tapraid5/zfn-xhp/zfn-xhp/zfn01016/zfn3486d15z | xppws | S=1 | 8/9/16 | 0:29 | Art: 2015-0811

APA NLM

Journal of Experimental Psychology:
Human Perception and Performance

0096-1523/16/$12.00 http://dx.doi.org/10.1037/xhp0000285

More Than a Boundary Shift: Perceptual Adaptation to Foreign-Accented Speech Reshapes the Internal Structure of Phonetic Categories

AQ: au

Xin Xie, Rachel M. Theodore, and Emily B. Myers
University of Connecticut

The literature on perceptual learning for speech shows that listeners use lexical information to disambiguate phonetically ambiguous speech sounds and that they maintain this new mapping for later recognition of ambiguous sounds for a given talker. Evidence for this kind of perceptual reorganization has focused on phonetic category boundary shifts. Here, we asked whether listeners adjust both category boundaries and internal category structure in rapid adaptation to foreign accents. We investigated the perceptual learning of Mandarin-accented productions of word-final voiced stops in English. After exposure to a Mandarin speaker's productions, native-English listeners' adaptation to the talker was tested in 3 ways: a cross-modal priming task to assess spoken word recognition (Experiment 1), a category identification task to assess shifts in the phonetic boundary (Experiment 2), and a goodness rating task to assess internal category structure (Experiment 3). Following exposure, both category boundary and internal category structure were adjusted; moreover, these prelexical changes facilitated subsequent word recognition. Together, the results demonstrate that listeners' sensitivity to acoustic–phonetic detail in the accented input promoted a dynamic, comprehensive reorganization of their perceptual response as a consequence of exposure to the accented input. We suggest that an examination of internal category structure is important for a complete account of the mechanisms of perceptual learning.

Keywords: perceptual learning, foreign-accented speech, acoustic cues, spoken word recognition, talker accommodation

Natural speech exhibits substantial acoustic–phonetic variation such that as speech context varies, different acoustic patterns may denote the same linguistic information. One prominent source of speech variability arises from talker-related characteristics (e.g., age, gender, idiolect, dialect, and accent). Recent research has revealed that when speakers produce ambiguous words that deviate from canonical pronunciations (including idiosyncratic pronunciations, dialects, or foreign accents; see Kleinschmidt & Jaeger, 2015, for a review), listeners use top-down lexical information to recalibrate phonetic boundaries. As such, atypical talker-related phonetic variation is accommodated (e.g., Kraljic & Samuel, 2005, 2006; Norris, McQueen, & Cutler, 2003). For instance, if native listeners hear ambiguous tokens (e.g., midway between /s/ and /f/) embedded in /f/-biased words (e.g., *belie?*), they would interpret the ambiguous sound as /f/. More importantly, they subsequently categorize more consonant tokens along an acoustic continuum (e.g., /ɛf/ -/ɛs/), especially intermediate ambiguous tokens, into the /f/ category. This boundary adjustment extends to shifts in sensitivity measured by discrimination (Clarke-Davidson, Luce, & Sawusch, 2008) and is maintained over time for a given speaker (e.g., Eisner & McQueen, 2006; Kraljic & Samuel, 2005).

Despite evidence that phonetic retuning bolsters rapid adaptation to atypical pronunciations, including foreign-accented speech (Reinisch & Holt, 2014; Sumner, 2011), the effects of adaptation on the perceptual structure of a phonetic category and further on online lexical access are not fully understood. A plethora of research has investigated how continuous variation in the speech signal is mapped onto discrete categories and further affects spoken word recognition. This body of work establishes that phonetic categories have a graded internal structure: Members of the same phonetic category are not perceptually equivalent (e.g., Pisoni & Tash, 1974) and some may be represented as better exemplars than others, as revealed by overt judgment and covert psychological responses (e.g., Kuhl, 1991; Samuel, 1982). Moreover, listeners readily adjust best exemplars of a phonetic category in the face of acoustic variation arising from contextual variables, such as speaking rate or place of articulation (Miller & Volaitis, 1989; Volaitis & Miller, 1992). In contrast, relatively little work has examined

---

Xin Xie, Rachel M. Theodore, and Emily B. Myers, Department of
AQ: 5 Speech, Language and Hearing Sciences, University of Connecticut.

This research was supported by National Institutes of Health Grant R01 DC013064 to Emily B. Myers and a doctoral dissertation fellowship awarded to Xin Xie from University of Connecticut. Materials in this article formed part of the first author's doctoral dissertation. We are grateful for Carol Fowler and James Magnuson for their valuable advice throughout the project. We thank Adam Sheya and Leonard Katz for helpful consultation on statistical analyses and John Bugden and Dhvani Patel for data collection. Portions of this work were accepted for poster presentation at 56th Annual Meeting of the Psychonomic Society. The content is the responsibility of the authors and does not necessarily represent official views of the National Institute on Deafness and Other Com-
AQ: 6 munication Disorders.

Correspondence concerning this article should be addressed to Xin Xie, Department of Speech, Language, and Hearing Sciences and Department of Psychological Sciences, University of Connecticut, Storrs, CT
AQ: 7 06269. E-mail: xin.xie@uconn.edu

APA NLM
tapraid5/zfn-xhp/zfn-xhp/zfn01016/zfn3486d15z | xppws | S=1 | 8/9/16 | 0:29 | Art: 2015-0811

internal category structure with respect to long-term, talker-related adaptation (but see Theodore, Myers, & Lomibao, 2015, for short-term talker learning effects). Investigations have almost exclusively measured phonetic retuning in terms of a recalibration of phonetic category boundaries (Kraljic & Samuel, 2005, 2006; Norris et al., 2003; Reinisch & Holt, 2014). The primary goal of the present study is to investigate whether phonetic adaptation instigates a more pervasive effect both within and between phonetic categories than has been previously investigated. We asked whether lexically guided phonetic adjustments entail changes beyond the phonetic boundary region to also change perceived goodness of tokens throughout the phonetic category.

AQ: 1

Listeners are known to track subphonemic detail in a talker-specific manner, for example, linking talker identity with talkers' tendency to produce word-initial stops with short versus long voice-onset times (VOTs; Allen & Miller, 2004; Theodore & Miller, 2010). In these studies, listeners were exposed to tokens of two words from each speaker, all of which were typical of the native language but showed talker differences in word-initial VOTs. If evidence can also be established that perceived goodness of speech tokens changes following talker adaptation driven by phonetic retuning, then it aligns the literature on native speech perception with that on adaptation to accented productions.

Such results are predicted by several recent theoretical frameworks, which have attempted to provide a unifying account of speech perception, adaptation, and speech category learning in general by positing top-down guided distributional learning as the underlying mechanism (e.g., Feldman, Griffiths, Goldwater, & Morgan, 2013; Kleinschmidt & Jaeger, 2015). These frameworks hypothesize that listeners dynamically track the probability distributions of acoustic–phonetic properties, as they are associated with native speech categories. Speech categorization then is viewed as a probability inference problem, which requires listeners to make estimations of the likelihood of different categories given a particular speech token. The probabilistic models predict that as listeners encounter a speaker with atypical speech variation (i.e., novel statistics that are inconsistent with prior experience), they fully track the acoustic distributions of that speaker and use top-down information to update the probability of observing an acoustic property given a category, which could be captured using a measurement of the perceived "goodness" of a token as a member of the phonetic category. Taken together, these theories suggest that talker adaptation should affect any region within the perceptual space as much as it affects the boundary region. Conversely, if phonetic retuning is restricted to the boundary region, then it raises the question of whether the top-down label information is only used when the bottom-up input is ambiguous (e.g., Vong, Navarro, & Perfors, 2016). This result would in turn challenge the probabilistic frameworks and call for further examination on the limits of speech adaptation. In the current investigation, we ask whether a novel phonetic contrast introduced by Mandarin-accented speech will show similarly pervasive effects on phonetic categorization and internal category structure ("goodness") and whether these shifts will extend to lexical activation of candidate words.

Adaptation to foreign-accented speech is well suited to address the question of perceptual reorganization because natural foreign accents usually not only deviate from the native categories along multiple acoustic dimensions (e.g., Flege, Munro, & Skelton, 1992), but also exhibit much larger intratalker variability relative to native pronunciations (Wade, Jongman, & Sereno, 2007). Hence, we expected foreign-accented words to exhibit distinctive acoustic patterns from native norms and elicit perceptual adjustment throughout the entire acoustic–phonetic space. Of interest, Eisner, Melinger, and Weber (2013) examined the process by which native English listeners adapted to Dutch-accented English in which, as in Dutch itself, final obstruents were devoiced. For instance, word-final /d/s in this accent are acoustically and perceptually similar to members of the unintended category /t/ (Warner, Jongman, Sereno, & Kemps, 2004) such that a word like *seed* will be produced similar to the word *seat*. In this study, exposure to multisyllabic words (e.g., *overload*) containing devoiced /d/ in word-final position produced changes in lexical access. Specifically, the accented production of *seed,* sounding like [si:t^h], primed the written form, *SEED* (auditory *seed*–visual *SEED*, henceforth *seed*–*SEED*) to a greater extent in listeners who had heard /d/-final words during exposure than in control listeners who did not have this exposure; namely, there was no significant identity *seed* ([si:t^h])–*SEED* priming for control listeners. This suggests that listeners accepted a Dutch-accented /d/ as a production of /d/ category only after the critical exposure.

We adapted the paradigm of Eisner et al. (2013) and examined perceptual adaptation to Mandarin-accented word-final stop consonants across three experiments. In Mandarin, all stops are voiceless (no /b/, /d/, or /g/ are found in Mandarin) and are categorically distinguished by aspiration, instead of voicing. That is, the two tokens [t^h] and [t] are phonologically contrastive in Mandarin, but are allophones of the same category /t/ in English. Because of this, Mandarin-accented stop productions demonstrate reduced or absent differences in vowel and closure durations (contrastive cues used by American English speakers), but maintain clear distinctions in burst durations (Bent, Bradlow, & Smith, 2008). As a consequence, voiced word-final stops in the Mandarin accent (e.g., /d/ as in *seed*) are typically heard by English listeners as /t/ because burst information is not readily exploited by native-English listeners (e.g., Flege et al., 1992).

Similar to Eisner et al. (2013), we assessed adaptation effects by examining group differences at test between participants who have the critical exposure to /d/-final tokens and participants who do not have such exposure. In each experiment, two groups of native-English listeners were exposed to naturally produced Mandarin-accented speech in an auditory lexical decision task during the exposure phase. The experimental group heard a set of critical /d/-final words that were devoiced in the Mandarin-accented speech, but the control group heard only replacement words that did not contain any example of /d/. The test task varied across experiments. Experiment 1 examined the effect of perceptual learning on spoken word recognition. Generalization of learning across the lexicon was examined by exposing participants to one set of words and testing them on a novel set. This study replicated the methods of Eisner et al. (2013) and extended this design to also ask whether perceptual adaptation results in changes in lexical competition. With a successful replication in Experiment 1, Experiments 2 and 3 were designed to provide a precise examination of perceptual changes throughout the phonetic category that lead to improved word recognition. Experiment 2 examined changes in the phonetic category boundary using a category identification task. Experiment 3 examined influences of learning on the internal

APA NLM
tapraid5/zfn-xhp/zfn-xhp/zfn01016/zfn3486d15z | xppws | S=1 | 8/9/16 | 0:29 | Art: 2015-0811

structure within the phonetic category by assessing listeners' goodness ratings of speech tokens as exemplars of each phonetic category (/d/ or /t/). Lastly, behavioral data were pooled across experiments and analyzed in combination with acoustic patterns of accented tokens to determine whether the reweighting of acoustic cues contributed to rapid perceptual adaptation to the foreign accent.

Experiment 1

In Experiment 1, we investigated whether native listeners can rapidly adapt to Mandarin-accented word-final /d/ pronunciations. Spoken word recognition was assessed by a cross-modal priming task. Notably, we expect an important difference in the acoustic–phonetic distribution of sounds in Mandarin-accented English compared to those sounds examined by past work (e.g., Kraljic & Samuel, 2005; McQueen, Cutler, & Norris, 2006; Reinisch & Holt, 2014; Sjerps & McQueen, 2010). In previous studies, ambiguous tokens of the to-be-adapted phoneme were created by artificially blending tokens from contrastive categories. As such, the ambiguous tokens were veridically "ambiguous" in that they fell in the middle of two contrastive phonetic distributions. In contrast, the burst dimension utilized by Mandarin talkers creates a perceptual contrast between a typical /t/ (intended as /d/ by the talker) and a perceptually hyperarticulated sound (intended as /t/ by the talker). Past research has shown that listeners can implicitly learn from the probabilistic or distributional cues in the speech signal in an unsupervised manner (e.g., Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kleinschmidt, Raizada, & Jaeger, 2015). Thus, if listeners are exposed to both /d/ and /t/ words, they may use /t/-final words as perceptual anchors, which may support adaptation to /d/-final words only because the accented /d/ tokens sound like poor exemplars of /t/ category. In this case, it would be difficult to pin down the informational source of adaptation: higher level lexical information or low-level acoustic distributional information. For this reason, we followed the design of Eisner et al. (2013) and did not include any /t/-final words during the exposure phase across all three experiments to examine the effects of lexically guided adaptation while controlling for contrastive, distributional information.

The fact that Mandarin-accented speakers produce clear distinctions between /d/ and /t/ along the burst dimension also led to one modification to the design of Eisner et al. (2013). To assess the actual amount of lexical competition during word recognition, we examined not only how auditory /d/-final words primed visual targets in an identity priming procedure (e.g., *seed–SEED*), but also how they primed phonological competitors of the intended targets (e.g., *seed–SEAT*). In Eisner et al. (2013), only *seat–SEAT* trials were examined, although one might infer that *seed–SEAT* condition would yield the same results, because word-final /d/s are devoiced and merged with /t/ as one category in Dutch-accented English. However, this is not the case for Mandarin-accented English. We hypothesized that exposure to Mandarin-accented /d/ tokens would increase the match between accented input and lexical forms, resulting in larger identity priming effect for Mandarin-accented /d/-final words (*seed–SEED* type) and smaller competitor priming (*seed–SEAT* type) by the experimental group compared to the control group. In addition, following learning, intended targets would have greater lexical activation than unintended competitors in the experimental group. Alternatively, because the lack of experience with a foreign accent provides listeners with no a priori prediction about the number of categories in the accent, it is hypothetically possible that exposure to /d/-final words (in the absence of /t/-final exposure words) causes listeners to merge the /d/ and /t/ categories. In this case, /d/ sounds will be treated equivalently to /t/ sounds, and then we would not only observe an increase of *seed–SEED* type of priming, but also an increase of *seed–SEAT* priming to a similar extent.

Method

Participants. Forty-eight monolingual English speakers with no hearing or visual problems were recruited from University of Connecticut. All participants were undergraduate students who were naïve to the Mandarin language and had no or minimal previous exposure to Mandarin-accented English. Participants were randomly assigned to one of the two exposure groups (experimental vs. control), with half of the participants assigned to each group. Across all experiments, participants received course credit or monetary reward for their participation, and gave informed consent according to the guidelines of the University of Connecticut Institutional Review Board.

Speech materials. A male native-Mandarin speaker recorded all speech materials. This speaker was a late second language learner of English and had resided in the United States for 18 months at the time of recording. Recordings were made in a soundproof room using a microphone onto a digital recorder, digitally sampled at 44.1 kHz and normalized for root mean square amplitude to 70 dB sound pressure level.

Exposure. For the experimental group, the exposure list consisted of 30 critical /d/-final words (e.g., *overload*), 60 filler words, and 90 nonwords. The list was identical for the control group except for the critical words. Instead of /d/-final words, there were 30 replacement words for the control group. The replacement words (e.g., *animal*) were matched to the critical -/d/ words in syllabic length and mean lemma frequency in CELEX (Baayen, Piepenbrock, & Gulikers, 1995). All words or nonwords were multisyllabic and contained three to four syllables. In both conditions, words were selected to meet the following criteria: (a) /d/ appeared only in word-final position, and only in critical words; (b) no other alveolar stops, no other voiced stops or dental fricatives, and no postalveolar affricates occurred; and (c) no voiceless stops (/p/ or /k/) occurred in word-final position. The same criteria were used in the selection of test stimuli.

Test. The test list was identical for both exposure groups. There were 60 monosyllabic /d/-final words (taken from /d/-/t/ minimal pairs such as *seed–seat*) and 180 monosyllabic filler words. Mean lemma frequencies in CELEX of the /d/- and /t/-final items were 83 (SD = 186) and 87 (SD = 126) per million, respectively, $t(59) = .159$, $p = .88$.

Procedure. Each participant completed an auditory lexical-decision task during exposure, which was immediately followed by a cross-modal priming task. A between-subjects design was used such that, during the exposure phase, the experimental group and the control group heard items from the experimental list and the control list, respectively. Items were presented in a random order. For the auditory lexical decision task, participants were instructed to decide whether each auditory stimulus was a real

English word and to press a yes/no button as quickly and accurately as possible.

The test phase was identical for both groups. Participants were told that they would see visual letter strings (targets) presented on the screen at the offset of the auditory words (primes) and they should respond with a yes/no button press whether the visual stimuli were real English words or not. On critical trials, 60 words from /d/-and /t/-final minimal pairs appeared as visual targets, in four different prime–target pairing types: /d/-final words as visual targets preceded by an identity prime (e.g., *seed–SEED*) or an unrelated prime (e.g., *fair–SEED*); /t/-final visual targets preceded by a minimal pair contrast (e.g., *seed–SEAT*) or an unrelated prime (e.g., *fair–SEAT*).Words in each set of minimal pair items were rotated over four counterbalanced lists. Within each list, there were equal proportions of the four pairing types. Noncritical trials were identical across counterbalanced lists: 30 filler words were paired with an identical prime (e.g., *foam–FOAM*) or an unrelated prime (e.g., *male–HORN*), and another 90 auditory filler words were paired with visual nonwords (e.g., *ring–WELF*). Thus, among the 180 trials in each list, half the targets were nonwords. The critical trials were evenly spaced in the test lists and for each list, there were two reverse test orders.

Stimuli were presented using Eprime 2.0.8 running on a desktop computer and delivered via Sennheiser HD280 headphones AQ: 2 (Sennheiser, USA) at a comfortable listening level that was held constant across participants; visual targets were shown in white Helvetica font in lower case on a black background in the center of the computer screen. Ten practice trials, which contained extra filler items, were given to the participants before the actual task in each phase. Exposure auditory items were presented with an interonset interval of 3,000 ms. During test, the intertrial interval was 1,400 ms, timed from the button press response to the onset of the next auditory prime. Visual targets were presented immediately at the offset of the auditory prime and stayed on the screen for 2 s unless terminated by a response. Reaction times (RTs) were measured from visual target onset. Participants were told to respond as fast as possible without sacrificing accuracy. Responses were made via keyboard with two buttons labeled "yes" and "no." Assignment of the "yes" button to the right or left hand was counterbalanced across participants.

Acoustic patterns. Figure 1 presents the distributional properties of the test stimuli (60 /d/-final and 60 /t/-final words) F1 produced by the Mandarin speaker (left panel) and a male native-English speaker (right panel; for comparison purpose). We measured three acoustic dimensions that are relevant for the voicing contrast. In the Mandarin-accented tokens, differences in vowel and closure durations were uninformative in cueing voicing (that is, /d/ and /t/ tokens were not separable along these two dimensions), whereas the difference in the burst release was striking. With minimal overlap between /d/ and /t/ distributions, bursts contained durational information that could be used to reliably differentiate voiceless from voiced tokens. This acoustic pattern stands in contrast with native-English productions (Figure 1, right panel) in which /d/ and /t/ tokens are reliably associated with long versus short preceding vowels, as well as discriminable closures, but not with very contrastive bursts. Note that /t/-final words only served as visual targets and never appeared as auditory primes in this experiment, but both /d/ and /t/ test items were presented in Experiments 2 and 3.

Results

Exposure. Response accuracy indicated that critical /d/-final words were largely judged to be real words by the experimental

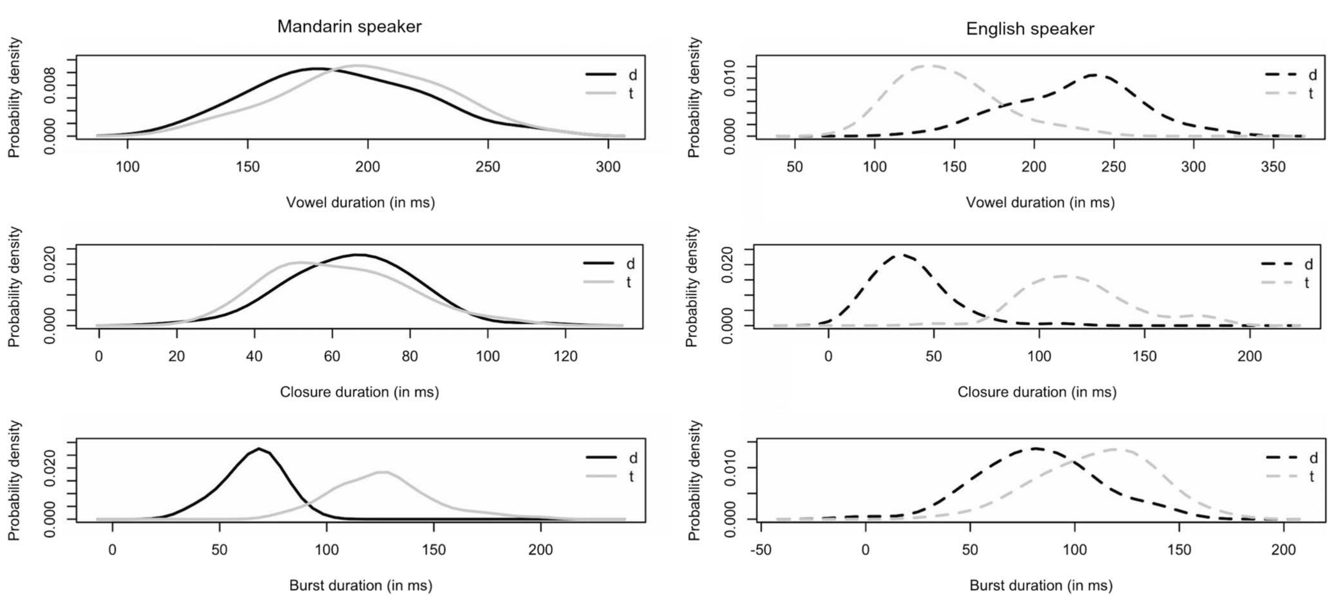

Figure 1. Probability density plots of acoustic measures (preceding vowel duration, closure duration and the length of burst and aspiration, respectively) across all 60 minimal pairs (/d/-final vs. /t/-final words). Left panel: Mandarin-accented speaker; right panel: native-English speaker. /t/-final words were only used in the test phase of Experiments 2 and 3.

APA NLM
tapraid5/zfn-xhp/zfn-xhp/zfn01016/zfn3486d15z | xppws | S=1 | 8/9/16 | 0:29 | Art: 2015-0811

group (M_1 = .81, SD_1 = .09) and that the replacement words were judged with equivalent accuracy by the control group ($M2$ = .81, SD_2 = .09). Response accuracy for filler words (M_1 = .80, SD_1 = .06; M_2 = .84, SD_2 = .05) and nonwords (M_1 = .77, SD_1 = .16; M_2 = .69, SD_2 = .17) were also comparable between the two groups.

T1 **Test.** Table 1 shows mean error rates and RTs in the test phase. Analysis of error rates did not reveal any group differences and were omitted from discussion here. RTs for correct responses were analyzed. Three words (*plod, moot, spate*) were discarded from further statistical analysis due to below-chance accuracy in related trials (remaining items had a mean of 93%). In addition, responses above or below 2 *SD*s from the mean of each prime type in each group were excluded from the RT analysis (4.5% of trials).

RT results were analyzed in a linear mixed-effects model, with *exposure group* (experimental vs. control), *target type* (/d/-final vs. /t/-final words), *prime type* (related vs. unrelated), and their interactions as fixed effects. Random effects included by-subject intercepts and by-item intercepts and slopes for priming type, which
Fn1 had the maximal random effect structure justified by the data[1] (Baayen, Davidson, & Bates, 2008). All the independent variables were contrast coded as follows: *exposure group*: experimental = 1, control = −1; *target type*: /d/-final targets = 1, /t/-final targets = −1; *prime type*: related = 1, unrelated = −1. As expected, related primes elicited faster responses than unrelated primes ($\beta = -23.23$, $SE = 2.48$, $p < .0001$). Meanwhile, /d/-final targets elicited slower responses than /t/-final targets ($\beta = 13.57$, $SE = 4.88$, $p < .01$). There was a significant three-way Exposure Group × Target Type × Prime Type interaction ($\beta = -4.11$, $SE = 2.11$, $p < .05$). No other effects were significant at the .05
F2 level. Figure 2 shows the RT priming magnitude (unrelated minus related) as a function of *exposure group* and *target type*.

To unpack the interaction, we fitted mixed-effects models for /d/-final targets and /t/-final targets separately, with *exposure group*, *prime type*, and their interaction as fixed effects. For /d/-final targets, there was a significant priming effect ($\beta = -27.44$, $SE = 3.27$, $p < .0001$); crucially, the priming effect was significantly larger in the experimental group than the control group, as revealed by the interaction effect ($\beta = -7.58$, $SE = 3.24$, $p < .05$). Thus, relative to control participants, participants in the experimental group who had been exposed to /d/-final words showed larger identity priming (e.g., *seed–SEED*) during test. Notably, critical test words were not heard during the exposure phase. Therefore, gains in identity priming reflected increased compatibility between accented /d/ tokens and phonetic representations, which generalized across the lexicon and facilitated subsequent word recognition. This result replicated findings of Eisner et al. (2013); they found an increase of identity priming as native-English listeners adapted to final-devoiced /d/ in Dutch-accented English. We added two novel types of critical trials (e.g., *seed–SEAT* and *fair–SEAT*) to Eisner et al.'s design to examine to what extent accented /d/ tokens activate the representation of /t/. For /t/-final targets, there was a priming effect ($\beta = -19.31$, $SE = 3.84$, $p < .0001$) but no Exposure Group × Prime Type interaction ($\beta = .46$, $SE = 2.73$, $p = .87$). This suggested that responses made to /t/-final words (e.g., *SEAT*) were faster following an auditory /d/-final word (e.g., *seed*) than following a phonologically unrelated word (e.g., *fair*), with no group difference in the absolute priming magnitude for /t/-final targets. Thus, exposure to critical /d/-final words increased priming for /d/-final targets in the experimental group, without decreasing priming for the voiceless competitor, /t/-final targets.

We further asked whether within each exposure group (experimental vs. control), auditory /d/-final words elicited larger priming for the intended item (e.g., *seed–SEED*) than for the phonological competitor (e.g., *seed–SEAT*). Starting with the control group, there was a significant priming effect ($\beta = -18.79$, $SE = 2.87$, $p < .0001$) but no Prime Type × Target Type interaction ($\beta = -.19$, $SE = 2.87$, $p = .95$). The absence of an interaction effect indicated that these listeners, who had not heard any examples of /d/ during exposure, activated the lexical representation of both the intended words (e.g., *seed*) and their phonological competitors (e.g., *seat*) almost equally. Thus, as predicted, Mandarin-accented /d/-final words were overall perceptually ambiguous for untrained native-English listeners. In contrast, in the experimental group, the main priming effect ($\beta = -26.98$, $SE = 3.65$, $p < .0001$) was modulated by a Prime Type × Target Type interaction ($\beta = -8.28$, $SE = 3.65$, $p < .05$), reflecting larger priming for the intended item (e.g., *seed–SEED*) than for the phonological competitor (e.g., *seed–SEAT*).

Discussion

In comparison with the control participants, the experimental participants, who had exposure to critical -/d/ words before test, showed stronger lexical activation of /d/-final words when hearing accented forms (e.g., *seed–SEED*); however, this increased activation was not at the expense of activation of /t/-final words (e.g., *seed–SEAT*). The influence of perceptual learning on the competing lexical target (e.g., *seed–SEAT*) was not assessed in Eisner et al. (2013). We now compare the current results to previous studies that investigated perceptual learning of atypical pronunciations embedded in one's *native* accent.

McQueen et al. (2006) showed a complete elimination of priming effect on phonological competitors (e.g., *doos–DOOF*, both are words in Dutch) after listeners adapted to the ambiguous fricatives (midway between /s/ and /f/) in a native Dutch accent (see also Sjerps & McQueen, 2010). The results were taken as evidence of thorough learning of the nonstandard sounds. However, we observed significant audio-to-visual priming for the phonological competitors of the intended words (*seed–SEAT* priming), indicating that the Mandarin-accented -/d/s did not fully function like native phonemes even after the critical exposure. The discrepancy between the current data and previous studies might arise for a number of reasons. First of all, in both McQueen et al. (2006) and Sjerps and McQueen (2010), the manipulated sound was the only unfamiliar sound that needed to be adapted to; the rest of stimuli were normal, clear native speech. In the current study, natural phonetic variation that deviates from the native norm was pervasive in the stimuli, in the sense that many other segments (e.g., vowels and other consonants) also bore traces of the non-native accent. The requirement of simultaneous adjustment to multiple phonetic categories may change the time course of complete

[1] The random effects in all models reported in the current study were determined by a stepwise variable selection procedure. The reported model always contained the maximal random effect structures justified by the data. We used the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015) to conduct the analysis. AQ: 8

Table 1
Mean Accuracy and Reaction Time (RT) Across Participants in the Cross-Modal Priming Task as a Function of Exposure Group in Experiment 1

	/d/-final		/t/-final	
Exposure condition	Related prime *seed–SEED*	Unrelated prime *fair–SEED*	Related prime *seed–SEAT*	Unrelated prime *fair–SEAT*
Mean error (%)				
Experimental	10 (7)	15 (13)	7 (7)	12 (10)
Control	9 (8)	18 (11)	6 (6)	13 (7)
Mean RT (ms)				
Experimental	598 (89)	672 (115)	593 (96)	630 (100)
Control	577 (89)	616 (92)	553 (87)	588 (81)

Note. *SD* is given in parentheses.

learning. In addition, in these two studies (cf. Eisner et al., 2013), listeners were exposed to unambiguous tokens of the contrasting segment in addition to the ambiguous sound. We intentionally excluded /t/-final words during exposure to control for any adaptation effects resulting from distributional learning in the presence of contrastive cues to /d/ versus /t/ words (as illustrated in Figure 1). Had listeners been exposed to /t/-final words as well, they may have learned to associate the acoustic distributions of /t/ tokens with the /t/ category, and potentially achieved faster and more complete adaptation. Finally, the overall intelligibility of our exposure words was relatively lower than native speech. As a result, increased effort in processing these words may have weakened the activation of these lexical entries, which would in turn weaken the retuning influence on the ambiguous phoneme. In Experiments 2 and 3, we sought to provide a more precise indication of prelexical changes, by examining changes in the location of phonetic boundary between categories and in the internal structure within the categories.

Experiment 2

In natural speech, category membership is often determined not by a single acoustic dimension, but by the combination of multiple acoustic cues. Each acoustic dimension has its own distributional characteristics and is differentially informative about phonetic segment identity (e.g., Flege et al., 1992). As shown in Figure 1, the acoustic properties of /d/- and /t/-final words were quite distinct from native speech when naturally produced by a Mandarin speaker. In Experiment 2, we determined whether exposure to nonnative accented tokens with different distributional properties than native-accented speech results in a shift in category boundary. The /d/-final and /t/-final minimal pairs served as test stimuli in Experiment 2. Following an exposure phase that was identical to that in Experiment 1, we assessed potential changes in the phonetic category boundary using a two-alternative, forced-choice category identification task during test. A phonetic boundary shift would be indicated by an increase in /d/ responses for /d/-final words for the experimental group compared to the control group.

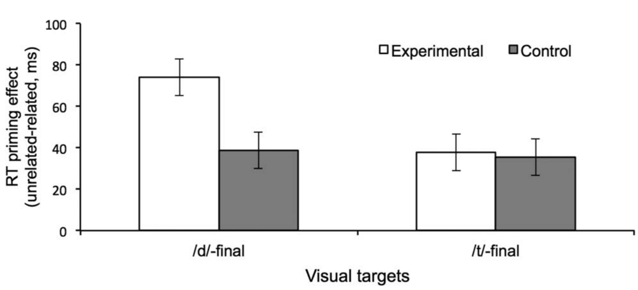

Figure 2. Experiment 1 test results: Priming of /d/-final words (reaction time [RT] in *fair-SEED* trials minus RT in *seed-SEED* trials) and /t/-final words (RT in *fair-SEAT* trials minus RT in *seed-SEAT* trials) for participants exposed to critical words (experimental group) or replacement words (control group). Error bars represent standard error of the mean.

Method

Participants. Forty-eight monolingual English speakers with no hearing or visual problems were recruited from University of Connecticut. Participants were randomly assigned to one of the two exposure groups (experimental vs. control, $n = 24$ in each condition).

Speech materials. The exposure stimuli were identical to those used in Experiment 1. The test list included 60 monosyllabic minimal pairs ending in /d/ or /t/ (e.g., *seed–seat*; identical to the /d/-final words that appeared in Experiment 1 during test as auditory primes). The test stimuli were organized into two blocks such that if *seed* appeared in block 1, *seat* appeared in block 2. Each counterbalanced block consisted of 30 /d/-final words and 30 /t/-final words; items were presented in random order within each block.[2] Compared to past studies that tested phonetic categorization of a speech continuum (e.g., /ada/ to /ata/; Kraljic & Samuel, 2006), the current test stimuli did not have an equal number of endpoints or equal spacing between members of a phonetic category. However, acoustic analyses of the speech tokens indicated that the stimuli spanned the /d/-/t/ continuum and had extreme tokens at both ends, which was perceptually confirmed with an identification pretest.

Procedure. During both phases, participants were told to respond as fast as possible without sacrificing accuracy. The exposure phase was identical to that used in Experiment 1. During the

[2] Note that while /t/-final words were never heard by listeners in Experiment 1, in this experiment listeners heard /t/-final words during the test phase (two-alternative, forced-choice task). However, because only minimal pairs were used at test, there was no top-down lexical information indicating the identity of the intended category (/d/ or /t/).

test phase, test items were presented with an intertrial interval of 2,000 ms. Participants were asked to identify the final consonant of each item as either /d/ or /t/ by pressing an appropriately labeled button. No feedback was provided.

Results

As shown in Figure 1, there was a fair amount of acoustic variability in the /d/ and /t/ tokens. Consistent with the acoustic data, the categorization results showed that the test items varied in the degree of their perceptual ambiguity (percent /d/ responses ranged from 0 to 100 for both /d/-final and /t/-final words). A mixed-effects logit model was used to analyze the category iden-
F3 tification data (see Figure 3). Mixed logit models predict the probability of a particular response (here, a /d/ response; Jaeger,
AQ: 3 2008). Fixed effects were *exposure group* (experimental vs. control) and *word type* (/d/-final vs. /t/-final) as well as their interaction. By-item intercepts and by-subject intercepts and slopes for word type were included as random effects. The independent variables were contrast coded as follows: *exposure group*: experimental = 1, control = −1; *word type*: /d/-final = 1, /t/-final = −1. For the dependent measures, /d/ responses were coded as 1 and /t/ responses were coded as 0. Positive log coefficients indicate a log odds ratio greater than 0 (corresponding odds ratio is greater than 1), which means that the level coded as 1 has greater probabilities of /d/ responses than the level coded as −1. Overall, /d/-final words (64%) elicited significantly more /d/ responses than /t/-final words (32%) across the two groups ($\beta = 1.00$, $SE = .16$, $p < .0001$). Crucially, there was a significant group effect ($\beta = .22$, $SE = .09$, $p < .05$): The experimental group reported significantly more /d/ responses (51%) than the control group (45%) overall. There was no *exposure* Group × Word Type interaction ($\beta = .08$, $SE = .08$, $p = .28$). The category identification results taken as a whole indicate that the experimental group tended to interpret more words (both /d/-final and /t/-final) as ending in /d/ than the control group, suggesting a general boundary shift toward the /t/-end along a /d/-/t/ continuum.

Discussion

Experiment 2 replicated previous findings on rapid perceptual learning of ambiguous sounds (Kraljic & Samuel, 2005; Norris et al., 2003; Reinisch & Holt, 2014): There was an increase in AQ: 4
identification of noncanoncial sounds as members of the recalibrated category. Crucially, because the only difference between the two groups was the presence versus absence of critical /d/ words during exposure, the increased /d/ reports in the experimental group relative to the control group reflected learning during the exposure phase. We did not observe any Exposure Group × Word Type interaction, suggesting that the learning is driven by the absorption of ambiguous tokens near the boundary into the /d/ category, rather than by enhanced discrimination between /d/ and /t/.

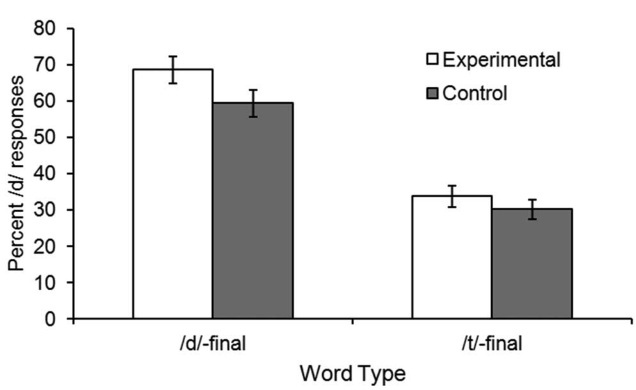

Figure 3. Mean percent /d/ responses for the two-alternative, forced-choice category identification task in Experiment 2 as a function of exposure group and word type. Error bars indicate standard errors of the mean.

Experiment 3

The objective of Experiment 3 was to investigate learning consequences on the internal structure of phonetic categories. Perceived goodness of accented tokens was assessed by a goodness rating task, which tapped into listeners' sensitivity to phonetic detail in a more graded way than categorical membership. Theodore and colleagues (2015) trained participants to associate a talker's characteristic voiceless stop consonant production (e.g., /k/ in *cane*) with either short or long VOTs. Listeners were subsequently asked to give goodness ratings of the speaker's tokens along a voiced–voiceless continuum. Their results showed that participants who were trained with long VOTs judged tokens with longer VOTs to be "better" exemplars than those who were trained with short VOTs. Importantly, such changes reflected reorganization within a category independent of any between-category boundary shifts. Here, using a rapid phonetic adaptation paradigm, we asked if perceptual learning of a non-native accent would have similar influences on the phonetic structure by changing the way in which fine-grained phonetic variation in foreign-accented tokens are perceived by native listeners. If so, we expected to see group differences with respect to goodness ratings of accented tokens as exemplars of the intended phonetic categories.

Method

Participants. Forty-eight monolingual English speakers with no hearing or visual problems participated in the experiment (with 24 participants in both the experimental and the control conditions).

Speech materials. The exposure and test stimuli were identical to those used in Experiment 2.

Procedure. The exposure phase was identical to that used in Experiments 1 and 2. Following the exposure phase, a practice phase performed that was designed to accustom listeners to performing goodness judgments on an isolated segment in the context of globally accented speech. Participants heard 10 words ending in consonants ranging from a good /m/ sound to a poor /m/ sound (perceived as /n/) spoken by the accented test talker. Participants were asked to rate the goodness of the final consonant on a scale from 1 (*very poor example of /m/*) to 7 (*very good example of /m/*). We reasoned that despite of the strong accent in all words, if a participant gave high ratings for words ending in clear /m/ but low rating for words that were perceived to end in /n/, then he or she understood the task. An experimenter was present during this practice phase to make sure that participants followed the procedure. Thus, the practice helped to ensure that listeners were rating

APA NLM
tapraid5/zfn-xhp/zfn-xhp/zfn01016/zfn3486d15z | xppws | S=1 | 8/9/16 | 0:29 | Art: 2015-0811

goodness of the final consonant of each item, and not the degree of accentedness of the whole word.

During the test phase, the test stimuli were divided into two sets and were administered in two blocks. The words from a minimal pair did not occur within the same block. To examine how exposure to /d/-final words influenced perceived goodness of /d/ tokens as members of the /d/ category, participants were asked to rate the final consonant of each item in terms of how good it was as an exemplar of a /d/ by pressing an appropriately labeled button. To examine potential influences from the critical exposure on the representation of phonetic category on the other end of the voicing continuum, we also asked listeners to rate each item for goodness as /t/. Goodness as /d/ served as the primary dependent measure, we consider results from goodness as /t/ a source to provide complementary information about the perceptual changes along the entire voicing continuum. Participants rated goodness as /d/ in one block and goodness as /t/ in another block. The allocation of test sets and the order of blocks were counterbalanced across participants. Each block consisted of 30 /d/-final words and 30 /t/-final words; within each block, items were presented in a random order. Participants were asked to rate each item on a scale from 1 (*very poor example*) to 7 (*very good example*). Auditory items were presented with an intertrial interval of 2,000 ms. No feedback was provided.

Results

To accommodate individual variability and potential rating bias and to understand the relative rating of each item, participants' raw ratings were transformed into standardized z-scores that were used
F4 in subsequent analyses. Figure 4 presents the mean standardized ratings for each task as a function of *exposure group*. Note that for each rating task (goodness as /d/ and goodness as /t/, separately), this within-subjects standardization procedure makes each participant's mean rating across all test items zero; the mean rating for /t/-final words is necessarily the additive inverse of that for /d/-final words. Thus, we only present the mean rating for /d/-final words in the goodness-as-/d/ task and mean rating for /t/-final words in the goodness-as-/t/ task. However, in the statistical analysis as reported below, both /d/-final words and /t/-final words were included for each task because all the reported effects were taken as random at the item level.

Goodness as /d/. A linear mixed-effects model was fitted with *exposure group, word type*, and Group × Word Type interaction as fixed effects. By-item intercepts were included as random effects. All independent variables were contrast-coded as in Experiment 2. There was no overall group difference in the ratings ($p = .99$). While /d/-final words received higher ratings than /t/-final words across both groups ($\beta = .42$, $SE = .04$, $p < .0001$), the learning effect took the form of an interaction between exposure group and word type ($\beta = .05$, $SE = .01$, $p < .001$): Relative to the control group, the experimental group rated /d/-final words as better examples of /d/ and rated /t/-final words as poorer examples of /d/.

Goodness as /t/. Similarly, /t/-final words received higher ratings than /d/-final words ($\beta = -.48$, $SE = .05$, $p < .0001$). There was no group effect ($p = .99$). Again, the learning effect was revealed in the Group × Word Type interaction ($\beta = -.04$, $SE = .01$, $p = .005$): Relative to the control group, the experimental group rated /t/-final words as better examples of /t/ and rated /d/-final words as poorer examples of /t/.

Discussion

In both goodness rating tasks, the experimental group (relative to the control group) assigned higher ratings to words in accordance with the intended word type. Crucially, although the experimental group was exposed to /d/-final words only, the exposure affected their judgment of speech tokens as exemplars of the /t/ category: /t/-final words were perceived to be better /t/s among experimental participants than control participants. Note that the categorization results in Experiment 2 showed that our test stimuli varied over a wide range in their ambiguity. Some items fell unambiguously into the unintended category; some fell into the ambiguous niche; others were clear, albeit non-native, tokens of the intended category. If the perceptual changes following exposure were limited to the boundary region, we would not be likely to observe a global improvement in the perceived goodness of speech tokens as members of the intended category (/d/ or /t/). In fact, in Experiment 2, exposure to /d/ productions led to more /d/ responses to both /d/-final and /t/-final words in categorization. If those ambiguous tokens near the boundary that drove the change in the boundary location also drove the group difference in the perceived goodness, /t/-final words would be rated as "better" as

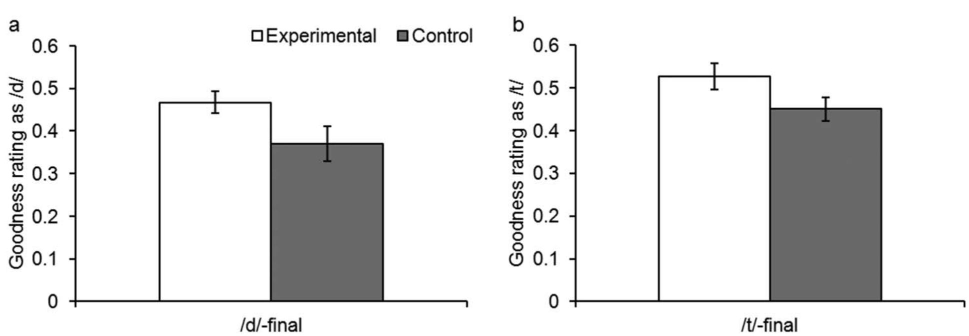

Figure 4. Mean goodness ratings (*z* transformed) of each word type (/d/-final words and /t/-final words) as exemplars of (a) the /d/ category and (b) the /t/ category as a function of each exposure group in Experiment 3. Error bars indicate standard errors of the mean.

APA NLM

tapraid5/zfn-xhp/zfn-xhp/zfn01016/zfn3486d15z | xppws | S=1 | 8/9/16 | 0:29 | Art: 2015-0811

members of the /d/ category. In contrast, in Experiment 3, the experimental group gave higher goodness as /t/ to /t/-final words than the control group. The fact that an overall improvement of perceived goodness was found in aggregate across the /d/ category and that the learning effect extended to untrained phonetic category (/t/) suggests that listeners were not merely incorporating /d/ tokens into the intended segmental category (e.g., Norris et al., 2003). Experiments 2 and 3 together revealed that perceptual adaptation to the Mandarin-accented speaker had consequences for the global phonetic structure of the alveolar stop contrast in word-final position: In concert with a phonetic boundary shift (Experiment 2), listeners also adjusted the internal structure within each phonetic category for the /d/-/t/ contrast.

Phonetic Adjustment: A Reweighting of Acoustic Cues

Note that exposure to /d/ tokens affected the phonetic representation of the /t/ category. Presumably, representations of internal structure of the two categories could be independent from each other and the goodness judgment task allows us to independently assess the structure within these two categories. That is, while in the categorization task, more /d/ responses necessarily translate to fewer /t/ responses, a token given lower ratings as /d/ does not necessarily receive higher ratings as /t/ if it is not typical of the speaker's productions. Given this, why is there a "carryover" effect?

Although it is implied in perceptual learning studies that listeners are sensitive to talker-specific distribution of acoustic cues, few studies have explicitly investigated the specific acoustic–phonetic properties associated with learning, other than showing that listeners learned general information such as "this speaker produces odd /d/ tokens" (Kraljic & Samuel, 2006). Thus, relatively little empirical data is available to inform the exact information source of the sound-to-category remapping process (see Reinisch, Wozny, Mitterer, & Holt, 2014, for an examination of specific acoustic cues in visually cued phonetic recalibration). Outside the domain of talker-related perceptual learning, studies of speech categorization suggest that listeners are generally sensitive to the statistical values of critical acoustic properties in the speech input (e.g., Clayards et al., 2008). More importantly, training with category-level feedback can shift listeners' attention to more informative acoustic cues over less informative ones as they learn non-native phonetic contrasts (Chandrasekaran, Sampath, & Wong, 2010; Francis, Baldwin, & Nusbaum, 2000; Francis & Nusbaum, 2002). Of note, several computational models of perceptual learning suggest that it is a general property of the perceptual system to weigh acoustic cues differentially as a function of their informativeness in distinguishing phonetic categories (Kleinschmidt & Jaeger, 2015; Toscano & McMurray, 2010). Here, we consider a similar mechanism that may underlie listeners' reorganization of phonetic structure of /d/ and /t/ in the rapid adaptation to the Mandarin-accented speaker.

Consistent with language-related production differences, English and Mandarin listeners differ in their use of temporal cues to identify voicing in stop consonants (e.g., Crowther & Mann, 1992). The appropriate use of informative cues has been linked to enhanced intelligibility of foreign-accented speakers (Xie & Fowler, 2013). We thus hypothesized that adaptation to the accent, and in particular, the adjustment of internal structures of /d/ and /t/ categories, is achieved via an adjustment in the weighting of various acoustic cues for the accent. To test this hypothesis, we assessed behavioral responses in the experimental and control groups of Experiments 2 and 3 as a function of the acoustic properties of the speech materials (see Figure 1).

We used mixed-effects regression models to predict the categorization and goodness judgment responses by including *exposure group* and the three temporal measures (duration of *vowel*, *closure* and *burst*) as well as their interactions (between exposure group and each acoustic measure) as predictors. Subjects and items were considered random effects. Predictors were standardized (converting into z scores) before they were entered into the regression model. Vowel duration was weakly correlated with closure duration, $r = -0.23$, $p < .05$, and was not correlated with burst duration, $r = .10$, $p = .27$; there was no significant correlation between closure and burst duration either, $r = -0.15$, $p = .09$. We present the regression results for the category identification and goodness ratings separately (see Table 2). The predictive power of T2

Table 2
Results Across Experiments 2 and 3: Estimated Probability of /d/ Responses (Experiment 2) and Goodness of /d/ and /t/ Tokens (Experiment 3) as a Function of Temporal Acoustic Cues

	Experiment 2		Experiment 3			
	Probability of /d/ responses		Goodness as /d/		Goodness as /t/	
Predictor	Log coefficient	p value	Coefficient	p value	Coefficient	p value
Intercept	−.08 (.16)	.63	−.002 (.04)	.96	−.002 (.05)	.97
Group	.21 (.09)	.02*	−5E05 (.01)	.99	–1E04 (.01)	.99
Vowel	.29 (.14)	.03*	.09 (.04)	<.05*	−.05 (.05)	.35
Closure	−.54 (.14)	<.001**	−.20 (.04)	<.0001**	.18 (.05)	<.001**
Burst	−.93 (.14)	<.0001**	−.42 (.04)	<.0001**	.43 (.05)	<.0001**
AQ: 9 Group × Vowel	−.05 (.03)	.15	−.01 (.02)	.50	–1E04 (.01)	.99
Group × Closure	−.04 (.04)	.25	−.01 (.02)	.52	−.008 (.01)	.59
Group × Burst	−.06 (.04)	.10	−.03 (.01)	.05*	.03 (.01)	.01*

AQ: 10 *Note.* *SE* is given in parentheses.
* $p < .05$. ** $p < .01$.

tapraid5/zfn-xhp/zfn-xhp/zfn01016/zfn3486d15z | xppws | S=1 | 8/9/16 | 0:29 | Art: 2015-0811
APA NLM

the acoustic cues reveals how informative they are (i.e., the perceptual weighting) in determining phonetic membership and category typicality; the interaction with exposure group reveals changes, if any, in the relative weighting as a result of exposure to critical words.

Category Identification (Experiment 2)

There was a main effect of exposure group, with the experimental group reporting more /d/s throughout the acoustic continuum ($p < .05$). The main effects of all three measures were significant: vowel duration, $p < .05$; closure, $p < .001$; burst, $p < .0001$. Although none of the interaction terms were statistically significant, there was a trend ($p = .10$) for the experimental group to rely more on the burst in making categorization decisions than the control group.

Goodness Ratings (Experiment 3)

Closure and burst information predicted goodness rating as both /d/ and /t/ (ps $< .001$); the main effect of vowel duration was significant for goodness as /d/ ($p < .05$) but not for goodness as /t/ ($p = .35$). Interestingly, there was a significant Exposure Group × Burst interaction for goodness ratings for both categories (/d/: $p = .05$; /t/: $p = .01$); the signs of coefficients suggest that the experimental group relied on the length of the burst more heavily than the control group in judging goodness.

Together, evidence suggested a heavier weighting of burst information by the experiment group across all three tasks, although goodness ratings were more sensitive than categorization responses in detecting changes in the cue-weighting functions. The results from the control group provided a gauge of initial cue use without exposure to the critical words: The main effects of the acoustic predictors revealed that listeners were generally sensitive to multiple cues, even the cues that are not typically used for native contrasts. For the experimental group, the variation of acoustic dimensions of critical exposure words further guided their attention to the most informative cue (burst durations) when tested with novel stimuli. The Group × Burst interaction on perceived goodness clearly suggested that the internal structure of the phonetic categories were reorganized as a result of a reweighting of acoustic parameters. Increased attention toward this acoustic dimension can explain the unexpected adjustment in the internal structure of the /t/ category: In other words, attention was increased toward a previously underweighted cue (burst duration) for identifying /d/ tokens as guided by corrective feedback (lexical information) during exposure; this simultaneously resulted in a relatively less reliance on vowel length as a predictor for voicing contrast (see Kruschke, 2003, 2011, for theoretical discussions on such attentional learning). This result parallels empirical findings of cue-weighting changes in second language acquisition, although learning of nonnative phonetic contrasts requires more training and occurs over longer time scales (see Francis et al., 2000). In the case of foreign accent adaptation, our results showed that exposure to only one of the contrastive categories was sufficient to elicit some changes in the perceptual weighting. These results are consistent with distributional learning models, which suggest that listeners make use of established native categories and adjust low-level cue-weighting as guided by top-down information (e.g., Kleinschmidt & Jaeger, 2015).

General Discussion

In the present research, we used the lexically informed perceptual learning paradigm (e.g., Norris et al., 2003) to study native listeners' adaptation to acoustic–phonetic variation in natural foreign accents. With respect to previous work that has focused on the recalibration of phonetic boundaries, our first finding is that listeners also reorganized the internal structure of a phonetic category during rapid perceptual learning of a specific talker's accent. In addition to the incorporation of *ambiguous* tokens into a recalibrated category, even *unambiguous* tokens beyond the boundary region were perceived as better exemplars of the intended categories (/d/ and /t/) as a result of talker-specific adaptation. This finding adds to existing literature that characterizes a highly adaptive native perceptual system; listeners can exhibit a small boundary shift to accommodate ambiguous tokens (e.g., Norris et al., 2003), a bigger boundary shift in the face of a sound that falls unambiguously into an unintended category (e.g., Sumner, 2011), or structural adjustment within a category as shown here. More broadly, the adaptive ability in updating within-category structure may enable listeners to readily adapt to unfamiliar idiolects, dialects or even nonnative accents in which speech tokens do not cause cross-category confusion. Reorganization within the category proper may be sufficient to improve online speech processing in these scenarios, enabling not only disambiguation of tokens near the category boundary, but facilitating lexical access for tokens near the category center.

Relatedly, our second finding is that listeners showed sensitivity to the precise acoustic cues that were used in phoneme specification and used them to reorganize the internal structure of phonetic categories. This finding is consistent with the "Attention to Dimension" model (Francis & Nusbaum, 2002), which suggests that listeners update attentional weights assigned to relevant acoustic dimensions when learning a new language. We provided novel evidence that native listeners dynamically update their own cue-weighting functions during rapid phonetic adaptation to foreign accents, and critically over much shorter time span than shown in previous studies of second language phoneme learning (e.g., Chandrasekaran et al., 2010; Francis & Nusbaum, 2002).

These findings together are compatible with the predictions of probabilistic frameworks (e.g., Feldman et al., 2013; Kleinschmidt & Jaeger, 2015), which postulate top-down guided probabilistic learning in speech perception and phonetic learning. A theoretically important question concerns the functional role of bottom-up input in phonetic adaptation. Of note, past research has demonstrated that listeners could fully adapt to ambiguous native-accented speech (cf. McQueen et al., 2006) within the time frame assessed in the current study; yet, we did not observe "complete" adaptation for Mandarin-accented English. The Mandarin-accented spoken words did not function fully as native words in the sense that robust lexical activation was still observed for phonetically similar words (Experiment 1). Following adaptation, accented /d/-final words (e.g., *seed*, pronounced like *seat*) activated the intended word form more strongly; however, phonetically related competitors (*seat* for *seed*) were not eliminated from consideration. A number of reasons might explain why potentially "more" adaptation is needed for foreign-accented speech than for native speech variants. Of particular interest is to what extent the low-level acoustic properties versus weaker top-down lexical support accounts for the "incomplete" adaptation. It is possible that the lower intelligibility of foreign-accented words, relative to native-accented words, provided poorer lexical support for the phonetic retuning to occur. Alternatively, it could be that case that if listeners had heard both /d/-final and /t/-final words during the exposure phase, the fine-grained differences in burst durations between /d/ and /t/ words would facilitate an attentional shift to this acoustic dimension and thus expedite the adaptation process. It is another empirical question whether similarly rapid adaptation would occur without top-down lexical guidance when the bottom-up input provides salient enough information; and when it does, what features of the bottom-up input (e.g., the amount of sampling in the perceptual space, the distance of deviant tokens from prototypical tokens) affect the speed of adaptation (Clayards et al., 2008; Kleinschmidt et al., 2015). These questions are important for a comprehensive understanding of how bottom-up and top-down mechanisms work together to support speech adaptation (see Sjerps & Reinisch, 2015).

To conclude, perceptual learning is not just a matter of adjusting phonetic boundaries in face of ambiguous tokens; it also instigates a reorganization of the internal structure of phonetic categories. Furthermore, for foreign-accented tokens, learning may result in changes in cue-weighting functions that can prepare listeners for adapting to similar variation in other acoustic environments. Future research should investigate to what extent the observed adjustment in the phonetic structure and cue-weighting strategy are context-specific (e.g., speakers, accents, phonemes).

References

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America, 115,* 3171–3183. http://dx.doi.org/10.1121/1.1701898

Baayen, H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4 (R package version 1.1–7) [Computer software]. Retrieved from http://CRAN. R-project.org/package=lme4

Bent, T., Bradlow, A. R., & Smith, B. L. (2008). Production and perception of temporal patterns in native and non-native speech. *Phonetica, 65,* 131–147. http://dx.doi.org/10.1159/000144077

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America, 128,* 456–465. http://dx.doi.org/10.1121/1.3445785

Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics, 70,* 604–618. http://dx.doi.org/10.3758/PP.70.4.604

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108,* 804–809. http://dx.doi.org/10.1016/j.cognition.2008.04.004

Crowther, C. S., & Mann, V. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *The Journal of the Acoustical Society of America, 92,* 711–722. http://dx.doi.org/10.1121/1.403996

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America, 119,* 1950–1953. http://dx.doi.org/10.1121/1.2178721

Eisner, F., Melinger, A., & Weber, A. (2013). Constraints on the transfer of perceptual learning in accented speech. *Frontiers in Psychology, 4,* 148. http://dx.doi.org/10.3389/fpsyg.2013.00148

Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review, 120,* 751–778. http://dx.doi.org/10.1037/a0034245

Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *The Journal of the Acoustical Society of America, 92,* 128–143. http://dx.doi.org/10.1121/1.404278

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics, 62,* 1668–1680. http://dx.doi.org/10.3758/BF03212164

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 349–366. http://dx.doi.org/10.1037/0096-1523.28.2.349

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122,* 148–203. http://dx.doi.org/10.1037/a0038695

Kleinschmidt, D., Raizada, R., & Jaeger, T. F. (2015). Supervised and unsupervised learning in phonetic adaptation. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th annual conference of the cognitive science society* (pp. 1129–1134). Austin, TX: Cognitive Science Society.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51,* 141–178. http://dx.doi.org/10.1016/j.cogpsych.2005.05.001

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13,* 262–268. http://dx.doi.org/10.3758/BF03193841

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science, 12,* 171–175. http://dx.doi.org/10.1111/1467-8721.01254

Kruschke, J. K. (2011). Models of attentional learning. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 120–152). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511921322.006

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50,* 93–107. http://dx.doi.org/10.3758/BF03212211

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30,* 1113–1126. http://dx.doi.org/10.1207/s15516709cog0000_79

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics, 46,* 505–512. http://dx.doi.org/10.3758/BF03208147

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47,* 204–238. http://dx.doi.org/10.1016/S0010-0285(03)00006-9

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Attention, Perception & Psychophysics, 15,* 285–290. http://dx.doi.org/10.3758/BF03213946

Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental*

Psychology: Human Perception and Performance, 40, 539–555. http://dx.doi.org/10.1037/a0034409

Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics, 45,* 91–105. http://dx.doi.org/10.1016/j.wocn.2014.04.002

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics, 31,* 307–314. http://dx.doi.org/10.3758/BF03202653

Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 195–211. http://dx.doi.org/10.1037/a0016803

Sjerps, M. J., & Reinisch, E. (2015). Divide and conquer: How perceptual contrast sensitivity and perceptual learning cooperate in reducing input variation in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 41,* 710–722. http://dx.doi.org/10.1037/a0039028

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition, 119,* 131–136. http://dx.doi.org/10.1016/j.cognition.2010.10.018

Theodore, R. M., & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *The Journal of the Acoustical Society of America, 128,* 2090–2099. http://dx.doi.org/10.1121/1.3467771

Theodore, R. M., Myers, E. B., & Lomibao, J. A. (2015). Talker-specific influences on phonetic category structure. *The Journal of the Acoustical Society of America, 138,* 1068–1078. http://dx.doi.org/10.1121/1.4927489

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science, 34,* 434–464. http://dx.doi.org/10.1111/j.1551-6709.2009.01077.x

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *The Journal of the Acoustical Society of America, 92,* 723–735. http://dx.doi.org/10.1121/1.403997

Vong, W. K., Navarro, D. J., & Perfors, A. (2016). The helpfulness of category labels in semi-supervised learning depends on category structure. *Psychonomic Bulletin & Review, 23,* 230–238. http://dx.doi.org/10.3758/s13423-015-0857-9

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica, 64,* 122–144. http://dx.doi.org/10.1159/000107913

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics, 32,* 251–276. http://dx.doi.org/10.1016/S0095-4470(03)00032-9

Xie, X., & Fowler, C. A. (2013). Listening with a foreign-accent: The interlanguage speech intelligibility benefit in Mandarin speakers of English. *Journal of Phonetics, 41,* 369–378. http://dx.doi.org/10.1016/j.wocn.2013.06.003

Received September 21, 2015
Revision received June 6, 2016
Accepted June 11, 2016 ■