Specificity and generalization in perceptual adaptation to accented speech

Jessica E. D. Alexander^{1,a)} and Lynne C. Nygaard²

¹*Department of Psychology, Centenary College of Louisiana, 2911 Centenary Boulevard, Shreveport, Louisiana 71105, USA*

²*Department of Psychology, Emory University, 36 Eagle Row, Atlanta, Georgia 30322, USA*

(Received 22 July 2018; revised 3 May 2019; accepted 10 May 2019; published online 10 June 2019)

The present study investigated the degree to which perceptual adaptation to foreign-accented speech is specific to the regularities in pronunciation associated with a particular accent. Across experiments, the conditions under which generalization of learning did or did not occur were evaluated. In Experiment 1, listeners trained on word-length utterances in Korean-accented English and tested with words produced by the same or different set of Korean-accented speakers. Listeners performed better than untrained controls when tested with novel words from the same or different speakers. In Experiment 2, listeners were trained with Spanish-, Korean-, or mixed-accented speech and transcribed novel words produced by unfamiliar Korean- or Spanish-accented speakers at test. The findings revealed relative specificity of learning. Listeners trained and tested on the same variety of accented speech showed better transcription at test than those trained with a different accent or untrained controls. Performance after mixed-accent training was intermediate. Patterns of errors and analysis of acoustic properties for accented vowels suggested perceptual improvement for regularities arising from each accent, with learning dependent on the relative similarity of linguistic form within and across accents. © 2019 Acoustical Society of America.

<https://doi.org/10.1121/1.5110302>

[TCB]

Pages: 3382–3398

I. INTRODUCTION

A fundamental challenge for accounts of speech perception is how listeners navigate the extensive variation that characterizes our sensory-perceptual experience. Multiple sources of variation influence the way speech is produced. Remarkably, listeners typically achieve perceptual constancy, retrieving linguistic structure despite multiple and overlapping sources of variation.

Listeners may cope with such extensive variation by tracking the systematic properties of that variation during perception. Perceptual learning of the properties that consistently vary in the speech signal allows listeners to dynamically restructure perceptual processing to accommodate informative variation. Previous research has shown that listeners perceptually adapt to voice-specific properties of spoken language, encoding, and using information about talker's voice to facilitate linguistic processing (Dahan *et al.*, 2008; Creel *et al.*, 2008; Creel and Tumlin, 2011; Johnsrude *et al.*, 2013; Levi *et al.*, 2011; Nygaard and Pisoni, 1998; Nygaard *et al.*, 1994; Papesh *et al.*, 2016). Thus, attending to and adapting to indexical characteristics of spoken language appears to tune the speech perceptual system to systematic variation introduced by individual talkers, helping listeners contend with the talker-specific variability present in natural speech.

Listeners also adapt to other sources of variation, such as accents or dialects that arise when groups of speakers share production characteristics. These classes of group variation crosscut variation from individual talkers and

complicate perceptual learning, requiring listeners to track variation associated with individual talkers as well as variation associated with the shared characteristics of the group (Bent and Holt, 2013; Zhang and Holt, 2018; Kleinschmidt and Jaeger, 2015). Foreign-accented speech in particular illustrates this challenge since it includes both the idiosyncratic characteristics associated with an individual talker's voice and the range of phonological characteristics that differ from native productions and are common to the group of accented speakers (Flege *et al.*, 1992; Flege *et al.*, 2003). Despite overlapping sources of variation, listeners' perception of accented speech improves with experience (Adank and Janse, 2010; Clarke and Garrett, 2004; Trude and Brown-Schmidt, 2012) and generalizes to novel speakers with the same accent, producing reliable effects of learning (Bradlow and Bent, 2008; Sidaras *et al.*, 2009; Tzeng *et al.*, 2016; Xie *et al.*, 2018).

Although there is considerable evidence that listeners perceptually adapt to foreign-accented speech, questions remain regarding the range of conditions under which listeners will or will not generalize and the nature of the learning that underlies patterns of specificity and generalization. The current study sought to examine these questions by investigating (1) the extent to which listeners learn and track properties specific to a particular accent, and (2) the range of linguistic properties in natural speech that undergo adjustment during the adaptation process. Listeners were exposed to foreign-accented speech produced by multiple talkers and generalization of learning was tested with a set of either the same or different foreign-accented speakers. Further, as in Baese-Berk *et al.* (2013), a condition in which listeners were

^{a)}Electronic mail: alexander.jed@gmail.com

exposed to multiple foreign-accented speakers was included to determine if learning would be facilitated by exposure to properties that may be characteristic of foreign-accented speech more generally.

With respect to the relative specificity of perceptual adaptation, one possibility is that listeners learn and represent specific properties that systematically vary for a particular accent, shifting their linguistic categories and processing toward particular accented-characteristics of speech. Although specificity of learning could result from several underlying changes in speech processing and categorization (e.g., Francis *et al.*, 2007; Goldinger, 1998; Johnson, 1997, 2006; Kleinschmidt and Jaeger, 2015; Pierrehumbert, 2003), learning to map accented items onto native categories might produce a kind of accent-specific filter for processing types or categories of systematic variation. Generalization to novel speakers with shared characteristics, but not to speakers with different accents (Bradlow and Bent, 2008) would support a specificity account. Indeed, Bradlow and Bent (2008) exposed listeners to sentence-length utterances produced by multiple Mandarin-accented talkers. At test, learning, as indexed by transcription performance, generalized to a novel talker with the same accent, but not to a novel talker with an unfamiliar accent (Slovakian-accented English). This pattern of generalization parallels findings suggesting that listeners generalize within but not across types of dysarthric speech (Liss *et al.*, 2002).

Alternatively, listeners may broaden or adjust their categorization criteria or expectations to incorporate examples of speech that differ in systematic ways from native speech. Adjusting in this way might allow listeners to generalize to unrelated accents because variation in the realization of spoken word forms overlaps with variation along similar dimensions across speakers and across a variety of accents (Baese-Berk *et al.*, 2013; Schmale *et al.*, 2015; van Heugten *et al.*, 2018). In support of the idea that listeners engage in perceptual learning that is not accent-specific, Baese-Berk *et al.* (2013) found better transcription performance for a novel foreign-accented speaker when listeners were trained with sentences produced by a group of foreign-accented speakers with different first languages. Listeners who were exposed to variation in foreign accent generalized learning to a novel speaker with a different, novel accent. These findings suggest that there may be shifts in perceptual processing that generalize across accent type. Listeners may learn characteristics of accented speech that crosscut particular accents, leading to generalization across accents given sufficient exposure to overlapping variation.

Finally, another possibility that is not mutually exclusive is that listeners may globally alter their attentional or cognitive processing strategies to compensate for the atypical or unfamiliar phonological or lexical form encountered in accented speech. Listeners may adopt or alter processing and expectation by shifting, for example, to a greater reliance on lexical, semantic, or contextual information during processing (Hanulíková *et al.*, 2012; Romero-Rivas *et al.*, 2015; Potter and Saffran, 2017).

The current study evaluates these possible mechanisms by replicating and extending previous work addressing the

relative specificity of accent learning, examining the extent to which listeners learn and generalize across different varieties of accented speech produced by multiple accented talkers. In contrast to previous work which assessed specificity of accent learning using sentence-length utterances (e.g., Baese-Berk *et al.*, 2013), the current study focused on learning from individual spoken words and examined accent learning across two different accent categories. By focusing on learning from spoken words, the aim was both to evaluate specificity of learning specifically for systematic *acoustic-phonetic* variation in accented speech and to evaluate if lexical characteristics of spoken words influenced learning and generalization.

Although talker-specific and talker-independent accent learning have been demonstrated for both individual words and sentence-length utterances (e.g., Baese-Berk *et al.*, 2013; Bradlow and Bent, 2008; Nygaard *et al.*, 1994; Sidaras *et al.*, 2009; Maye *et al.*, 2008; Xie and Myers, 2017), there is some evidence that what is being learned may differ across utterance lengths or types (e.g., Nygaard and Pisoni, 1998). For example, Sidaras *et al.* (2009) trained and tested native English-speaking listeners with both sentences and words produced by Spanish-accented talkers and found evidence for perceptual learning with both types of stimuli. However, Nygaard and Pisoni (1998) found that talker-specific learning of native speech, at least, may be dependent on a match between utterance type (sentences vs words) from exposure to test. When listeners were exposed to a set of voices from sentence-length utterances, recovery of linguistic structure was better for sentences produced by familiar than by unfamiliar talkers, but this advantage did not extend to word-length stimuli produced by familiar talkers. When exposed to word-length stimuli during training, learning generalized to words produced by familiar talkers, suggesting that the nature of the talker-specific learning depended on the type of materials presented during training. Thus, when listeners were trained with sentence-length utterances as in Baese-Berk *et al.* (2013), they may have adapted to properties of accented speech gleaned from multiple levels of linguistic and non-linguistic structure. In addition to lower-level phonological information, sentences contain prosodic characteristics, including temporal properties, that vary as a function of accent, as well as semantic and syntactic information that could constrain the recovery of the linguistic structure of the utterance. Training listeners with sentences makes it difficult to ascertain what aspects of the accented speech listeners may be learning. Using word-length utterances allowed us to assess specificity of learning at the segmental level and also allowed us to manipulate the lexical characteristics of our stimuli. The word stimuli differed in word frequency and neighborhood density (Luce and Pisoni, 1998) in order to determine if lexical factors influenced either the extent of or relative specificity of learning.

In addition, generalization of learning was evaluated within and across two different target accents, Spanish- and Korean-accented speech, and generalization was assessed to multiple rather than single talkers at test. The aim was to evaluate whether relative degree of specificity was similar across different accent types and to determine if learning

was independent of the characteristics of a single test speaker. Previous research has generally examined learning of accented speech by assessing the benefits of different types of exposure conditions on the intelligibility (e.g., Bradlow and Bent, 2008) or processing (e.g., Xie *et al.*, 2018) of utterances from a single target accented speaker at test. By examining generalization of learning to multiple accented talkers at test and by comparing generalization of learning to both Spanish- and Korean-accented speech within the same experiment, we were able to minimize within-test adaptation to particular talker-specific characteristics, maximize our ability to assess talker-independent learning, and evaluate patterns of learning and generalization across different types of variation.

Spanish- and Korean-accented speech was chosen for two reasons. First, in previous research, we found perceptual adaptation and generalization to novel words produced by unfamiliar talkers of Spanish-accented speech (Sidasar *et al.*, 2009). Using both types of foreign-accented speech provided an opportunity for replication and a basis to explore the specificity question. Second, both Spanish and Korean differ from English in phonological and acoustic-phonetic characteristics and these L1 differences seem to carry over into the accented-speech of the L2. For example, Spanish-accented speech systematically differs from native English speech in temporal and spectral characteristics of vowels (Flege *et al.*, 1992; Sidasar *et al.*, 2009), and the equivalents of the English vowels /æ/, /ʌ/, /ɔ/, /ʊ/, or /ɪ/ are not found in Spanish (Spanish vowels include /i e a o u/; Nash, 1977). Spanish also differs from English in consonant characteristics such as voice onset time (VOT; Nash, 1977; Flege *et al.*, 2003).

Like Spanish, Korean vowels differ from native English averages and contribute to the realization of Korean-accented English. For example, Korean-accented speech contains fewer differences in vowel duration than native English (Lee *et al.*, 2006). Some Korean vowels map easily onto English vowels (/i e ε a ʌ o u/) in F1/F2 space, but there are English vowels, which do not have a clear Korean equivalent (e.g., /ɪ æ ʊ ɔ/; Yang, 1996). Korean differs from both English and Spanish in consonant pronunciation as well, exhibiting differences in place and manner of articulation (Nissen *et al.*, 2004; Rice, 2002), and these characteristics are found in Korean-accented speech as well (Nissen, *et al.*, 2004; Tsukada *et al.*, 2004). These differences among native English, Korean-accented, and Spanish-accented speech produce a distinct set of systematic regularities that listeners may use during perceptual adaptation. Assuming listeners generalize from one accent to another to the degree that the specific acoustic-phonetic properties are similar, we hypothesized that if perceptual adaptation is specific to the accented properties present during exposure, we would see little generalization across accent groups.

Generalization of learning was also evaluated for listeners trained with a mixed group of accented speakers. Here, the inventory of accent characteristics was assumed to be more variable and might be more likely to overlap with either the Spanish- or Korean-accented properties at test. Thus, we predicted that some generalization of learning might be observed with mixed accented training depending

on the type of accented speech at test (Spanish versus Korean). However, if listeners are generally adjusting the type of perceptual processing (e.g., reliance on context) or tuning to general properties of accented speech (e.g., slower speaking rate), then we expected benefits of accented exposure to generalize regardless of condition.

II. EXPERIMENT 1

Experiment 1 replicated and extended research on perceptual adaptation to foreign-accented English by evaluating the extent to which listeners would perceptually adapt to Korean-accented speech from word-length utterances varying in lexical characteristics and produced by the same versus different talkers as heard during training, in preparation for assessing specificity of accent learning in Experiment 2. Lexical characteristics were varied such that training and test stimuli consisted of *easy* and *hard* words, which differ in neighborhood density and word frequency (Luce and Pisoni, 1998; Vitevitch *et al.*, 1999). Easy words are high frequency words with relatively few, low frequency neighbors (e.g., *shirk*, *lurk*, and *murk*). Hard words are low frequency words with many high frequency neighbors (e.g., *beat*, *lead*, and *feed*). Lexical characteristics such as word frequency and neighborhood density have been shown to affect the time course and accuracy of spoken word recognition (Bradlow and Pisoni, 1999; Magnuson *et al.*, 2007; Remez *et al.*, 2011).

Overall, listeners were expected to transcribe Korean-accented speech more accurately than untrained controls who were unfamiliar with the accent. We predicted that listeners would learn systematic segmental properties of the accented utterances from exposure to word-length stimuli. Indeed, in order for improved intelligibility to result from training with words, listeners would need to modify the mapping from the acoustic realizations in the speech signal to linguistic representations of particular segments.

With respect to performance for familiar and unfamiliar talkers, we hypothesized that learning would generalize to *all* accented talkers at test. Although accent learning has been hypothesized to stem from talker-specific learning, with immediate generalization of learning found primarily for specific similar talkers (Xie and Myers, 2017), talker-independent perceptual adaptation has been shown for accented speech (Bradlow and Bent, 2008; Sidasar *et al.*, 2009; Tzeng *et al.*, 2016; Xie *et al.*, 2018). If perceptual adaptation relies on the similarity between properties of particular talkers' idiolect from study to test (see Xie and Myers, 2017), performance for familiar talkers at test should be better than for unfamiliar talkers. If listeners are adapting to the systematic properties that characterize the particular accent, then performance for both familiar and unfamiliar talkers should be better relative to untrained controls.

Finally, with respect to lexical characteristics, we hypothesized that hard words with many high frequency neighbors would be more difficult to distinguish from other words, ultimately requiring more detailed processing of acoustic-phonetic structure (or even more experience or repetition) in order to correctly map accented segments to existing acoustic-phonetic categories. If so, listeners may show

differential learning for *easy* and *hard* words (as an interaction between training condition and word type). Listeners trained with accented speech could show increased performance for easy words relative to controls because smaller adjustments might be needed to demonstrate differentiation from other lexical items. Alternatively, we might observe the largest difference in performance as a result of training for *hard* words at test as listeners fine-tune linguistic category structure for accented speech.

Lexical characteristics may impact degree of perceptual learning in other ways as well. High frequency words may have more robust lexical representations (e.g., stronger priors in ideal adapter frameworks, Kleinschmidt and Jaeger, 2015, or more exemplars in exemplar-based lexicon frameworks, Goldinger, 1998) than low frequency words and therefore, show less influence of talker- or accent-specific learning. Indeed, talker-specific effects have been shown to be most robust for low frequency words (Goldinger, 1998) and prior exposure through frequency of occurrence of specific lexical forms may impact the specificity of learning. Finally, of course, it may also be the case that properties of frequency and neighborhood density will not interact with accented training, suggesting that perceptual adaptation to accented pronunciations may not differentially impact word recognition processes.

A. Method

1. Participants

Seventy-two native speakers of American English with no reported history of hearing or speech disorders participated and received \$15 or course credit for participation. Participants were excluded if they reported experience with Korean as a language in their home or if they reported frequent exposure to Korean-accented speech.

2. Materials

Twelve native Korean speakers (six males, six females) from Seoul, South Korea living in the Atlanta area were recorded reading 144 monosyllabic words (72 *easy*, 72 *hard*; from Luce and Pisoni, 1998). The words differed on characteristics of neighborhood density and word frequency but were all highly familiar (Nusbaum *et al.*, 1984). Speakers were between the ages of 20 and 37, with the mean age of 26.7 years at the time of recording. To ensure the native Korean speakers spoke a similar dialect, all 12 speakers were born in South Korea and lived in Seoul before coming to the US. They all began learning English at a mean age of 13 years, with a range from 10 to 15 years of age. Their mean age of arrival in the US was 24.3, and the range of age of arrival was from 15 to 35 years of age. All stimuli were recorded in a sound-attenuated room with a SONY Digital Audio Tape-corder TCD-D7. The recordings were re-digitized on an iMac and edited for presentation using Sound Studio 3.

Baseline intelligibility was obtained for all speakers. A separate group of 120 participants (ten per speaker) who were unfamiliar with the accent transcribed all 144 recorded

words. Intelligibility was calculated as the proportion words correct. The mean intelligibility was 0.57 (proportion words correct). Two groups of six speakers (three males, three females) equated for intelligibility were constructed to create counterbalances (see Table 1 for speaker intelligibility for experiments 1 and 2). There were no significant intelligibility differences between groups, $t(10) = 0.024$, $p = 0.98$.

3. Procedure

Participants were trained in one session with words spoken by six speakers (three males, three females) and were tested with new words from either the same speakers or six new Korean-accented speakers (three males, three females) from the other group.

a. Training phase. Training entailed four Comparison blocks and three Variability blocks, alternating. In Comparison blocks, listeners heard each accented word twice (six per speaker) and rated accentedness, from non- to heavily-accented (1–7). In Variability blocks, participants transcribed each word (24 per speaker, repeated once with new word/speaker pairings), then saw the correct answer and heard the word again. Half of the training stimuli were hard words and half easy words. Sidaras *et al.* (2009) used Variability and Comparison blocks to deliver high-variability training with items and speakers varying trial-to-trial as well as training grouped by words so that listeners might compare different speakers producing the same item. The use of two learning environments ultimately led to robust learning (for the unique contribution of training structure to perceptual learning, see Tzeng *et al.*, 2016). Training lasted for approximately 35–40 min.

b. Generalization test. At test, participants transcribed 48 randomly-presented novel words (half easy and half hard words) spoken by either the same six speakers as heard during training or by six new Korean-accented speakers and received no feedback. A no training control group completed only the test portion of the experiment.¹ Speaker groups were counterbalanced across listeners. There were 24 listeners in each training condition (same, different, control).

B. Results and discussion

Logistic mixed-effects models were used to assess the effects of training condition on test transcription performance using the lme4 package (Bates *et al.*, 2015) in R. *Post hoc* tests were completed using the multcomp package (Hothorn *et al.*, 2008). Accuracy of word transcription was the dependent variable with misspellings and homophone spellings counted as correct. Random effects of subject and words were included in the model. We added fixed effects of word type (easy/hard), training condition (same speakers, different speakers, and no training controls), and the interaction term, stepping forward and comparing model fit using log-likelihood ratio tests (Baayen *et al.*, 2008). Training conditions were dummy coded, with controls as the reference level.

TABLE I. Mean Intelligibility (% correct) for Korean-, Spanish-, and Mixed-accented Talkers.

Speaker Group	Female Speakers		Male Speakers	
	Place of Origin (first language)	Intelligibility	Place of Origin (first language)	Intelligibility
Korean 1	Seoul, South Korea (Korean)	53.3	Seoul, South Korea (Korean)	59.2
		58.3		67.5
		64.3		81
Korean 2	Seoul, South Korea (Korean)	68.5	Seoul, South Korea (Korean)	68.5
		60.1		60.1
		56.8		56.8
Spanish 1	Mexico City, Mexico (Spanish)	32.9	Mexico City, Mexico (Spanish)	54.5
		39.7		58.3
		68.8		49.2
Spanish 2	Mexico City, Mexico (Spanish)	48.9	Mexico City, Mexico (Spanish)	42.9
		35.2		60.3
		53.5		52.8
Mixed 1	Korce, Albania (Albanian)	57.7	Bucharest, Romania (Romanian)	76.6
	Vlaadingen, Holland (Dutch)	64.7	Chittagong, Bangladesh (Bengali)	88.9
	Shizuoka, Japan (Japanese)	79.5	Sagar, India (Hindi)	70.2
Mixed 2	Alès, France (French)	73.7	Kazan, Russia (Russian)	90.9
	Bamberg, Germany (German)	72.5	Beijing, China (Mandarin)	53.7
	Mogadishu, Somalia (Somali)	92.7	Izmir, Turkey (Turkish)	72.2

The final model included word type, $\chi^2(3)=26.89$, $p < 0.001$, and training condition, $\chi^2(4)=12.86$, $p=0.012$, as fixed factors, both of which significantly improved model fit, suggesting that transcription accuracy differed across conditions and word type (see Fig. 1 for a comparison with Sidas et al., 2009). Including the interaction of training condition and word type did not significantly improve model fit, $\chi^2(2)=0.029$, $p=0.99$. The significant effect of word type suggests that listeners were significantly more accurate transcribing easy than hard words, $\beta = -1.59$, $z = -5.34$, $p < 0.001$. In training group comparisons, listeners trained and tested with the same speakers performed significantly better than controls, $\beta = 0.30$, $z = 2.77$, $p = 0.011$, as did those trained and tested with different speakers, $\beta = 0.39$, $z = 3.55$, $p = 0.001$. There was no significant difference between listeners trained with same and different speakers, $\beta = -0.09$, $z = -0.824$, $p = 0.41$. See Table II for means and standard deviations at test.

FIG. 1. (a) *Exp 1. Korean-accented test*: Proportion correct transcription accuracy at test for Same Voices, Different Voices, or No-Training. (b) *Spanish-accented test*: Proportion correct transcription accuracy at test for Same Voices, Different Voices, or No-Training for Spanish-accented words. Data reproduced from Sidas et al. (2009). Error bars represent one standard error of the mean.

These findings indicate that very brief training with Korean-accented speech allowed listeners to perceptually adapt to the systematic variation present in Korean-accented English. This adaptation facilitated transcription at test for Korean-accented novel words and novel talkers, indicating that listeners learned systematic segmental properties of Korean-accented English to better extract the linguistic structure from the word-length utterances of novel talkers with that same accent.

Regarding lexical properties, although easy words were more accurately transcribed than hard words, there was little evidence of an interaction between the lexical properties of words and the perceptual learning process, at least in this brief training paradigm. This may indicate perceptual adaptation, even with such brief training, was sufficiently specific or detailed to benefit performance for both easy and hard words. Alternatively, lexical characteristics such as neighborhood density or word frequency could have more gradient effects on perceptual learning, which were not captured by the particular task (e.g., transcription) or by the manipulation of lexical characteristics used in this experiment. Nevertheless, since learning generalized to novel words as well as novel speakers, perceptual adaptation appeared to be occurring at the sublexical level, perhaps influencing the mapping from the acoustic speech signal to segmental

TABLE II. Experiment 1: Means and standard deviations for proportion words correct at generalization test for Korean-accented speech.

Test Condition	Easy Words		Hard Words		All Words	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Same Voices	0.821	0.092	0.634	0.102	0.727	0.074
Different Voices	0.851	0.088	0.609	0.086	0.730	0.050
No Training	0.799	0.090	0.563	0.094	0.681	0.066

representation rather than influencing the dynamics and resolution of lexical processing.

Similarly, little evidence was found for the presence of talker-specific learning during training (in contrast to Xie and Myers, 2017). Listeners tested with the same speakers from training performed no differently than those tested with different speakers. Presumably if listeners were learning the very particular way in which certain accented speakers produced the English targets, then performance with the same speakers would be superior to that of different speakers. However, perhaps the lack of talker-specific learning is not surprising given that listeners received relatively more experience with the Korean accent overall than with any particular speaker. That is, during training, listeners heard multiple examples of Korean-accented speech, but relatively few tokens of *any particular* talker's utterances. Thus, the structure of training may have facilitated comparison *across* talkers for the extraction of speaker-independent properties (Tzeng *et al.*, 2016) rather than providing sufficient experience with a particular talker to facilitate the extraction of talker-specific properties. This type of highly variable training may have highlighted properties of the Korean-accent that were similar across all the speakers rather than promoting registration of the particular characteristics of any one individual speaker.

These findings replicate the perceptual adaptation to Spanish-accented words found in previous studies (Sidasar *et al.*, 2009), illustrating that listeners can perceptually adapt to accented speech from both Spanish- and Korean-accented speakers.

III. EXPERIMENT 2

Experiment 2 directly examined the nature of the perceptual adaptation process during exposure to accented speech by assessing specificity of learning within and across distinct accents. Listeners were trained with either Spanish-accented speech, Korean-accented speech, or speech from multiple first language groups and were tested with either Korean- or Spanish-accented speech to address whether learning with one accent generalizes to a novel accent or whether learning is specific to the accent presented during training.

The inclusion of a condition in which listeners were trained with a mixture of accents (Mixed-accent group) was designed to determine if exposure to extensive variation in phonological word form generally would facilitate adaptation to accented speech. The Mixed-accent speaker groups consisted of accented speech from twelve different native languages drawn from a variety of language families. There should be less *systematic* and *consistent* overlap in the acoustic-phonetic properties of these speakers' accented productions of English words, allowing for a strong test of the notion that non-specific, high-variability exposure might lead to better performance for accented speech. Indeed, Baese-Berk *et al.* (2013) found that exposure to sentence-length utterances from multiple foreign-accented speakers improved intelligibility of a talker with a different foreign accent. Listeners in their study generalized learning from

high variability, multiple accent exposure to novel utterances produced by a foreign-accented speaker with a different L1. Baese-Berk *et al.* (2013) concluded that listeners learned properties of accented speech that were similar and systematic *across* accents. In the current experiment, we sought to build upon this finding by evaluating whether listeners may also learn properties that are specific to a particular accent group and specific to the segmental realization of a particular accent. Thus, unlike in Baese-Berk *et al.* (2013), listeners in the current experiment were trained and tested with word-length utterances in order to determine if listeners can learn accent-specific, talker-independent *segmental* properties of accented speech. By limiting our training materials to word-length utterances, we hoped to isolate exposure to variation in acoustic-phonetic segmental properties.

To assess learning, listeners were presented with novel utterances at test that were produced by multiple speakers of a single accent. The purpose was to both provide a strong test of generalization by presenting multiple novel talkers at test and to ensure that the characteristics of the test items remained constant across training conditions so that any inadvertent contribution of a particular test talker or test item would be consistent across training conditions. Thus, the characteristics of the test remained constant, with only the properties of the training changing across conditions (Spanish-, Korean-, Mixed-accent and no-training training conditions).

We hypothesized that if listeners are learning the specific and systematic properties of the accent group during training, there should be little transfer across accent groups. Consistent with previous work examining learning from accented sentences (Bradlow and Bent, 2008), listeners should only outperform controls when trained and tested on the same accent. However, if perceptual learning generalizes (e.g., listeners are generally shifting attention or processing strategies, relaxing their categorization criteria), then any type of accent training should generalize at test and show benefits relative to untrained controls. Finally, if high-variability exposure to multiple accents during training leads to generalization, promoting learning of properties that are general to all types of accented speech, then mixed-accent training should result in better performance than either training with a single different accent or with no accent training at all (Baese-Berk *et al.*, 2013; Schmale *et al.*, 2015; van Heugten *et al.*, 2018).

A. Method

1. Participants

The same participant criteria and incentives were used in Experiment 2. Listeners were excluded if they reported experience as a language in their home or if they report frequent exposure with any of the first languages of the speakers used in Experiment 2. 160 listeners participated.

2. Materials

The Korean speaker stimuli used in Experiment 1 were also used in Experiment 2, along with two other sets of

accented stimuli. One set consisted of the Spanish-accented stimuli from [Sidas et al. \(2009\)](#). Twelve native Spanish speakers from Mexico City living in the Atlanta area were recorded reading the list of 144 (72 easy and 72 hard) words. The native Spanish speakers had a mean age of 32.75 years at the time of recording, with a range of 26–39 years of age. Their mean age of arrival in the US was 26.42 years, with a range of 21–34 years of age, and they had begun speaking English at approximately 16.67 years of age on average, with a range of 2–28 years of age. Spanish-accented stimuli were recorded in a sound-attenuated room with a SONY Digital Audio Tape-corder TCD-D7. The recordings were re-digitized on an iMac and edited for presentation using Sound Studio 3.

A third group of speakers with a variety of first languages (Mixed-accented speakers) were also recorded reading the same set of 144 easy and hard words. These speakers were all members of the Emory University community and were living in Atlanta. The first languages of the speakers were from a variety of language families. They included speakers of Albanian, Dutch, Japanese, Romanian, Bengali, Hindi, French, German, Somali, Russian, Mandarin, and Turkish (see [Table I](#) for place of origin information). They were recorded in the same manner as the Spanish-accented speakers. The Mixed accent speakers had a mean age of 25.09 years at the time of recording, with a range of 19–37 years of age. Their mean age of arrival in the US was 19.91 years, with a range of 3–31 years of age, and they had begun speaking English at approximately 10.18 years of age on average, with a range of 3–16 years of age.

A separate set of 120 participants (ten for each speaker) transcribed all 144 recorded words for baseline intelligibility. A one-way analysis of variance (ANOVA) showed a significant difference in intelligibility among groups, $F(2, 35) = 13.05, p = 0.001$. Spanish-accented speakers ($M = 49.75$) were significantly less intelligible than Korean-accented speakers ($M = 64.80; p = 0.004$), and than Mixed-accent speakers ($M = 74.44; p < 0.001$). Korean-accented speakers were marginally less intelligible than Mixed-accent speakers, $p = 0.056$. These intelligibility differences are likely a factor of individual speakers' English proficiency and potentially related to factors such as education, income, and age of acquiring English ([Ingvalson et al., 2011; Piske, 2012](#)).

Two groups of six speakers (three males, three females) were constructed for each accent group to create counterbalances. For the Mixed-accent groups, speakers were assigned to groups so that there were languages from multiple language families in each counterbalance group. For all groups, we attempted to equate the intelligibility of the set of speakers in each counterbalance group. There were no significant intelligibility differences between speaker groups for each accent [[Korean, \$t\(10\) = 0.024, p = 0.98\$; Spanish, \$t\(10\) = 2.23, p = 0.56\$; Mixed, \$t\(10\) = 0.41, p = 0.69\$ \]. See \[Table I\]\(#\) for intelligibility information for all speakers.](#)

3. Procedure

The same training paradigm was used as in Experiment 1, but listeners were trained with the set of six speakers from

one of the accent conditions (Korean, Spanish, or Mixed) and were tested with new words from six new speakers. The accent during training varied by condition, and the test contained either Spanish- or Korean-accented speech. All listeners were presented with unfamiliar talkers at test to address talker-independent generalization to novel accents. For the Same-accent conditions, listeners were trained on the test accent, and for the Different, Mixed, and No-Training conditions, listeners had not received training on the test accent. The talker group within each accent was fully counterbalanced across listeners. There were 20 listeners in each training condition (same accent, different accent, mixed accent, and no training) for each test accent (Spanish or Korean).

B. Results and discussion

Logistic mixed-effects models were used as in Experiment 1. Accuracy of word transcription was the dependent variable. Subjects and words were entered as random intercepts. [Table III](#) reports means and standard deviations. Fixed effects of test accent (Korean or Spanish), training condition (same accent, different accent, mixed accent, or no training), and word type (easy and hard), were entered into the model, stepping forward to find the best model fit. The final model included fixed effects of test accent, $\chi^2(1) = 216.67, p < 0.001$, training condition, $\chi^2(3) = 11.92, p = 0.008$, and word type, $\chi^2(1) = 9.32, p = 0.002$, and random intercepts for subjects and words. Model fit did not significantly improve when interaction terms were entered for test accent by word type, $\chi^2(1) = 1.53, p = 0.217$, test accent by training condition, $\chi^2(3) = 3.85, p = 0.278$, word type by training condition, $\chi^2(3) = 6.43, p = 0.093$, or the three-way interaction, $\chi^2(10) = 13.06, p = 0.220$. Including any random slopes prevented the model from converging. This model indicates that there were main effects of test accent, training condition, and word type, and that there were no significant interactions among variables. Listeners transcribed Korean-accented test items significantly more accurately than Spanish-accented test items, $\beta = -1.20, z = -20.01, p < 0.001$, and they transcribed easy words more accurately than hard words, $\beta = -1.63, z = -3.19, p = 0.001$. To examine the main effect of training condition, we performed all pairwise comparisons, with a Holm-Bonferroni correction. Listeners trained and tested on the same accents performed significantly better than untrained controls, $\beta = 0.253, z = 3.06, p = 0.013$, and better than listeners trained and tested with different accents, $\beta = 0.236, z = 2.80$,

TABLE III. Experiment 2: Means and standard deviations for proportion words correct at generalization tests by training condition.

Test Accent	Training Condition	Easy Words		Hard Words		All Words	
		M	SD	M	SD	M	SD
Korean	Same Accent	0.800	0.055	0.621	0.109	0.710	0.067
	Different Accent	0.752	0.075	0.550	0.129	0.651	0.083
	Mixed Training	0.758	0.066	0.575	0.105	0.667	0.079
	No Training	0.781	0.066	0.540	0.102	0.660	0.045
Spanish	Same Accent	0.602	0.063	0.390	0.105	0.496	0.037
	Different Accent	0.577	0.088	0.379	0.099	0.478	0.045
	Mixed Training	0.613	0.087	0.377	0.107	0.495	0.057
	No Training	0.592	0.063	0.319	0.103	0.455	0.049

FIG. 2. (a) *Exp 2. Korean-accented test*: Proportion correct transcription accuracy at test for Same-Accent, Different-Accent, Mixed-Accent, or No-Training. (b) *Exp 2. Spanish-accented test*: Proportion correct transcription accuracy at test for Same-Accent, Different-Accent, Mixed-Accent, or No-Training. Error bars represent one standard error of the mean.

$p = 0.026$ (see Fig. 2). No other pairwise comparisons were significant, all p 's > 0.230 .

These findings provide evidence for both accent specificity and generalization in perceptual learning of systematic variation. Listeners appeared sensitive to the particular properties of the Korean-accented as distinct from Spanish-accented utterances with superior performance when accent matched from the training to test phase of the experiment. Listeners appeared to derive no benefit from exposure to multiple speakers of an accent that was different from the accent presented at test. This specificity in adaptation to a variety of accented speech is consistent with evidence for specificity of perceptual learning in a variety of other domains (Ahissar and Hochstein, 1997; Eisner and McQueen, 2005; Kraljic and Samuel, 2007; Nygaard and Pisoni, 1998) and consistent with previous demonstrations of specificity of learning foreign-accented speech (Bradlow and Bent, 2008). The current findings extend previous demonstrations to word-length utterances suggesting that listeners were specifically tracking segmental properties of each accent and perhaps modifying the acoustic-phonetic mappings. In addition, the extent of perceptual learning did not differ as a function of lexical characteristics suggesting signal-dependent modifications rather than changes to the process of lexical competition. Indeed, whatever learning occurred during training, any tuning to accent-specific acoustic-phonetic detail benefited both words with few and many similar sounding neighbors. Finally, listeners did not appear to rely on global shifts in attention or linguistic processing strategies when encountering any type of accented speech but rather appeared to engage in perceptual learning of acoustic-phonetic features characteristic of the particular set of accented materials.

That performance for the mixed accent training group did not differ significantly in our analyses from either untrained controls or from the same accent training conditions suggests that listeners derived some benefit from exposure to multiple accented speakers, but that that benefit was weak at best for these types of stimuli. Although this evidence is certainly not definitive, it suggests that listeners were not generally shifting categorization criteria or processing strategies to include any type of accented realizations of

English speech sounds. One possibility for this pattern is that listeners may have been tuning their acoustic-phonetic processing to include the particular or specific features of the speech they encountered during training and generalized to the extent that that tuning overlapped with and were similar to accented features encountered at test, at least in the context of this learning paradigm (see Baese-Berk *et al.*, 2013). Across types of variation, materials, and methodologies, perceptual learning appears to extend to items that vary in ways that are consistent with training stimuli. The current findings suggest that this type of perceptual adaptation extends to *classes* of predictable variation that may crosscut speakers and styles of speech in spoken language and that listeners are sensitive to the particular range of segmental variation encountered during exposure.

C. Supplemental analyses

In order to further examine the possibility that listeners were tuning or adapting to particular segmental realizations that crosscut accents, we conducted a set of supplemental, exploratory analyses to determine what sets of properties might have been available to listeners as a function of mixed training. Since the counterbalance groups in the mixed training condition contained different L1 accents, we were interested in whether exposure to the particular sets of accented speakers in the Mixed accent counterbalance groups may have resulted in different patterns of adaptation. To that end, we compared the Mixed-accent training groups with the No Training controls for Spanish- and Korean-accented tests. As a reminder, listeners in Mixed group 1 were trained with speakers of Albanian, Dutch, Japanese, Romanian, Bengali, and Hindi. Listeners in Mixed group 2 were trained with speakers of French, German, Somali, Russian, Mandarin, and Turkish. If exposure to Mixed accent speakers generalized to test in only some conditions, the intermediate effect we saw in the overall analysis may be explained by generalization across particular similarities between speech from the Mixed accent speakers and test stimuli. However, if training with either of the counterbalanced mixed speaker groups generalized to both Spanish- and Korean-accented tests, it would provide stronger evidence for accent-independent generalization of learning.

The initial logistics mixed effects model included Test accent (Spanish or Korean) and training groups (Mixed group 1, Mixed group 2, No training) as fixed effects, as well as the interaction term between the training and test conditions. Subjects and words were entered as random intercepts. Including the interaction term caused convergence errors. Since the pattern of results might differ for Spanish- and Korean-accented tests and the interaction term could not be included in the model, the test accents were analyzed in separate models. For Spanish-accented tests, training condition (Mixed group 1, Mixed group 2, No training) was entered into the model as a fixed effect, with subjects and words as random effects. Model fit significantly improved after adding the fixed effect of training group to the random effects intercepts, $\chi^2(2) = 8.97$, $p = 0.011$, suggesting that transcription accuracy differed significantly by training group. There was no significant difference

between the performance of listeners trained with Mixed group 1 [$M = 0.475$, standard deviation (SD) = 0.058] and untrained controls ($M = 0.455$, $SD = 0.049$), $\beta = -0.17$, $z = -1.09$, $p = 0.27$, but listeners trained with Mixed group 2 ($M = 0.519$, $SD = 0.047$) performed significantly better than controls, $\beta = -0.52$, $z = -3.17$, $p = 0.005$. There was also no significant difference between the performance of listeners trained with Mixed group 1 and Mixed group 2, $\beta = 0.35$, $z = 1.93$, $p = 0.11$. Thus, at least relative to control subjects, listeners exposed to one set of mixed-accent speakers showed generalization to test, and the other group did not show clear generalization.

For Korean-accented tests, training condition (Mixed group 1, Mixed group 2, No training) was entered into the model as a fixed effect, with subjects and words as random effects. Model fit did not significantly improve after adding the fixed effect of training group to the random effects intercepts, $\chi^2(2) = 0.20$, $p = 0.91$, indicating that transcription accuracy did not differ between Mixed group 1 ($M = 0.662$, $SD = 0.076$), Mixed group 2 ($M = 0.671$, $SD = 0.085$), and no training controls ($M = 0.660$, $SD = 0.045$). Learning did not seem to generalize for the Mixed groups in Korean-accented speech.

For Spanish-accented tests, one mixed training group including accented speech from French, German, Somali, Russian, Mandarin, and Turkish talkers produced improved test performance, helping to explain the intermediate results in the overall analysis for the mixed accent training conditions. Although these analyses were exploratory and hence need to be interpreted with caution, the differential pattern of results for Spanish- versus Korean-accented words and for the two mixed speaker groups suggest that listeners may have generalized from overlapping properties of the accented speech presented during training with mixed accent group 1 to the Spanish-accented speech presented at test. If there was simply a general broadening of acoustic-phonetic categories or a shift in processing strategy with mixed accent training, generalization should have been found across all mixed-accent conditions relative to test and strongly for both test accents.

Error and acoustic analyses in the next sections attempt to identify possible candidates for segmental attunement that might underlie our hypothesized overlap in training to test features. We chose to focus on the effects of training condition on the identification of vowels in order to determine if patterns of confusions among particular vowels or areas of the vowel space could be predicted by acoustic measures of vowel realization across accent groups (see Sidaras *et al.*, 2009).

1. Error analyses

To further examine the nature of learning, we analyzed the identification of vowels from the word transcriptions of listeners in experiment 2. We examined correct vowel identification for each of the training conditions in response to Spanish or Korean accented target vowels. Correct identification of the intended vowels as well as the distribution of

errors were examined.² The analyses included target words that contained the vowels /i/, /ɪ/, /e/, /æ/, /ʌ/, and /a/.

The percent identification of target vowels for Spanish and Korean accented speech are presented in a confusion matrix (see Table IV) representing target vowels and response choices for both trained and untrained listeners. In general, for Spanish-accented speech, as in Sidaras *et al.* (2009), the high front vowels /i/ and /ɪ/ were confusable, while /e/ was readily identified. The low back vowels /æ/, /ʌ/, and /a/ were confusable for both untrained and trained listeners. For Korean-accented speech, /i/ and /e/ were readily identified. The vowel /ɪ/ was much more confusable than the other high front vowels. The vowels /æ/ and /ʌ/ were moderately confusable, while /a/ was more readily identified.

To examine how vowel identification differed as a function of training for each type of accented speech, we compared vowel identification performance between listeners presented with the same, different, and mixed accent from study to test and no training controls for each test accent (Spanish-accented and Korean-accented). Although we only examined a subset of English vowels (and adaptation in the overall learning task was likely driven by properties of both consonants and vowels), the error analyses show interesting patterns across test accents and training conditions. The patterns of accuracy and errors across vowels suggest that the more readily identified, less confusable vowels, such as /æ/, /a/, and /e/, may be acoustically more distinct across foreign-accented and English-accented speech, allowing even the untrained listeners to differentiate them from the other vowels. However, more confusable vowels could be expected to be acoustically similar in a particular accent. More specifically, the patterns of differences based on training show an interesting mix of training effects from same, different, and mixed accent training. These varying effects show evidence of adaptation from one accent to another for specific vowels and for particular areas of vowel space. First, patterns of confusions and the impact of training condition differed for the Spanish-accented and Korean-accented vowels. Second, in general, training on the same accent led to improvement in vowel identification relative to training with a different accent group with improvement for particular vowels depended on the conjunction of training accent and test accent. Third, generalization across conditions was found as well. For example, identification of the Korean-accented vowel /ɪ/ improved with Spanish-accented training, suggesting that exposure to particular kinds of variation benefitted listeners at test regardless of accent. Taken together, these patterns of identification performance suggest that listeners were perceptually adapting to specific types of segmental variation rather than to an overall pattern of systematic change. To the extent that training variation overlapped with test variation, evidence for perceptual adaptation and generalization was found. The learning appeared to take place at a vowel by vowel or dimension by dimension level, providing a mosaic benefit to the listener. Taken together with the evidence from the followup analyses of the mixed accent training groups, it seems that adaptation across accents may occur based on the specific properties of the speech being encountered.

TABLE IV. Confusion matrix of vowel percent responses based on word transcription performance. Note: Cells represent percent responses to target.

Spanish Accented Test															
No Accented Training								Same/Spanish Accented Training							
Intended Targets	Listeners' Responses							Intended Targets	Listeners' Responses						
	/i/	/ɪ/	/e/	/æ/	/ʌ/	/a/	Other		/i/	/ɪ/	/e/	/æ/	/ʌ/	/a/	Other
/i/	43.3	44.2	5.8	0.8			5.8	/i/	41.7	46.7	6.7				5.0
/ɪ/	27.5	59.2	8.3	0.8			4.2	/ɪ/	20.8	70.8	8.3				
/e/			95.8	0.8			3.3	/e/	0.8	0.8	95.8				2.5
/æ/		1.7		73.3	18.3	5.0	1.7	/æ/		1.7		85.0	10.0	1.7	
/ʌ/		1.4	0.7	7.1	38.6	24.3	27.9	/ʌ/				6.4	44.3	22.9	26.4
/a/	11.7			1.7	13.3	63.3	10.0	/a/	3.3				11.7	80.0	5.0
Different/Korean Accented Training								Mixed Accented Training							
Intended Targets	Listeners' Responses							Intended Targets	Listeners' Responses						
	/i/	/ɪ/	/e/	/æ/	/ʌ/	/a/	Other		/i/	/ɪ/	/e/	/æ/	/ʌ/	/a/	Other
/i/	45.8	40.0	5.8	0.8			7.5	/i/	47.5	39.2	4.2				9.2
/ɪ/	24.2	65.0	9.2				1.7	/ɪ/	25.8	63.3	6.7	0.8			3.3
/e/	0.8		97.5			0.8	0.8	/e/	1.7		95.8	0.8			1.7
/æ/				80.0	15.0	1.7	3.3	/æ/			1.7	78.3	15.0	1.7	3.3
/ʌ/				7.1	37.9	21.4	33.6	/ʌ/				7.1	40.7	20.0	32.1
/a/	8.3				11.7	76.7	3.3	/a/	3.3				18.3	76.7	1.7
Korean Accented Test															
No Accented Training								Same/Korean Accented Training							
Intended Targets	Listeners' Responses							Intended Targets	Listeners' Responses						
	/i/	/ɪ/	/e/	/æ/	/ʌ/	/a/	Other		/i/	/ɪ/	/e/	/æ/	/ʌ/	/a/	Other
/i/	92.5	6.7					0.8	/i/	92.5	4.2	0.8				2.5
/ɪ/	35.8	50.0	2.5			0.8	10.8	/ɪ/	28.3	58.3	1.7	1.7			10.0
/e/			95.0			0.8	4.2	/e/			99.2				0.8
/æ/			3.3	68.3	5.0		23.3	/æ/				80.0	8.3		11.7
/ʌ/				1.4	66.4	27.1	5.0	/ʌ/					72.9	20.0	7.1
/a/		1.7			13.3	85.0		/a/					8.3	91.7	
Different/Spanish Accented Training								Mixed Accented Training							
Intended Targets	Listeners' Responses							Intended Targets	Listeners' Responses						
	/i/	/ɪ/	/e/	/æ/	/ʌ/	/a/	Other		/i/	/ɪ/	/e/	/æ/	/ʌ/	/a/	Other
/i/	43.3	44.2	5.8	0.8			5.8	/i/	89.2	10.8					
/ɪ/	27.5	59.2	8.3	0.8			4.2	/ɪ/	18.3	66.7	3.3	0.8			10.8
/e/			95.8	0.8			3.3	/e/		0.8	96.7				2.5
/æ/		1.7		73.3	18.3	5.0	1.7	/æ/			6.7	73.3	10.0		10.0
/ʌ/		1.4	0.7	7.1	38.6	24.3	27.9	/ʌ/				2.9	67.9	22.9	6.4
/a/	11.7			1.7	13.3	63.3	10.0	/a/				3.3	11.7	83.3	1.7

To further examine how specific vowel properties may explain some of the effects of adaptation, we performed acoustic analyses on the six target vowels. For the vowels that showed significant improvement in identification performance, listeners may be learning specific temporal and spectral characteristics that distinguish one vowel from another, either that correspond to an accent-category (e.g., Spanish to Spanish) or that produce reliable changes in vowel identification performance across accents. In order to assess the potential basis of listener performance across Spanish- and Korean-accented training, acoustic properties of the accented

and native utterances were analyzed and compared. We also measured acoustic vowel properties for the mixed-accent talkers as a group to examine how mixed-accent training might facilitate identification of Spanish- or Korean-accented tokens at test.

2. Acoustic analyses

First (F1) and second (F2) formant center frequencies and the duration of the vowels assessed in the error analyses were analyzed using Praat (Boersma and Weenink, 2019)

and compared to native English tokens. English vowels from six native English speakers, the 12 native Spanish speakers, and the 12 native Korean speakers were analyzed. We also analyzed acoustic characteristics of the 12 mixed accent speakers used in the training experiments in order to evaluate similarities within and across accent groups. For these analyses, only words containing the target vowels /i/, /ɪ/, /e/, /æ/, /ʌ/, and /a/ were used. Each of these vowels had 12–16 tokens per speaker and were embedded in the stimulus words with varying consonant contexts. The native English and native Spanish speech was originally analyzed in Sidaras *et al.* (2009). The same criteria for determining vowel onset and offset were used in the current study. Those criteria were taken from Munson and Solomon (2004) and provided guidelines for finding the onset and offset of the vowel given the surrounding consonants. The first and second formants were measured halfway between the onset and the offset of the vowel at the midpoint. The vowel duration was determined by subtracting the onset time from the offset time. Table V shows means and standard deviations for F1, F2, and duration.

a. Spectral characteristics. To examine F1 and F2 across accents, vowels were grouped into the three high front vowels /i/, /ɪ/, and /e/, and three low vowels /æ/, /ʌ/, and /a/. These vowels and the accent groups were compared on F1 and F2.²

For /i/, /ɪ/, and /e/, native English, Spanish, and Korean speakers differentiated these vowels on F1. However, for the Mixed group of accented speakers, only /i/ vs /ɪ/ and /i/ vs /e/ were significantly different on F1, and /ɪ/ and /e/ were not. While all the vowels were significantly different from one another in Spanish and Korean accented speech, the patterns of F1 values were different from those in the speech of native English speakers. For native English speakers, the /ɪ/ vowel had a higher F1 value than /e/ did,³ but for Spanish and Korean speakers, the F1 of /ɪ/ was significantly lower than /e/. With regard to F2, native English speakers differentiated /i/, /ɪ/, and /e/, but native Spanish and Korean speakers did not. For the Mixed group of accented speakers, all vowels, /i/, /ɪ/, and /e/, were significantly different on F2, with a similar pattern to the native English speakers. Figure 3 shows mean F1 and F2 values for each speaker group for /i/, /ɪ/, and /e/.

For /æ/, /ʌ/, and /a/, native English speakers differentiated all vowels on F1. Native Spanish speakers showed differences between /æ/ and /ʌ/ and between /æ/ and /a/, but not between /ʌ/ and /a/. Native Korean speakers showed no significant differences among the vowels, /æ/, /ʌ/, and /a/, for F1. The Mixed group of accented speakers only showed differences between /æ/ and /a/ on F1. Thus, while the F1 values of all the /æ/, /ʌ/, and /a/ vowels were significantly different from one another in native English speech, the F1 values of these vowels were not as distinct in the other accents. Spanish accented speech showed a similar pattern to native English speakers but with smaller differences. Both the Korean-accented speech and the Mixed-accent speech had very small F1 differences among these three vowels. The F2 values of the /æ/, /ʌ/, and /a/ vowels were

TABLE V. Mean values and standard deviations for duration, F1, and F2 for native English speakers, Spanish accented speakers, Korean accented speakers, and Mixed accent speakers.

Speaker Group	Vowel	Duration		F1		F2	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
English	/i/	193.63	33.04	347.39	49.14	2582.53	234.50
	/ɪ/	157.25	26.64	542.76	128.77	2056.82	159.30
	/e/	221.22	40.55	451.36	101.22	2458.94	162.19
	/æ/	225.35	31.05	849.90	204.44	1220.24	70.09
	/ʌ/	181.09	29.78	691.22	189.40	1853.79	109.96
	/a/	237.01	46.18	761.70	203.36	1410.58	133.74
Spanish	/i/	178.19	36.64	356.76	53.84	2433.35	229.45
	/ɪ/	169.32	42.08	376.34	46.34	2426.63	301.20
	/e/	240.48	42.46	441.35	55.30	2374.18	267.31
	/æ/	215.15	35.53	784.88	146.43	1312.00	142.82
	/ʌ/	185.12	41.75	633.13	84.12	1646.25	144.18
	/a/	195.91	33.97	673.43	87.70	1346.96	197.82
Korean	/i/	186.70	43.67	362.43	39.73	2348.90	255.40
	/ɪ/	146.68	31.04	416.46	62.79	2273.40	270.14
	/e/	214.89	39.22	481.64	45.30	2254.96	300.99
	/æ/	185.85	56.66	741.90	122.01	1228.12	83.17
	/ʌ/	157.05	23.75	706.32	104.46	1808.19	181.60
	/a/	191.27	47.72	719.65	93.22	1250.45	154.13
Mixed	/i/	203.46	27.07	328.90	65.07	2420.23	253.37
	/ɪ/	150.13	28.67	449.22	55.92	2116.54	228.26
	/e/	201.27	36.47	451.17	55.55	2298.02	225.05
	/æ/	174.62	28.75	757.31	77.18	1192.04	126.84
	/ʌ/	182.38	28.09	722.18	86.74	1716.87	133.12
	/a/	176.83	26.85	691.30	84.15	1339.22	145.78

differentiated from one another in native English speech and in the Mixed accent group, but Korean and Spanish accented speech showed little difference in F2 between their productions of /ʌ/ and /a/. While the Mixed accent group differentiated between /ʌ/ and /a/ in F2, the difference was smaller than for native English speakers. Figure 4 shows mean F1 and F2 values for each speaker group for /æ/, /ʌ/, and /a/.

b. Temporal characteristics. For native English speakers and native Korean speakers, all vowels, /i/, /ɪ/, and /e/, were significantly different on duration. For native Spanish speakers, /i/ and /e/ differed in duration, as did /ɪ/ and /e/, but /i/ and /ɪ/ were not significantly different. For the Mixed group of accented speakers, /i/ and /ɪ/ differed in duration, as did /ɪ/ and /e/, but /i/ and /e/ were not significantly different. Figure 5(a) shows duration information for /i/, /ɪ/, and /e/.

For the vowels /æ/, /ʌ/, and /a/, native English speakers showed significant differences in /æ/ and /ʌ/ as well as /ʌ/ and /a/, but not /æ/ and /a/. For native Spanish speakers, all vowels, /æ/, /ʌ/, and /a/, were significantly different on duration. For native Korean speakers, /ʌ/ and /a/ were significantly different in duration but not the other pairs of vowels. For the Mixed group of accented speakers, there were no significant differences among vowels on duration. For duration for the /æ/, /ʌ/, and /a/ vowels, native English speakers showed longer durations for /æ/ and /a/ than for /ʌ/. Native Korean and native Spanish speakers showed the same pattern

FIG. 3. Mean F1 and F2 values in Hz for /i/, /ɪ/, and /e/ for native English speakers, Spanish accented speakers, Korean accented speakers, and mixed accent speakers.

as native English speakers, but with smaller differences. Figure 5(b) shows duration information for /æ/, /ʌ/, and /a/.

Taken together, the analysis of the temporal and spectral properties suggests that relative to the properties of native English vowels, each group of accented speakers distinguished among the vowel trios (high front and back) in distinct ways. Certainly, the lack of differentiation in spectral characteristics (particularly F2) in the Spanish- and Korean-accented speakers' production of high front vowels yielded compressed and shifted vowel spaces that differed from the native English vowel space and from each other. The average vowel space across the mixed accent group for these vowels was more expanded and thus more similar to the native English speakers' productions than the other two accent groups, but clear differences were still observed. The vowel spaces for the back vowels were similarly compressed

FIG. 4. Mean F1 and F2 values in Hz for /æ/, /ʌ/, and /a/ for native English speakers, Spanish accented speakers, Korean accented speakers, and mixed accent speakers.

FIG. 5. Mean durations for native English speakers, Spanish accented speakers, Korean accented speakers, and mixed accent speakers for (a) vowels /i/, /ɪ/, and /e/, and (b) for vowels /æ/, /ʌ/, and /a/.

for all accent groups compared to native English productions. Here, the Spanish-accented productions were more similar to native English productions, albeit shifted, while the other two accent groups differed strikingly (particularly in F1). Thus, with respect to spectral characteristics, the acoustic analyses suggest distinct patterns of productions across accent groups, perhaps providing the basis for some accent specificity in the perceptual adaptation process.

Relative to the native English productions, the pattern of differences in temporal characteristics among the high front vowels was most similar for the Korean-accented productions and perhaps least similar for the Spanish- and Mixed-accented productions (particularly the durations of /i/ and /ɪ/ for Spanish and /i/ and /e/ for Mixed). Likewise, the pattern for the back vowels was again most similar for the Korean-accented productions, but less so for both the Spanish- and Mixed-accent group productions. Again, with respect to duration, each accent group and the average of the mixed accented productions yielded patterns distinct both from English and from the other accents.

These acoustic analyses highlight both that each accent group has a unique set of acoustic properties associated with particular vowels, as well as certain similarities of specific features across accent groups. To speculate, the impact of the particular sets of properties and the particular degree of overlap in acoustic-phonetic features may have driven specificity and generalization, as least for vowels. For example, we observed poor identification of Spanish-accented front vowels and little improvement in identification across conditions. Acoustically, the Spanish-accented front vowel space is highly compressed with little differentiation in temporal properties and exposure to the expanded and shifted front vowel spaces of either Korean- or mixed accents would be expected to provide little benefit as a function of training. In contrast, exposure to productions of /a/, which are more similar to native realizations, in Korean- and mixed-accented groups may have served to shift identification of the Spanish-accented realization of the vowel along the relevant spectro-temporal dimensions. The degree to which any particular set of training and test materials align on certain segmental properties may explain generalization from one accent to another. One caveat of course is that the different accented conditions did not vary exclusively in vowel properties. A complete understanding of the relationships among the accents would necessarily need to include examination of properties of

consonants and perhaps word-level stress patterns (see Baese-Berk *et al.*, 2013 for a similar discussion).

IV. GENERAL DISCUSSION

This study examined the extent to which perceptual adaptation to accented speech is specific to the regularities in pronunciation associated with a particular set of foreign-accented talkers and with a particular type of accent. The goal was to determine what any specificity of learning might imply for the mechanisms underlying perceptual learning of variation in spoken language. Experiment 1 replicated and extended previous research (e.g., Sidaras *et al.*, 2009) by showing perceptual adaptation to Korean-accented speech, both for word-length utterances produced by the set of speakers encountered during training and for a new group of Korean-accented speakers. These results suggest that perceptual adaptation extends to segmental variation found in word-length utterances and that talker familiarity did not significantly impact the degree and extent of generalization of learning. Generalization occurred to words produced by speakers of the familiar *accent*, regardless of the familiarity of particular talker-specific instantiations. Experiment 2 demonstrated that perceptual adaptation to accented words generalized to novel speakers with the same but not a different accent, suggesting a degree of specificity of the adaptation process, similar to that found by Bradlow and Bent (2008). However, some evidence of generalization across accents was found with exposure to multiple types of accented speech, suggesting that exposure to mixed accent types might be beneficial, depending on the degree to which the particular accents share acoustic-phonetic properties. Followup analyses of the mixed accent conditions as well as analyses of vowel identification performance and acoustic characteristics were consistent with the idea that segmental properties that crosscut accents may serve as one basis for cross-accent generalization.

A. Talker-independent perceptual adaptation

These findings suggest that adaptation to an accent involves learning systematic properties independent of idiosyncratic pronunciations associated with specific talkers. This is consistent with previous research showing talker-independent perceptual adaptation across a variety of accent types, utterance types, and specific paradigms, including offline transcription tasks and on-line processing tasks (Bradlow and Bent, 2008; Sidaras *et al.*, 2009; Xie *et al.*, 2018). Here, however, we also observed no particular benefit for test materials produced by the same set of accented talkers relative to a different set of same accent talkers. Sidaras *et al.* (2009) reported a similar finding, and at first glance, the lack of an effect of talker familiarity seems at odds with recent work showing that generalization of learning for an accented segment is conditioned by similarity in the pronunciation of the individual talker(s) encountered across training and test (Reinisch and Holt, 2014; Xie and Myers, 2017), rather than to exposure to group-wide variation. For example, while Xie and Myers (2017) found generalization to a new speaker with the same accent, that generalization depended on

similarity of productions between a specific talker presented during training and the talker at test. In our current study, the same-talker group would necessarily include talkers that were the most similar from training to test. As such, we would have expected a benefit for the repetition of those particular talker characteristics. That we did not find an advantage for familiar talkers at test suggests that aspects of the learning environment and presentation and types of stimulus materials may play a key role in the extent to which learning will or will not generalize based on individual talker productions.

Alternatively, perhaps the same versus different talker groups were sufficiently similar to one another in the distribution of relevant acoustic-phonetic characteristics that learning generalized across sets of speakers regardless of talker-specific familiarity. On this view, any dissimilarities between the same versus different talkers at test may have manifested along dimensions that were irrelevant for the extraction of linguistic structure (e.g., Levi *et al.*, 2011) and therefore, same and different talker groups may have been effectively equally similar at test. Further work examining the specific ways in which particular speakers differ along each of the putatively relevant and irrelevant dimensions of variation would be needed to create a full assessment of the nature of linguistically-relevant talker-specific similarity.

With respect to possible differences across assays of talker-independent learning that may have affected generalization performance, in our experiments, listeners were asked to transcribe multiple accented talkers at both training and test. This type of design limits listeners' exposure to any particular talker, either during exposure or at test, and as such, may have limited the registration of particular, idiosyncratic realizations of acoustic-phonetic form in favor of adaptation to the particular range and distribution of variation. In addition, in contrast to previous work (e.g., Reinisch and Holt, 2014), our materials contained multiple types of segments and presumably multiple shifts in segmental realization across the stimulus set. That is, our stimuli were highly variable and naturalistic. Tracking variation across multiple segments with this kind of stimuli may have changed the nature of the learning task, leading to the registration of accent-general as opposed to talker-specific variation.

Regardless, the relative specificity and generalization that was found suggests that this type of perceptual adaptation is a highly complex task for the learner. For perceptual learning to occur, listeners must correctly parse and attribute variation due to the shared accent characteristics from variation from other sources. Listeners' adaptation to properties of an accent that generalize to novel talkers suggests that they extracted properties of the accent that are common across talkers and that they did not use idiosyncratic characteristics that do not aid in retrieving linguistic structure across speakers. Our task may have facilitated this learning by exposing listeners to multiple speakers of each accent during both training and test, leading to greater sensitivity to properties that were consistent *across* talkers. High-variability training might allow listeners to detect cross-talker similarities and highlight contrastive properties between and across accents (Goldstone, 1994, 1995; Tzeng *et al.*, 2016). This type of learning may be dependent, however, on the range of variation within and across the categories

to be learned. [Wade et al. \(2007\)](#) found that high-variability training can be beneficial within a certain range of category overlap and variation, but detrimental with high variation and overlap of linguistic category structure (see also [Perrachione et al., 2011](#)).

B. Relative specificity and generalization of accent learning

In examining transfer of training to novel utterances produced by novel speakers, we found evidence for both specificity and generalization in the perceptual adaptation process. That training on either Spanish- or Korean-accented speakers did not improve transcription performance for words produced by a set of different accented speakers (Korean- and Spanish-accented respectively) suggests that listeners were adapting to segment-level variation that was characteristic of a particular accent. This finding is consistent with the classic work of [Bradlow and Bent \(2008\)](#). However, when listeners were presented with talkers with mixed foreign accents, we saw intermediate effects, with some evidence of learning. That the results of mixed accent training neither patterned with same accent learning nor with no training controls suggests that there were some properties of the mixed accented productions that benefited learning. Although our follow-up analyses examining performance after each of the mixed accented groups should be interpreted with caution given that the test accents differed in a variety of ways (most notably in baseline intelligibility and acoustic-phonetic properties), these data suggested that one group of mixed accented speakers generalized across accents while the other did not. Likewise, benefits of mixed accent training were found for Spanish-accented but not Korean-accented tokens. Taken together with the pattern of improvement as a function of training for individual vowels and comparisons of vowel space across accented types, these findings suggest that listeners were tuning to segmental variation wherever it arose, generalizing learning of segmental instantiation across the particular accent or group of accents.

Why then did accented training with the same accent from study to test result in the most robust learning? One possibility is that same accented training provided increased exposure to the particular set of accented variations that would be relevant at test. Learning was specific to the particular segmental variation associated with training stimuli, which resulted in adaptation for listeners trained and tested with the same accent, but also resulted in some more limited evidence for generalization across accents. Because Spanish- and Korean-accented speech differ from native English speech in specific, distinct, and systematic ways, adaptation necessitated learning patterns of variation shared by the speakers of each accent group. Thus, rather than learning being associated with a particular accent category or class, learning involved adaptation to a mosaic of characteristics that were or were not useful during subsequent accented speech perception. This explanation is consistent with other research demonstrating generalization from learning with multiple accents to improved intelligibility of a single different accented speaker ([Baese-Berk et al., 2013](#)). The current

study extends this work by providing evidence that generalization may be driven by specific similarities between properties of accented speech from study to test. For example, properties of accented speech such as differences in duration or timing and variation in spectral characteristics can be shared across accents and accented speakers ([Xie and Myers, 2017](#)), providing a basis for examination of patterns of specificity and generalization during adaptation.

We should note, however, that evidence for generalization in the current experiment was not as conclusive as that found in [Baese-Berk et al. \(2013\)](#). There may be several reasons why we did not observe the same level of generalization. In addition to the potential differences in overlap of segmental content across accents mentioned above, our study used words rather than sentence-length utterances. It may be that properties such as slower speaking rate or certain kinds of vowel reduction which are characteristic of accented speech in general may have been less prominent, salient, or present in our set of stimuli ([Grohe and Weber, 2016](#)). By presumably restricting perceptual adaptation to segmental properties, the degree of generalization may have depended on the specific realizations of sets of segments. Alternatively, the nature of the training procedure may have impacted the extent of generalization. [Baese-Berk et al. \(2013\)](#) trained listeners over two days, providing an opportunity for consolidation from sleep. [Xie et al. \(2017\)](#) found that an episode of sleep facilitated generalization of accent learning to a novel talker. Similarly, the opportunity to sleep in [Baese-Berk et al. \(2013\)](#) may have facilitated generalization of accent learning from the mixed accent presentation not only to a novel speaker but to a novel accent. Future work will be needed to fully evaluate this possibility.

Although our findings do suggest improvement in identification of particular vowel segments as a function of the type of exposure and the overlap between the variation encountered during training and that encountered at test, the patterns of improvement were not always straightforward and the nature of the acoustic overlap in spectral and temporal properties was complex. Nevertheless, the analysis of error patterns and the distribution of acoustic characteristics for vowels across accents and training conditions suggest that listeners were encoding distributions of variation associated with particular segments and particular accents and, to the extent that those distributions were similar from training to test, we observed benefits of perceptual adaptation. Consistent with research examining lexically-driven perceptual learning of particular segmental contrasts (e.g., [Kraljic and Samual, 2007](#)), listeners appear to track and adapt variation in acoustic-phonetic form and category structure. Future studies will need to extend examination of learning to other segment types while striking a balance between the examination of the full range and complexity of variation associated with natural foreign-accented speech and precise control over which segments vary and in what way (e.g., [Wade et al., 2007](#)).

C. Perceptual adaptation to foreign-accented speech

The patterns of perceptual learning observed in our task raise questions about the nature of representational change

as a function of perceptual adaptation and how and when listeners track different sources of variation in spoken language. Certainly, representational or processing changes must have occurred to accommodate the listeners' accumulated experience with accented speech. However, the nature of those changes remains unclear. One clue comes from our finding that lexical factors did not interact with the degree or type of learning. Easy words were uniformly identified better than hard words and both types of lexical items improved as a function of experience. This finding suggests that the changes to segmental processing or representation may have occurred prelexically with benefits propagating across easy and hard words. Whatever the nature of those changes, the necessity for higher or lower resolution of phonetic form for the two types of words did not differentially impact the effects of perceptual adaptation on intelligibility. This finding is consistent with research demonstrating contributions of both signal-dependent and signal-independent properties to perceptual resolution of sinewave speech (Remez *et al.*, 2011) and argues against any general shift to a different kind of perceptual process or strategy, at least with these stimuli and task structure.

What then changed as a function of learning? One possibility is that listeners restructured the mapping from acoustic-phonetic form to phonological or lexical category (Dupoux and Green, 1997; Francis *et al.*, 2000; Greenspan *et al.*, 1988; Liss *et al.*, 2002; McQueen *et al.*, 2006) in a segment-specific manner. With experience, listeners might develop segment-specific filters to facilitate the mapping of accented speech onto existing speech categories. Alternatively, listeners may amass exemplars of accented utterances that extend their current phonological or lexical representations in particular ways. As listeners accumulate experience with accented items, the boundaries of relevant linguistic categories shift in an accent- or segment-specific manner to include new, accent-consistent items as category members (Ettlinger, 2007; Goldinger, 1998). On this view, although representational structures retain the range of variation, depending on the task and the test conditions, the type and extent of generalization might appear abstract, reflecting the effects of multiple exemplars. A third possibility is that listeners track the statistics of systematic variation, registering the distributional properties of segments, voices, accents, and languages (Kleinschmidt and Jaeger, 2015; McMurray and Jongman, 2011; Zhang and Holt, 2018). Here, too, listeners would be sensitive to variation at different levels and use that sensitivity to modify their subsequent linguistic processing.

In any case, it did not appear that listeners necessarily tracked properties of particular accents, but rather seemed to register variation in a segment- or word-specific manner. That is, our findings did not provide clear evidence that listeners were associating constellations of properties with a particular accent, in essence forming a category or class specific to a particular foreign accent. However, our findings did suggest that listeners were tracking the range of variation associated with particular segments. Recent work suggests that listeners in other contexts may be able to track multiple sources, patterns, or distributions of variation during speech perception, associating patterns of variation with particular

talkers (Zhang and Holt, 2018) or with particular causes (Liu and Jaeger, 2018). These studies raise the intriguing possibility that listeners may have demonstrated more specificity of learning and less overall generalization if accented utterances were associated with particular speaker or speaker-group characteristics, as in Zhang & Holt (2018).

V. CONCLUSION

The speech perception system is remarkable in its ability to both readily adapt to new types of speech and maintain perceptual constancy in the face of enormous variation. That listeners appear to extract and encode classes of systematic accent variation is consistent with accounts of spoken language processing that posit perceptually-detailed representations for spoken language and/or tracking of systematic variation (Goldinger, 1998; Johnson, 1997; Jusczyk, 1993; Kleinschmidt and Jaeger, 2015; Pisoni, 1993, 1997; Sumner *et al.*, 2014). That listeners engage in perceptual adaptation to accented speech suggests that the perceptual system may be able to produce stable linguistic perception specifically because it adapts so readily to the ever-changing characteristics of speech input at multiple levels of organization—from segmental-specific to talker-specific to accent-specific adjustments.

The current study examined specificity in perceptual adaptation to accented speech, replicating and extending previous work by showing adaptation to Korean-accented speech regardless of speaker familiarity as well as patterns of specificity and generalization as the result of training with the same, different, or mixed accent groups. Perceptual learning of one accent did not transfer to another but exposure to multiple speakers of one accent facilitated processing of novel speakers of that same accent and exposure to multiple accents may have produced generalization to novel talkers and accents. Generalization seemed to occur for specific segments and specific overlap between training and test stimuli. The findings support an account of perceptual adaptation for accented speech based on transfer of learning across similar specific acoustic-phonetic properties of speech.

ACKNOWLEDGMENTS

Portions of this work were presented at the 50th Annual Meeting of the Psychonomic Society, Boston, Massachusetts and were based on Jessica Alexander's doctoral dissertation. We thank Laura Namy, Donald Tuten, Harold Gouzoules, and Stella Lourenco for helpful comments on this research. We also thank Kathy Jernigan, Melanie Hammet, Melanie Tumlin, and Sabrina Sidaras for their labor and sage advice. This research was supported by a grant from the National Institute of Deafness and Other Communication Disorders (Grant No. DC008108).

¹Trained listeners were compared only to no training controls. Previous work has been mixed regarding whether control conditions using native English utterances during training yield better test performance than no training controls (e.g., Bradlow & Bent, 2008; Sidaras *et al.*, 2009; Tzeng *et al.*, 2016). Work in our laboratory has shown minimal differences with our paradigm (single words and multiple talkers at test) between no training controls and native English trained controls. Therefore, in the current experiment, English trained controls were not included.

²See supplementary material at <https://doi.org/10.1121/1.5110302> for documentation of statistical analyses for acoustic analyses.

³For this set of native English speakers (three male and three female), average F1 for /t/ appeared somewhat higher than has typically been found (e.g., Hillenbrand *et al.*, 1995) although not out of the range of recorded values. This difference may be due to the subset of consonant environments in our stimulus set or to the particular talker characteristics of these native English speakers.

- Adank, P., and Janse, E. (2010). "Comprehension of a novel accent by young and older listeners," *Psychol. Aging* **25**(3), 736–740.
- Ahissar, M., and Hochstein, S. (1997). "Task difficulty and the specificity of perceptual learning," *Nature* **387**, 401–406.
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). "Mixed-effects modeling with crossed random effects for subjects and items," *J. Mem. Lang.* **59**, 390–412.
- Baese-Berk, M. M., Bradlow, A. R., and Wright, B. A. (2013). "Accent-independent adaptation to foreign accented speech," *J. Acoust. Soc. Am.* **133**(3), EL174–EL180.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.* **67**(1), 1–48.
- Bent, T., and Holt, R. F. (2013). "The influence of talker and foreign-accent variability on spoken word identification," *J. Acoust. Soc. Am.* **133**(3), 1677–1686.
- Boersma, P., and Weenink, D. (2019). "Praat: Doing phonetics by computer (version 6.0.46) [computer program]," <http://www.praat.org/> (Last viewed 3 January 2019).
- Bradlow, A. R., and Bent, T. (2008). "Perceptual adaptation to non-native speech," *Cognition* **106**(2), 707–729.
- Bradlow, A. R., and Pisoni, D. B. (1999). "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors," *J. Acoust. Soc. Am.* **106**, 2074–2085.
- Clarke, C. M., and Garrett, M. F. (2004). "Rapid adaptation to foreign-accented English," *J. Acoust. Soc. Am.* **116**, 3647–3658.
- Creel, S. C., Aslin, R. N., and Tanenhaus, M. K. (2008). "Heeding the voice of experience: The role of talker variation in lexical access," *Cognition* **106**(2), 633–664.
- Creel, S. C., and Tumlin, M. A. (2011). "On-line acoustic and semantic interpretation of talker information," *J. Mem. Lang.* **65**(3), 264–285.
- Dahan, D., Drucker, S. J., and Scarborough, R. A. (2008). "Talker adaptation in speech perception: Adjusting the signal or the representations?," *Cognition* **108**(3), 710–718.
- 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**(3), 914–927.
- Eisner, F., and McQueen, J. M. (2005). "The specificity of perceptual learning in speech processing," *Percept. Psychophys.* **67**(2), 224–238.
- Ettlinger, M. (2007). "Shifting categories: An exemplar-based computational model of chain shifts," UC Berkeley Phonology Lab Annual Report, Berkeley, CA, pp. 177–182.
- Flege, J. E., Munro, M. J., and Skelton, L. (1992). "Production of the word-final English /t/-d/ contrast by native speakers of English, Mandarin, and Spanish," *J. Acoust. Soc. Am.* **92**(1), 128–143.
- Flege, J. E., Schirru, C., and MacKay, I. R. A. (2003). "Interaction between the native and second language phonetic subsystems," *Speech Commun.* **40**(4), 467–491.
- Francis, A. L., Baldwin, K., and Nusbaum, H. C. (2000). "Effects of training on attention to acoustic cues," *Percept. Psychophys.* **62**(8), 1668–1680.
- Francis, A. L., Nusbaum, H. C., and Fenn, K. (2007). "Effects of training on the acoustic-phonetic representation of synthetic speech," *J. Speech Lang. Hear. Res.* **50**(6), 1445–1465.
- Goldinger, S. D. (1998). "Echoes of echoes? An episodic theory of lexical access," *Psychol. Rev.* **105**(2), 251–279.
- Goldstone, R. L. (1994). "Influences of categorization on perceptual discrimination," *J. Exp. Psychol. Gen.* **123**, 178–200.
- Goldstone, R. L. (1995). "Effects of categorization on color-perception," *Psychol. Sci.* **6**(5), 298–304.
- Greenspan, S. L., Nusbaum, H. C., and Pisoni, D. B. (1988). "Perceptual-learning of synthetic speech produced by rule," *J. Exp. Psychol. Learn. Mem. Cogn.* **14**(3), 421–433.
- Grohe, A.-K., and Weber, A. (2016). "Learning to comprehend foreign-accented speech by means of production and listening training," *Lang. Learn.* **66**(3), 187–209.
- Hanulíková, A., van Alphen, P. M., van Goch, M. M., and Weber, A. (2012). "When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing," *J. Cogn. Neurosci.* **24**, 878–887.
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**(5), 3099–3111.
- Hothorn, T., Bretz, F., and Westfall, P. (2008). "Simultaneous inference in general parametric models," *Biometrical J.* **50**(3), 346–363.
- Ingvalson, E. M., McClelland, J. L., and Holt, L. L. (2011). "Predicting native English-like performance by native Japanese speakers," *J. Phon.* **39**(4), 571–584.
- Johnson, K. (1997). "Speech perception without speaker normalization: An exemplar model," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic Press, New York), pp. 145–165.
- Johnson, K. (2006). "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology," *J. Phon.* **34**, 485–499.
- Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., and Carlyon, R. P. (2013). "Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice," *Psychol. Sci.* **24**(10), 1995–2004.
- Juszyk, P. W. (1993). "From general to language-specific capacities—The WRAPSA model of how speech-perception develops," *J. Phon.* **21**, 3–28.
- Kleinschmidt, D. F., and Jaeger, T. F. (2015). "Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel," *Psychol. Rev.* **122**, 148–203.
- Kraljic, T., and Samuel, A. G. (2007). "Perceptual adjustments to multiple speakers," *J. Mem. Lang.* **56**, 1–15.
- Lee, B., Guion, S. G., and Harada, T. (2006). "Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals," *Stud. Sec. Lang. Acq.* **28**(3), 487–513.
- Levi, S. V., Winters, S. J., and Pisoni, D. B. (2011). "Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible?," *J. Acoust. Soc. Am.* **130**(6), 4053–4062.
- Liss, J. M., Spitzer, S. M., Caviness, J. N., and Adler, C. (2002). "The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria," *J. Acoust. Soc. Am.* **112**(6), 3022–3030.
- Liu, L., and Jaeger, T. F. (2018). "Inferring causes during speech perception," *Cognition* **174**, 55–70.
- Luce, P. A., and Pisoni, D. B. (1998). "Recognizing spoken words: The neighborhood activation model," *Ear Hear.* **19**(1), 1–36.
- Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., and Aslin, R. N. (2007). "The dynamics of lexical competition during spoken word recognition," *Cogn. Sci.* **31**, 1–24.
- Maye, J., Aslin, R., and Tanenhaus, M. (2008). "The weckud wetch of the wast: Lexical adaptation to a novel accent," *Cogn. Sci.* **32**(3), 543–562.
- McMurray, B., and Jongman, A. (2011). "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations," *Psychol. Rev.* **118**(2), 219–246.
- McQueen, J. M., Cutler, A., and Norris, D. (2006). "Phonological abstraction in the mental lexicon," *Cogn. Sci.* **30**(6), 1113–1126.
- Munson, B., and Solomon, N. P. (2004). "The effect of phonological neighborhood density on vowel articulation," *J. Speech Lang. Hear. Res.* **47**, 1048–1058.
- Nash, R. (1977). *Comparing English and Spanish: Patterns in Phonology and Orthography* (Regents Publishing Co., New York).
- Nissen, S. L., Dromey, C., and Wheeler, C. (2004). "First and second language tongue movements in Spanish and Korean bilingual speakers," in *Proceedings of the Annual Convention of the American Speech-Language-Hearing Association*, November 20, Philadelphia, PA.
- Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (1984). "Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words," *Res. Speech Percept. Prog. Report* **10**, 357–372.
- Nygaard, L. C., and Pisoni, D. B. (1998). "Talker-specific learning in speech perception," *Percept. Psychophys.* **60**(3), 355–376.
- Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (1994). "Speech-perception as a talker-contingent process," *Psychol. Sci.* **5**(1), 42–46.
- Papesh, M. H., Goldinger, S. D., and Hout, M. C. (2016). "Eye movements reveal fast, voice-specific priming," *J. Exp. Psychol. Gen.* **145**(3), 314–337.
- Perrachione, T. K., Lee, J., Ha, L. Y., and Wong, P. C. (2011). "Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design," *J. Acoust. Soc. Am.* **130**, 461–472.

- Pierrehumbert, J. B. (2003). "Phonetic diversity, statistical learning, and acquisition of phonology," *Lang. Speech*, **46**, 115–154.
- Piske T. (2012). "Factors affecting the perception and production of l2 prosody: Research results and their implications for the teaching of foreign languages," in *Pragmatics and Prosody in English Language Teaching: Educational Linguistics, Vol. 15*, edited by J. Romero-Trillo (Springer, Dordrecht, the Netherlands).
- Pisoni, D. B. (1993). "Long-term-memory in speech-perception—Some new findings on talker variability, speaking rate and perceptual-learning," *Speech Commun.* **13**, 109–125.
- Pisoni, D. B. (1997). "Some thoughts on 'normalization' in speech perception," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic Press, San Diego, CA).
- Potter, C. E., and Saffran, J. R. (2017). "Exposure to multiple accents supports infants' understanding of novel accents," *Cognition* **166**, 67–72.
- Reinisch, E., and Holt, L. L. (2014). "Lexically guided phonetic retuning of foreign-accented speech and its generalization," *J. Exp. Psychol. Hum. Percept. Perform.* **44**(2), 539–555.
- Remez, R. E., Dubowski, K. R., Broder, R. S., Davids, M. L., Grossman, Y. S., Moskalenko, M., Pardo, J. S., and Hasbun, S. M. (2011). "Auditory-phonetic projection and lexical structure in the recognition of sine-wave words," *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 968–977.
- Rice, K. (2002). *Vowel Place Contrasts* (Praeger, Westport, CT).
- Romero-Rivas, C., Martin, C. D., and Costa, A. (2015). "Processing changes when listening to foreign-accented speech," *Front. Hum. Neurosci.* **9**, 15.
- Schmale, R., Seidl, A., and Cristia, A. (2015). "Mechanisms underlying accent accommodation in early word learning: Evidence for general expansion," *Dev. Sci.* **18**(4), 664–670.
- Sidaras, S. K., Alexander, J. D., and Nygaard, L. C. (2009). "Perceptual learning of systematic variation in Spanish-accented speech," *J. Acoust. Soc. Am.* **125**(5), 3306–3316.
- Sumner, M., Kim, S. K., King, E., and McGowan, K. B. (2014). "The socially weighted encoding of spoken words: A dual-route approach to speech perception," *Front. Psychol.* **4**, 1–13.
- Trude, A. M., and Brown-Schmidt, S. (2012). "Talker-specific perceptual adaptation during online speech perception," *Lang. Cog. Process.* **27**(7–8), 979–1001.
- Tsukada, K., Birdsong, D., Mack, M., Sung, H. Y., Bialystok, E., and Flege, J. (2004). "Release bursts in English word-final voiceless stops produced by native English and Korean adults and children," *Phonetica* **61**(2–3), 67–83.
- Tzeng, C. Y., Alexander, J. E. D., Sidaras, S. K., and Nygaard, L. C. (2016). "The role of training structure in perceptual learning of accented speech," *J. Exp. Psychol. Hum. Percept. Perform.* **42**, 1793–1805.
- van Heugten, M., Paquette-Smith, M., Krieger, D. R., and Johnson, E. K. (2018). "Infants' recognition of foreign-accented words: Flexible yet precise signal-to-word mapping strategies.," *J. Mem. Lang.* **100**, 51–60.
- Vitevitch, M. S., Luce, P. A., Pisoni, D. B., and Auer, E. T. (1999). "Phonotactics, neighborhood activation, and lexical access for spoken words," *Brain Lang.* **68**(1–2), 306–311.
- Wade, T., Jongman, A., and Sereno, J. (2007). "Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds," *Phonetica* **64**(2–3), 122–144.
- Xie, X., Earle, F. S., and Myers, E. B. (2017). "Sleep facilitates generalisation of accent adaptation to a new talker," *Lang. Cogn. Neurosci.* **33**(2), 196–210.
- Xie, X., and Myers, E. B. (2017). "Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers," *J. Mem. Lang.* **97**, 30–46.
- Xie, X., Weatherholtz, K., Bainton, L., Rowe, E., Burchill, Z., Liu, L., and Jaeger, T. F. (2018). "Rapid adaptation to foreign-accented speech and its transfer to an unfamiliar talker," *J. Acoust. Soc. Am.* **143**, 2013–2031.
- Yang, B. G. (1996). "A comparative study of American English and Korean vowels produced by male and female speakers.," *J. Phon.* **24**(2), 245–261.
- Zhang, X., and Holt, L. L. (2018). "Simultaneous tracking of coevolving distributional regularities in speech.," *J. Exp. Psychol. Hum. Percept. Perform.* **44**(11), 1760–1779.