of errors or confusions for vowel identification for trained and untrained listeners were calculated. The error analyses then served as a guide for the acoustic analyses to determine how the native Spanish speakers were producing the English vowels and which cues the English listeners were using to perceptually learn the systematic variation in the accented speech.

A. Error analyses

Analyses of vowel identification and confusions were calculated using the word transcription responses of listeners who participated in experiment 2. Evaluations of listeners' responses at test were thus necessarily limited by the orthographic constraints of written English. It should be noted, however, that these constraints were the same for both the trained and untrained groups. Listeners trained on accented voices, whether same or different, were grouped together ($n=49$) and listeners not trained with Spanish-accented speech were grouped together ($n=49$). The vowel identification analyses were carried out for target words with the vowels /i/, /ɪ/, /e/, /æ/, /ʌ/, and /a/. Other vowels were excluded either because they were less frequent in our set or because the initial or final consonant heavily influenced the vowel (e.g., words with /r/ immediately following the vowel and words that began with /r/ or /w/). The data used for the analyses included listener responses to multiple words in each vowel category; 490 responses to /i/, 588 responses to /ɪ/, 588 responses to /e/, 294 responses to /æ/, 686 responses to /ʌ/, and 294 responses to /a/.

Table II shows confusion matrices of target vowels for trained and untrained listeners. Cells reflect percent identifications, which take into account the number of possible tokens. Regardless of training, the high front vowels /i/ and /ɪ/ were frequently confused with one another, while the /e/ vowel was relatively well identified. These confusions follow from a mapping between Spanish and English vowels with /ɪ/, a vowel not found in Spanish, being confused with other high front vowels /i/ and /e/. Likewise, for both trained and untrained listeners, the low vowels /æ/, /ʌ/, and /a/ were highly confusable. The accented /ʌ/, a vowel not in the Spanish inventory, was particularly difficult for the native English-speaking listeners. The pattern of confusions suggests that the native Spanish speakers had difficulty producing vowels that fell outside their vowel inventory (/ɪ/, /æ/, and /ʌ/) and that speakers were referencing their own vowel categories in order to approximate the non-native vowels (/i/ and /a/).

In addition to the overall pattern of confusions, the results show that listeners transcribed at least a subset of accented vowels more accurately after training with accented speech. Targeted comparisons of identification performance for listeners who did and did not receive accented training were completed for each of the accented vowels /i/, /ɪ/, /e/, /æ/, /ʌ/, and /a/ that were analyzed. The vowels /i/, /æ/, and /a/ showed significantly higher accuracy for trained than untrained listeners, p's<0.05. The vowels /ɪ/, /e/, and /ʌ/ did not show a significant difference between trained and untrained listeners, all p's>0.05. The improvement in identification for particular accented vowels indicates that listeners might have been learning specific information during training that allowed them to better discriminate and identify particular vowels.

The vowel-specific nature of the learning guided the analysis of the acoustic-phonetic correlates to identification performance. If training with Spanish-accented speech reduced the confusability of vowels such as /i/, /æ/, and /a/, then acoustic-phonetic characteristics of these vowels should distinguish them from other vowels in the listeners' repertoire. In particular, temporal and spectral characteristics of the Spanish-accented vowels were examined to determine

TABLE III. Mean values and standard deviations for duration, F1, and F2 for native English and Spanish speaker groups.

Speaker group	Vowel	Duration		F1		F2	
		M	SD	M	SD	M	SD
English	/i/	194.16	32.31	347.79	48.36	2580.52	237.60
	/ɪ/	161.24	24.69	526.87	126.61	2079.26	162.67
	/e/	227.50	45.09	444.87	97.62	2462.76	164.09
	/æ/	228.52	30.92	837.04	208.61	1852.45	109.34
	/ʌ/	188.21	28.33	686.44	190.17	1413.56	132.22
	/a/	235.61	44.23	768.69	198.51	1221.27	72.73
Spanish	/i/	178.19	36.63	356.76	53.84	2433.36	229.45
	/ɪ/	169.59	40.71	392.07	48.71	2395.58	282.15
	/e/	235.87	40.51	439.96	52.00	2367.90	263.70
	/æ/	217.56	35.55	777.84	143.82	1637.71	144.31
	/ʌ/	189.02	41.32	634.20	85.69	1352.06	208.50
	/a/	195.63	34.55	664.78	89.72	1310.77	157.48

Mean values represent averages of tokens of each vowel for each speaker and standard deviations represent the variance of the means of these tokens.

which properties might be contributing both to the overall identification of these vowels and to the improvement that listeners achieve with training.

B. Acoustic analyses

Acoustic analyses of vowel duration and first (F1) and second (F2) formant center frequencies were carried out using PRAAT sound analysis software (Boersma and Weenink, 2006) for the English vowels embedded in words produced by the 12 native Spanish and 6 native English speakers from experiment 2. Only words with the target vowels /i/, /ɪ/, /e/, /æ/, /ʌ/, and /a/ were analyzed. For each of these vowels, between 12 and 16 tokens were analyzed for each speaker (for a total of 144–192 tokens per vowel). Three trained coders completed all acoustic analyses, with a single coder completing all analyses for vowels produced by a single talker. Recall that the vowels were embedded in words that contained a variety of consonant contexts. Although the context varied, it was consistent across both the Spanish-accented and native English speakers. Criteria for determining vowel onset and offset were taken from Munson and Solomon (2004). Vowel duration was determined from onset and offset times. Measurements of the first and second formant frequencies were taken at the midpoint of the vowel. Inter-rater reliability for vowel onset and offset measures was assessed using a subset of six vowels for all talkers (12 Spanish-accented talkers, 6 native English talkers; 108 tokens). Reliability was good with 86% agreement among all three coders. Table III reports mean values and standard deviations of duration, F1, and F2 for each vowel.

Based on the differences in the identification scores and patterns of confusion for each vowel, we expected temporal or spectral overlap for vowels that were confusable (e.g., /i/ and /ɪ/) and less overlap for those that were not confusable (e.g., /i/ and /e/). Further, we expected that the specific vowels that were better identified after learning (/i/, /æ/, and /a/) would have temporal or spectral properties that distinguished them from other intended vowels. Separate focused analyses

were conducted on the three high front vowels /i/, /ɪ/, and /e/ and on the three low vowels /æ/, /ʌ/, and /a/. All follow-up comparisons used a Bonferroni corrected alpha of 0.0125.

Temporal characteristics. Figures 3(a) and 3(b) show mean duration measures for each English vowel produced by the native English and Spanish speaker groups. Separate ANOVAs were performed on duration with speaker group (English or Spanish) as a between group factor and either vowel group /i/, /ɪ/, and /e/ or vowel group /a/, /æ/, and /ʌ/ as the within group factor. For the /i/, /ɪ/, and /e/ vowel grouping, a main effect of vowel, $F(2,32)=63.90$, $p<0.001$, partial $\eta^2=0.80$, but no main effect of speaker group or interaction was found. The pattern of duration differences across the three vowels, /i/, /ɪ/, and /e/, for native Spanish speakers was similar to those of native English speakers. Native Spanish speakers produced the English vowel /e/ with longer durations than either /i/ or /ɪ/, both p's<0.001. In addition, the vowel /i/ had longer durations than /ɪ/, $p<0.001$. The relative differences in duration among these vowels are consistent with previous findings (e.g., Flege *et al.*, 1997) and suggest that duration is a reliable cue that listeners may use to distinguish among Spanish-accented productions of these vowels.

For the vowels /a/, /æ/, and /ʌ/, the ANOVA revealed a significant interaction between speaker group and vowel, $F(2,32)=10.80$, $p<0.001$, partial $\eta^2=0.40$, indicating that the pattern of durations across vowels differed as a function of speaker group. Comparisons across vowels for the native Spanish speakers revealed significant differences in duration between productions of /æ/ and /ʌ/, $p<0.001$, and between /æ/ and /a/, $p<0.001$, but not between /ʌ/ and /a/, $p=0.113$. Spanish speakers did not distinguish the vowels /ʌ/ and /a/ in terms of duration and exhibited a pattern of durations across vowels that differed from the native English speakers. Recall that listeners who were trained with Spanish-accented speech showed better identification of the /æ/ vowel than those that were not trained. The pattern of differences in duration sug-
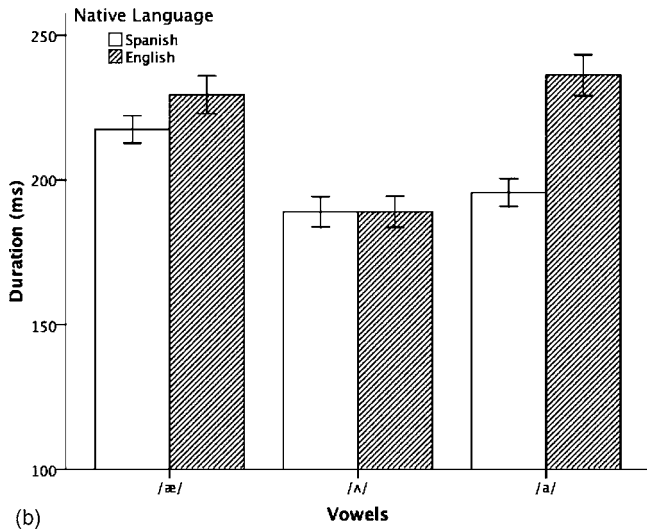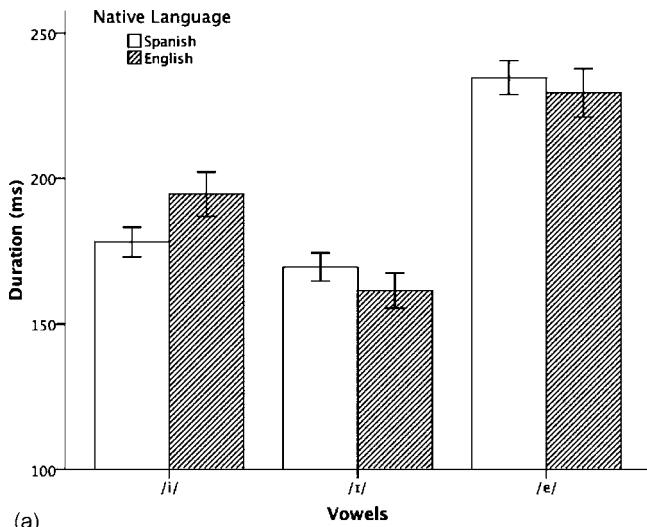
FIG. 3. Mean durations for native English and Spanish speaker groups for (a) vowels /i/, /ɪ/, and /e/ and (b) /æ/, /ʌ/, and /ɑ/.

gests that this property could serve as one cue for the English listeners that distinguishes /æ/ from similar vowels for the native Spanish speakers.

Spectral characteristics. Figure 4 shows mean F1 and F2 values for the vowels /i/, /ɪ/, and /e/ for native English and Spanish speakers. Separate ANOVAs were performed on F1 and F2 with speaker group (English or Spanish) and vowel group (/i/, /ɪ/, and /e/) as factors.

For measures of F1, a significant interaction between speaker group and vowel was found, $F(2,32)=20.41$, $p<0.001$, partial $\eta^2=0.56$. Comparisons among vowels for the native Spanish speakers revealed that all pairwise comparisons were significant; /i/ and /e/, $p<0.001$, /ɪ/ and /e/, $p<0.001$, and /ɪ/ and /i/, $p<0.01$. Although the Spanish speakers were distinguishing among the three vowels, the pattern was very different from that produced by the native English speakers. For the Spanish speakers, F1 values for /ɪ/ fell between values for /i/ and /e/. For English speakers, F1 values for /e/ fell between /i/ and /ɪ/. The overlap in F1 frequencies among the three vowels coupled with the lower F1 frequency for /ɪ/ may have contributed to the confusability of /i/ and /ɪ/.
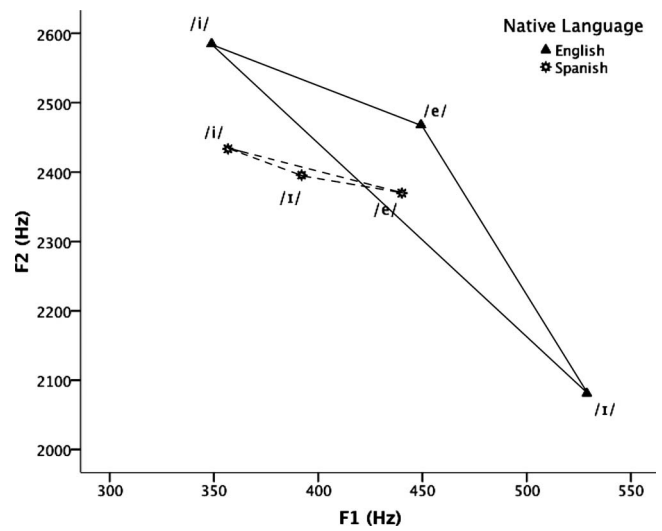

FIG. 4. Mean F1 and F2 values in hertz for /i/, /ɪ/, and /e/ for native English and Spanish speakers.

Turning to F2, a significant interaction between speaker group and vowel was found, $F(2,32)=38.57$, $p<0.001$, partial $\eta^2=0.71$. Comparisons across vowels for Spanish speakers revealed significant differences only between /e/ and /i/, $p=0.011$. It appears that accented speakers had difficulty producing /ɪ/, realizing the vowel with lower F1 and higher F2 values than native English speakers. These modified spectral characteristics overlapped with adjacent vowel categories and may have contributed to the confusability of /i/ and /ɪ/.

Figure 5 shows mean F1 and F2 values for the vowels /æ/, /ʌ/, and /ɑ/ for native English and Spanish speakers. Again, separate ANOVAs were performed on F1 and F2 with speaker group (English or Spanish) and vowel (/æ/, /ʌ/, and /ɑ/) as factors. For F1, there was a significant main effect of vowel $F(2,32)=23.28$, $p<0.001$, partial $\eta^2=0.59$, but no effect of speaker group and no interaction. For both groups of speakers, F1 values for /ʌ/ were significantly lower than for /ɑ/, $p<0.003$, which in turn had lower F1 values than for the vowel /æ/, $p<0.003$. These results show that native
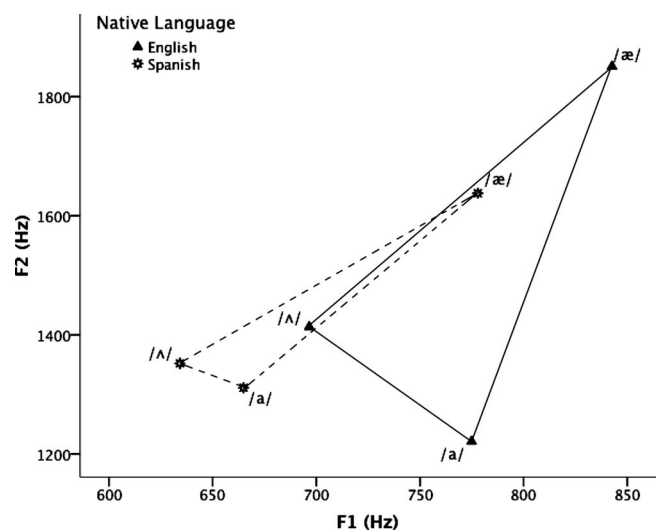

FIG. 5. Mean F1 and F2 values in hertz for /æ/, /ʌ/, and /ɑ/ for native English and Spanish speakers.

Spanish speakers distinguished among these low vowels with respect to F1, approximating the pattern produced by native English speakers. In turn, these vowels were relatively less confusable than the high front vowels.

For measures of F2, there was a significant interaction between speaker group and vowel, $F(2,32)=8.36$, $p<0.001$, partial $\eta^2=0.34$. Comparisons of relative vowel differences for the native Spanish speakers revealed differences between /æ/ and /ʌ/, $p<0.001$, and between /æ/ and /a/, $p<0.001$. No significant difference was found between /a/ and /ʌ/, $p=0.221$.

Summary. These findings suggest that acoustic characteristics of the Spanish-accented English vowels may be related to the perceptual confusions observed for the native English listeners. Particular vowels or particular sets of vowels that were confusable to the English listeners also had temporal and/or spectral characteristics that overlapped in acoustic-phonetic space.

In particular, the vowels /i/ and /ɪ/ were found to be highly confusable in the error analyses (see Table II), which seems to correspond to the observed overlapping spectral characteristics of these vowels. Recall, however, that identification of the highly confusable /i/ was better for those listeners who received accentedness training than for those that did not. To speculate, listeners who received training with accented speech may have begun to distinguish among high front vowels within the Spanish speakers' relatively crowded vowel space by attending to the properties of accented productions of both /i/ and /e/ that proved to be similar to the native English productions, resulting in /i/ being susceptible to perceptual learning and /e/ being *a priori* less confusable.

For the trio of vowels /æ/, /ʌ/, and /a/, the vowel /ʌ/ was found to be highly confusable in the error analyses and the vowel /æ/ showed significant improvement as a function of training with accented speech. Likewise, the native Spanish speakers did not distinguish /ʌ/ from /a/ either with respect to duration or with respect to F2, perhaps making /ʌ/ less distinct perceptually. In contrast, the native Spanish speakers produced the vowel /æ/ with temporal and spectral properties that were significantly different from either /ʌ/ or /a/. Although the accented /æ/ was produced with a significantly lower F2, listeners with accentedness training may have been sensitive to the distinctive constellation of cues that set /æ/ apart, at least in this limited set of productions and analyses.

V. GENERAL DISCUSSION

The objective of this study was to investigate the nature and extent of perceptual learning of foreign accented speech. Perceptual learning of Spanish-accented sentence- and word-length utterances was examined in a high-variability training and test paradigm. We sought to determine whether listeners learn the systematic variation specific to accent by examining generalization of learning to multiple familiar and unfamiliar accented talkers. The results showed that after only a brief training period with sentences or words, listeners showed an increased ability to transcribe novel accented words and sentences produced by familiar talkers. Most re-markably, listeners generalized, showing increased transcription performance for a group of six unfamiliar talkers from the same accent group.

Previous research has demonstrated perceptual learning in speech and language processing in general (Davis *et al.*, 2005; Dupoux and Green, 1997; Greenspan *et al.*, 1988; Nygaard and Pisoni, 1998) and for accented speech in particular (e.g., Bradlow and Bent, 2008). However, studies of accommodation to accented speech have focused on generalization to a single novel accented talker (Bradlow and Bent, 2008; Weil, 2001; Clarke and Garrett, 2004). The current findings demonstrate generalization to multiple talkers, suggesting that perceptual learning occurs for accent-general properties of speech and is not tied to particular talker- or item-specific characteristics.

Our findings also begin to pinpoint the nature of the perceptual learning process. Previous studies have almost exclusively used sentence-length utterances to evaluate perceptual adaptation to accented speech (Bradlow and Bent, 2008; Clarke and Garrett, 2004; Weil, 2001) leaving open the question of whether listeners are learning global prosodic features or regularities in phonological form. In the current investigation, listeners used information present in both sentence- and word-length utterances, suggesting a sensitivity to regularities in the acoustic-phonetic structure of accented speech.

In order to confirm that learning was taking place at a segmental level, perceptual confusions for a subset of vowels were examined for listeners who did and did not receive training. The results showed that those listeners who received training with accented speech showed better identification of certain accented vowels (/i/, /æ/, and /a/) than untrained listeners. It appears that listeners learned specific segmental information during training that allowed them to better discriminate and identify particular vowels.

In addition to perceptual confusions, acoustic analyses were performed to investigate which acoustic-phonetic cues of the accented vowels listeners may have learned. The pattern of listeners' vowel confusions suggests that, not surprisingly, the native Spanish speakers had difficulty producing vowels that fall outside their native vowel inventory (/ɪ/, /æ/, and /ʌ/). However, although vowels that were highly confusable to listeners had overlapping temporal and/or spectral characteristics, the native Spanish speakers did appear to produce systematic segmental acoustic-phonetic variation that may have contributed, at least in part, to the perceptual learning of the Spanish-accented speech. For instance, the low vowels /ʌ/ and /a/ were not distinct with respect to duration, but were distinguished by the native Spanish speakers with spectral properties. Thus, with training English listeners may have learned to rely to a greater extent on particular spectral cues for these vowels.

These findings are generally consistent with previous experiments that have shown perceptual adjustments of phoneme categories as a result of experience with unusual pronunciations (Eisner and McQueen, 2005; Kraljic and Samuel, 2006; Norris *et al.*, 2003). The present findings confirm that when listeners are exposed to variation in accented speech, they are able to extract specific systematic information on a segmental level that generalizes to novel talkers'

voices. Although in some studies (Kraljic and Samuel, 2006, 2007; Norris *et al.*, 2003) perceptual learning of alternate pronunciations generalizes to different talkers' voices, other research with different contrasts has found that learning seems to be talker-specific (Eisner and McQueen, 2005). In the present experiment, listeners did generalize, indicating that listeners were able to learn which characteristics of the accented speech should be attributed to consistent properties of talker's voice and which characteristics are due to cross-speaker regularities in accent.

Exposure to extensive variability during training may be necessary for listeners to extract the systematicities present in accented speech. Previous studies have shown that high stimulus variability during training facilitates second language vocabulary learning (Barcroft and Sommers, 2005; Sommers and Barcroft, 2007) as well as the learning of non-native phonetic categories (Logan *et al.*, 1991; Lively *et al.*, 1993, 1994). In the present experiments, although training with accented speech was extremely brief, listeners were exposed to many novel voice-word pairings during both training and at test. The opportunity both to compare tokens across the training blocks as well as from multiple talkers may have allowed listeners to generalize learning to novel accented utterances and speakers.

It should be noted that all this variability, while potentially necessary for robust learning, made the listeners' task extremely difficult both during training and at test. Recall that listeners encountered spoken utterances produced by *multiple* familiar or unfamiliar accented talkers, and consequently, were forced to readjust to a new talker's voice on a trial-by-trial basis. Previous research has established that changes from trial to trial in characteristics of spoken language such as talker's voice incur a processing cost (Mullennix *et al.*, 1989). Nevertheless, listeners learned to parse multiple sources of variability, dynamically attributing variance in the speech signal to changes in the linguistic, talker-specific, and accent-general structure of speech.

These findings are consistent, in a broad sense, with accounts that assume that representation of spoken language includes both perceptual and linguistic properties of speech (Goldinger, 1998; Johnson, 1997; Jusczyk, 1997; Nygaard *et al.*, 1994; Pisoni, 1997). In this sense, perceptual learning of accented speech may be a form of perceptual expertise or automaticity that relies on the accumulation of representations which include the lawful variation in Spanish-accented speech (see Logan, 1988; Ettlinger, 2007). Alternatively, listeners may be tuning their procedural memory or normalization routines in an accent-general fashion (Kolers and Roediger, 1984; Nusbaum and Morin, 1992). Rather than explicitly representing perceptual details of spoken language, listeners may engage in a normalization procedure that becomes tuned to unravel the combined contributions of a particular accent, talker's voice, and other sources of variation.

Taken with previous findings, our data suggest that listeners appear to be exquisitely sensitive to systematic variation in speech and alter their processing or representation of linguistic structure accordingly (e.g., Eisner and McQueen, 2005; Norris *et al.*, 2003; Kraljic and Samuel, 2006, 2007). Perceptual processing and representation of spoken language

appear to include and utilize surface characteristics of speech in linguistic processing. Listeners perceptually adapt as they build up a repertoire of experiences with accented speech that in turn facilitates later processing of the linguistic structure of speech. By engaging in perceptual learning of the lawful variation inherent in accented speech, listeners appear to be sensitive to the details of segmental variability resulting from the complex relationship between linguistic environment, idiosyncratic talker-specific variability, and variation due to properties of the accent itself.

[1]Spanish speakers were given all stimulus materials before the date of recording to familiarize themselves with the materials in order to decrease the chance of making production errors during recording.

[2]Easy and hard words were used in order to evaluate the effects of lexical properties on perceptual learning. In this experiment, lexical properties influenced overall performance level but did not interact with any other variables in experiment 2. Participant (F_1) and item (F_2) ANOVAs with training groups (same, different, English, and no training) and word types (easy vs hard) factors revealed no interaction between training groups and easy/hard word performance, $F_1(3,94)=0.275$, $p=0.844$, partial $\eta^2 =0.009$ and $F_2(1,70)=0.413$, $p=0.523$, partial $\eta^2=0.006$, but there were main effects of both word type and training groups. Transcription performance for hard words ($M=30.5$, SD$=8.2$) was significantly worse than for easy words ($M=60.9$, SD$=7.8$), $F_1(1,94)=660.9$, $p<0.001$, partial $\eta^2 =0.875$ and $F_2(1,70)=19.98$, $p<0.001$, partial $\eta^2=0.12$. The main effect of word type indicates that neighborhood density and word frequency contributed overall to transcription performance, but did not seem to affect or be affected by the learning process in this task. As such, these properties are not discussed further in the current investigation.

Allen, J. S., and Miller, J. L. (**2004**). "Listener sensitivity to individual talker differences in voice-onset-time," J. Acoust. Soc. Am. **115**, 3171–3183.

Barcroft, J., and Sommers, M. S. (**2005**). "Effects of acoustic variability on second language vocabulary learning," Stud. Second Lang. Acquis. **27**, 387–414.

Boersma, P., and Weenink, D. (**2006**). "Praat: Doing phonetics by computer," from http://www.praat.org (Last viewed January, 2006), Version 5.0.23, computer program.

Bohn, O.-S., and Flege, J. (**1992**). "The production of new and similar vowels by adult German learners of English," Stud. Second Lang. Acquis. **14**, 131–158.

Boula de Mareüil, P., and Vieru-Dimulescu, B. (**2006**). "The contribution of prosody to the perception of foreign accent," Phonetica **63**, 247–267.

Bradlow, A. R. (**1995**). "A comparative acoustic study of English and Spanish vowels," J. Acoust. Soc. Am. **97**, 1916–1924.

Bradlow, A. R., and Bent, T. (**2008**). "Perceptual adaptation to non-native speech," Cognition **106**, 707–729.

Bradlow, A. R., Nygaard, L. C., and Pisoni, D. B. (**1999**). "Effects of talker, rate, and amplitude variation on recognition memory," Percept. Psycho-

J. Acoust. Soc. Am., Vol. 125, No. 5, May 2009

Sidaras *et al.*: Perceptual learning of accented speech   3315

phys. **61**, 206–219.

Clarke, C. M., and Garrett, M. F. (**2004**). "Rapid adaptation to foreign-accented English," J. Acoust. Soc. Am. **116**, 3647–3658.

Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (**1993**). "PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers," Behav. Res. Methods Instrum. Comput. **25**, 257–271.

Davis, M. H., Johnsrude, I. S., Hervaise-Adelman, A., Taylor, K., and McGettigan, C. (**2005**). "Lexical information drives perceptual learning distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**, 222–241.

Dupoux, E., and Green, K. (**1997**). "Perceptual adjustment to highly compressed speech: Effects of talker and rate changes," J. Exp. Psychol. Hum. Percept. Perform. **23**, 914–927.

Eisner, F., and McQueen, J. M. (**2005**). "The specificity of perceptual learning in speech processing," Percept. Psychophys. **67**, 224–238.

Ettlinger, M. (**2007**). "Shifting categories: An exemplar-based computational model of chain shifts," Paper presented at the 29th Annual Meeting of the Cognitive Science Society, Nashville, TN.

Evans, B. G., and Iverson, P. (**2004**). "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences," J. Acoust. Soc. Am. **115**, 352–361.

Flege, J., Bohn, O.-S., and Jang, S. (**1997**). "The effect of experience on nonnative subjects' production and perception of English vowels," J. Phonetics **25**, 437–470.

Flege, J., MacKay, I., and Meador, D. (**1999**). "Native Italian speakers' production and perception of English vowels," J. Acoust. Soc. Am. **106**, 2973–2987.

Flege, J., Schirru, C., and MacKay, I. (**2003**). "Interaction between the native and second language phonetic subsystems," Speech Commun. **40**, 467–491.

Flege, J. E., and Fletcher, K. L. (**1992**). "Talker and listener effects on the degree of perceived foreign accent," J. Acoust. Soc. Am. **91**, 370–389.

Frick, R. W. (**1985**). "Communicating emotion: The role of prosodic features," Psychol. Bull. **97**, 412–429.

Goggin, J., Thompson, C., Strube, G., and Simental, L. (**1991**). "The role of language familiarity in voice identification," Mem. Cognit. **19**, 448–458.

Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychol. Rev. **105**, 251–279.

Green, K. P., Kuhl, P. K., Meltzoff, A. N., and Stevens, E. B. (**1991**). "Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect," Percept. Psychophys. **50**, 524–536.

Greenspan, S., Nusbaum, H. C., and Pisoni, D. B. (**1988**). "Perceptual learning of synthetic speech produced by rule," J. Exp. Psychol. Learn. Mem. Cogn. **14**, 421–433.

IEEE Subcommittee (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

Johnson, K. (**1997**). in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego, CA), pp. 145–166.

Jusczyk, P. W. (**1997**). *The Discovery of Spoken Language* (MIT Press, Cambridge, MA).

Kolers, P. A., and Roediger, H. L. III (**1984**). "Procedures of mind," J. Verbal Learn. Verbal Behav. **23**, 425–449.

Kraljic, T., and Samuel, A. G. (**2006**). "Generalization in perceptual learning for speech," Psychon. Bull. Rev. **13**, 262–268.

Kraljic, T., and Samuel, A. G. (**2007**). "Perceptual adjustments to multiple speakers," J. Mem. Lang. **56**, 1–15.

Kučera, H., and Francis, W. N. (**1967**). *Computational Analysis of Present-Day American English* (Brown University Press, Providence, RI).

Labov, W. (**1972**). *Sociolinguistic Patterns* (University of Pennsylvania Press, Philadelphia, PA).

Ladefoged, P., and Broadbent, D. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (**1994**). "Training Japanese listeners to identify English /r/ and /l/. III: Long-term retention of new phonetic categories," J. Acoust. Soc. Am. **96**, 2076–2087.

Logan, G. D. (**1988**). "Toward an instance theory of automatization," Psy-

chol. Rev. **95**, 492–527.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–885.

Luce, P. A., and Pisoni, D. D. (**1998**). "Recognizing spoke words. The neighborhood activation model," Ear Hear. **19**, 1–36.

Magnuson, J. S., and Nusbaum, H. C. (**2007**). "Acoustic differences, listener expectations, and the perceptual accommodation of talker variability," J. Exp. Psychol. Hum. Percept. Perform. **33**, 391–409.

McLennan, C. T., and Luce, P. A. (**2005**). "Examining the time course of indexical specificity effects in spoken word recognition," J. Exp. Psychol. Learn. Mem. Cogn. **31**, 306–321.

Mullennix, J. M., Pisoni, D. B. , and Martin, C. S. (**1989**). "Some effects of talker variability on spoken word recognition," J. Acoust. Soc. Am. **85**, 365–378.

Mullennix, J. W., and Pisoni, D. B. (**1990**). "Stimulus variability and processing dependencies in speech perception," Percept. Psychophys. **47**, 379–390.

Munro, M. J. (**1993**). "Production of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings," Lang. Speech **36**, 39–66.

Munro, M. J. (**1998**). "The effects of noise on the intelligibility of foreign-accented speech," Stud. Second Lang. Acquis. **20**, 139–154.

Munro, M. J., and Derwing, T. M. (**1995**). "Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech," Lang. Speech **38**, 289–306.

Munson, B., and Solomon, N. P. (**2004**). "The effect of phonological neighborhood density on vowel articulation," J. Speech Lang. Hear. Res. **47**, 1048–1058.

Norris, D., McQueen, J. M., and Cutler, A. (**2003**). "Perceptual learning in speech," Cognit. Psychol. **47**, 204–238.

Nusbaum, H. C., and Magnuson, J. S. (**1997**). in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego, CA), pp. 109–132.

Nusbaum, H. C., and Morin, T. M. (**1992**). in *Speech Perception, Speech Production, and Linguistic Structure*, edited by Y. Tohkura, Y. Sagisaka, and E. Vatikiotis-Bateson (OHM, Tokyo), pp. 113–134.

Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (**1984**). "Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words," Research on Speech Perception: Progress Report **10**, 357–372.

Nygaard, L. C., Burt, S. A., and Queen, J. S. (**2000**). "Surface form typicality and asymmetric transfer in episodic memory for spoken words," J. Exp. Psychol. Learn. Mem. Cogn. **26**, 1228–1244.

Nygaard, L. C., and Pisoni, D. B. (**1998**). "Talker-specific perceptual learning in spoken word recognition," Percept. Psychophys. **60**, 355–376.

Nygaard, L. C., Sommers, M., and Pisoni, D. B. (**1994**). "Speech perception as a talker-contingent process," Psychol. Sci. **5**, 42–46.

Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (**1993**). "Episodic encoding of voice attributes and recognition memory for spoken words," J. Exp. Psychol. Learn. Mem. Cogn. **19**, 309–328.

Pisoni, D. B. (**1997**) in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullenni (Academic, San Diego, CA), pp. 9–32.

Schmid, P. M., and Yeni-Komshian, G. H. (**1999**). "The effects of speaker accent and target predictability on perception of mispronunciations," J. Speech Lang. Hear. Res. **42**, 56–64.

Schwab, E. C., Nusbaum, H. C., and Pisoni, D. B. (**1985**). "Some effects of training on the perception of synthetic speech," Hum. Factors **27**, 395–408.

Sommers, M. S., and Barcroft, J. (**2007**). "An integrated account of the effects of acoustic variability in first language and second language: Evidence form amplitude, fundamental frequency, and speaking rate variability," Appl. Psycholinguist. **28**, 231–249.

Van Lancker, D., Kreiman, J., and Emmorey, K. (**1985**). "Familiar voice recognition: Patterns and parameters. Part I. Recognition of backward voices," J. Phonetics **13**, 19–38.

van Wijngaarden, S. J., Steeneken, H. J., and Houtgast, T. (**2002**). "Quantifying the intelligibility of speech in noise for non-native listeners," J. Acoust. Soc. Am. **111**, 1906–1916.

Weil, S. A. (**2001**). "Foreign accented speech: Encoding and generalization," J. Acoust. Soc. Am. **109**, 2473 (A).

Yonan, C. A., and Sommers, M. S. (**2000**). "The effects of talker familiarity on spoken word identification in younger and older listeners," Psychol. Aging **15**, 88–99.