Journal of Experimental Psychology:
Human Perception and Performance
2015, Vol. 41, No. 6, 1680–1695

© 2015 American Psychological Association
0096-1523/15/$12.00 http://dx.doi.org/10.1037/xhp0000113

Sleep and Native Language Interference Affect Non-Native Speech Sound Learning

F. Sayako Earle
University of Connecticut

Emily B. Myers
University of Connecticut and Haskins Laboratories, New Haven, Connecticut

Adults learning a new language are faced with a significant challenge: non-native speech sounds that are perceptually similar to sounds in one's native language can be very difficult to acquire. Sleep and native language interference, 2 factors that may help to explain this difficulty in acquisition, are addressed in 3 studies. Results of Experiment 1 showed that participants trained on a non-native contrast at night improved in discrimination 24 hr after training, while those trained in the morning showed no such improvement. Experiments 2 and 3 addressed the possibility that incidental exposure to perceptually similar native language speech sounds during the day interfered with maintenance in the morning group. Taken together, results show that the ultimate success of non-native speech sound learning depends not only on the similarity of learned sounds to the native language repertoire, but also to interference from native language sounds before sleep.

Keywords: speech perception, perceptual learning, sleep, memory consolidation, second language acquisition

Non-native speech sounds are difficult for adults to perceptually disambiguate, particularly if these sounds are similar to sounds in the existing native language phonology (see Strange, 1995, for review). For example, the Hindi dental /d̪/ and retroflex /ɖ/ sounds are often perceived by native English speakers as variants of the English alveolar /d/ category (Werker & Lalonde, 1988). Previous accounts have focused on limitations in processing these sounds, suggesting that similarity to native-language perceptual or articulatory representations may prevent listeners from distinguishing novel non-native tokens from native speech sounds (Best, 1995; Flege, 1995; Kuhl & Iverson, 1995). However, little is known about difficulties that may arise due to failures in encoding learned variants into long-term memory following speech sound training. The formation of novel speech sound categories requires that listeners both encode details of these sounds in memory, as well as abstract away from episodic details to recognize new instances of the sound (see Earle & Myers, 2014, for review). Given this, understanding the role of consolidation, that is, the memory process that facilitates these qualitative changes to the memory trace, not only contributes to accounts of speech sound learning, but provides broader insight into the emergence of perceptual categories.

Sleep in Memory Consolidation

The contribution of sleep to memory consolidation is supported by a growing literature (see Rasch & Born, 2013, for review), but few studies have directly investigated how sleep affects perceptual learning as it relates to speech, arriving at different conclusions depending on the aspect of speech learning that is assessed (see Earle & Myers, 2014, for review; Eisner & McQueen, 2006; 2013; Fenn et al., 2003; Roth et al., 2005). In particular, some studies show no significant sleep-mediated influences on the maintenance/stability of learned phonetic information. For example, Eisner and McQueen (2006) found that shifts in category boundary to accommodate speaker idiosyncrasies emerged immediately after perceptual training, and remained stable over a posttraining interval of 24 hours irrespective of when sleep occurred in relation to training. Similarly, Roth et al. (2005) found that a period of restful wake, as well as sleep, stabilized the training-induced performance gain on the identification of syllables in noise. It should be noted that the sleep group alone showed a trend toward higher performance at the delayed posttest, suggesting that a larger sample size may have yielded a statistically significant improvement as a function of sleep.

In contrast, a separate set of studies suggests that sleep facilitates the recovery of learned perceptual information, and assists in generalization to new instances. Fenn et al. (2003; 2013) trained individuals to identify synthetically generated words, a task which requires that these nonstandard phonetic tokens be mapped onto the listener's native phonology. In their case (Fenn et al., 2003),

This article was published Online First August 17, 2015.

F. Sayako Earle, Department of Speech, Language, and Hearing Sciences, University of Connecticut; Emily B. Myers, Department of Speech, Language, and Hearing Sciences and Department of Psychology, University of Connecticut; and Haskins Laboratories, New Haven, Connecticut.

This work was supported by National Institutes of Health, National Institute on Deafness and Other Communication Disorders (NIH NIDCD) Grants R03 DC009495, R01 DC013064, and P01 HD001994. The content is the responsibility of the authors and does not necessarily represent official views of the NIH, NIDCD, or The Eunice Kennedy Shriver National Institute of Child Health and Human Development.

Correspondence concerning this article should be addressed to F. Sayako Earle, Department of Speech, Language, and Hearing Sciences, University of Connecticut, 850 Bolton Road, Unit 1085, Storrs, CT 06269. E-mail: Frances.Earle@uconn.edu

sleep exerted either a protective or restorative effect on posttraining performance. In a further investigation, the variability of tokens during training determined whether sleep would promote improved performance on the trained tokens (limited token set) or facilitate generalization (expanded set of training tokens; Fenn et al., 2013).

Thus, it appears that sleep does not ubiquitously improve performance on trained perceptual tasks when it comes to speech. Rather, sleep effects appear to be more pronounced when the task requires a reorganization of the preexisting phonological system (e.g., Fenn et al., 2003). Previous studies have tended to uncover patterns that suggest maintenance of perceptual task performance (Eisner & McQueen, 2006; Fenn et al., 2003; Roth et al., 2005), rather than overnight improvement with one exception to our knowledge (Fenn et al., 2013). It is important to note that these studies have all addressed how sleep affects perceptual adjustments made within one's native language; thus, it is not yet clear how sleep might assist the acquisition of novel (non-native) acoustic-phonetic features.

For the formation of non-native speech sound categories, two bodies of work, word learning, and auditory skill learning literatures, suggest that sleep plays a crucial role in at least two qualitatively different ways (see Earle & Myers, 2014, for review). The collective literature on word learning show that sleep facilitates the integration of learned verbal or orthographic forms into the existing lexicon (Bowers, Davis, & Hanley, 2005; Clay, Bowers, Davis, & Hanley, 2007; Davis et al., 2009; Dumay & Gaskell, 2007; Dumay, Gaskell, & Feng, 2004). Moreover, sleep appears to facilitate generalization to untrained items, particularly in online tasks (Tamminen, Davis, Merkx, & Rastle, 2012). Insights from this word learning literature lead to the prediction that phonetic information may undergo a similar sleep-induced change in status within the mental phonology, resulting in generalization away from the trained instances in order to recognize the contrast spoken by new talkers or in new vowel contexts. In contrast, the literature on auditory (nonspeech) skill learning suggests that sleep enhances performance on tasks that assess learned skills (e.g., Atienza, Cantero, & Stickgold, 2004; Brawn, Nusbaum, & Margoliash, 2010). Therefore, sleep may also promote improved performance on perceptual tasks in which the assessment tokens are identical to those used in training.

The first of these two predictions is supported by a recent study in our lab, in which generalization of training to an untrained talker occurred after sleep, but not before (Earle & Myers, 2015). Of note, this sleep effect on talker generalization was observed only in the identification task, whereas performance on discrimination of the non-native contrast, across trained and untrained conditions, remained stable over time. Furthermore, there was no significant improvement in identification on the trained talker, suggesting that sleep effects on performance in the identification task applied only to the generalization of training across talkers, and did not facilitate improved performance with the trained tokens.

A lack of sleep-related improvement in discrimination contradicted the prediction generated by the auditory skill learning literature. This discrepancy between our expectation and our findings motivated a more careful consideration of the demands of the phonetic identification and discrimination tasks, and in particular a consideration of how these demands recruit declarative and procedural memory systems, which are themselves differently affected by sleep (see Marshall & Born, 2007, for review).

Tasks Used to Assess Speech Perception: Differential Effects of Sleep

An individual's performance on different perceptual tasks, such as identification and discrimination of non-native speech tokens, is often assumed to reflect the quality of common perceptual representations of the target contrast. However, within-individual performance on different perceptual tasks are often found to diverge (e.g., Earle & Myers, 2015; MacKain, Best, & Strange, 1981); furthermore, it has been proposed that different sources of information contribute to task performance (e.g., Antoniou, Best, & Tyler, 2013; Antoniou, Tyler, & Best, 2012). For example, Antoniou et al. (2012) assessed a group of Greek–English bilinguals on category goodness ratings and discrimination along a voice onset time (VOT; /p/–/b/ and /d/–/t/) continuum of word-initial stops. The authors found that, while category goodness ratings given by the bilinguals were consistent with English and Greek monolinguals respective to the language mode of the target tokens, discrimination judgments aligned with the VOT boundaries common in the dominant language of the bilinguals' linguistic environment. Similarly, Antoniou et al. (2013) found that Greek–English bilinguals' categorization judgments on a non-native (Ma'di) contrast differed according to the language in which the instructions were given, but that language mode did not affect discrimination performance across subgroups. This set of studies suggests that performance on category goodness ratings and categorization tasks are more sensitive to language-specific phonetic knowledge than discrimination performance. We have argued similarly for the task-specific recruitment of different perceptual information following categorization training (Earle & Myers, 2014). Specifically, for the sake of generating predictions utilizing the wider memory consolidation literature, we have discussed this separation of task performance in terms of declarative and procedural knowledge.

Identification tasks, in which listeners map the acoustic input onto a visual or motoric label (such as choose *A* vs. *B*, or click *left* or *right*), require the explicit recall of cross-modal information. Therefore, changes to task performance across time may reflect the different stages of memory encoding in the declarative memory system (see Earle & Myers, 2014, for review). The benefit of sleep to declarative knowledge is associated with the hippocampal-cortical transfer of information thought to occur during slow-wave sleep (see Diekelmann & Born, 2010; for review; Wilson & McNaughton, 1994; Ji & Wilson, 2007), often referred to as "systems consolidation." Systems consolidation (*complementary systems account of learning*; McClelland, McNaughton, & O'Reilly, 1995) predicts the offline abstraction and integration of the episodic trace with preexisting information. This leads to the prediction that the effects of sleep-mediated abstraction of acoustic phonetic features from the training tokens will be more salient for tasks that directly assess declarative recall of token-label mapping. This is consistent with the sleep-mediated talker generalization effect that we observed in our previous work (Earle & Myers, 2015).

In contrast, perceptual discrimination may not require the explicit recall of category label, but is often observed to improve as

a result of categorization training (McCandliss et al., 2002; Swan & Myers, 2013). We have therefore argued (Earle & Myers, 2014) that improvement on discrimination requires an implicitly acquired ability to attend selectively to the relevant acoustic–phonetic details of the signal (see Francis & Nusbaum, 2002, for an attention-based model on non-native speech learning); in other words, training-induced changes to performance in this case may reflect procedural learning. For procedural learning, sleep effects have been more consistently observed in the improvement of an acquired skill as opposed to the generalization or abstraction of skill to new input. It has been suggested that the mechanism underlying such skill enhancement in perceptual tasks is the localized strengthening in the primary sensory cortex of selective synapses engaged during perceptual learning (Schwartz, Maquet, & Frith, 2002), which may behaviorally manifest as an increased automaticity (as might be measured by decreased reaction time [RT] or increased accuracy) in perceptual tuning (e.g., Atienza, Cantero, & Stickgold, 2004). This process is thought to reflect latent synaptic consolidation during REM sleep, occurring as a complementary, but distinct, process to systems consolidation (see Diekelmann & Born, 2010, for review).

To reiterate, the perceptual skill that is proposed to be acquired implicitly through categorization training is the ability to attend selectively to features that disambiguate the target tokens. Whereas the effects of systems consolidation (that leads to the generalization of skill to new instances) may be limited to tasks that assess declarative recall (such as identification), synaptic consolidation might be expected to facilitate improved performance whenever the task uses familiar (trained) tokens. Therefore, sleep is predicted to facilitate improvement in discrimination, as well as identification, of the trained tokens. However, prior work failed to show any improvements in discrimination or identification on the trained tokens (Earle & Myers, 2015). It is important to note that because that study was designed specifically to assess generalization to new instances, the stimulus test set included a large degree of variability. The token set used during assessment included trained and untrained vowels, and trained and untrained speakers; perhaps, as result, the task undermined participants' ability to retain consistent acoustic-phonetic features particular to the training tokens.

In the current investigation, two questions are examined. First, we ask whether, with reduced variability in the training and test set, sleep will facilitate improvements in both discrimination and identification of trained tokens following training (Experiment 1). The current study therefore differs from the previous in two ways: the variability of the assessment tokens was reduced, and the number of assessment trials was increased. As a result, sleep is predicted to facilitate improvement on identification and discrimination of the trained tokens, but not in discrimination of the untrained tokens. Second, we ask whether exposure to similar native-like tokens may interfere with sleep-mediated improvements in consolidation (Experiments 2 and 3).

Experiment 1

Changes in discrimination performance were tracked after identification training over 24 hr after training on a non-native (Hindi dental vs. retroflex stop) contrast. Participants were trained in the morning or the evening, and maintenance was assessed at approximate 12-hr intervals. On the basis of an analogy with the auditory skill learning literature, improvement in discrimination and identification was expected during the overnight interval.

Method

Participants. Sixty-nine undergraduate students (48 female, 21 male) between the ages of 18 and 24 were recruited from the University of Connecticut community, and were given course credit in exchange for their participation. This experiment was advertised to monolingual speakers of American English only; upon enrollment, nine participants were excluded on the basis of reporting that they were bilingual, or had grown up in a multilingual household. Six participants did not finish the study. Data from the remaining 54 participants (40 female, 14 male) were processed for further analyses. Participants gave informed consent in accordance with the guidelines of the University of Connecticut institutional review board.

Stimuli. Five exemplars of each 'word' (minimal pairs /ɖ̪ug/ and /d̪ug/; /ɖ̪ig/ and /d̪ig/) were produced by an adult male native speaker of Hindi. Auditory stimuli were recorded using a digital recorder (Roland Corporation, Los Angeles, CA) in a sound-proof booth. Tokens were trimmed to the onset of the stop burst, and mean amplitude was normalized across stimuli using Praat (Boersma & Weenink, 2013). The same set of 20 tokens was used for all participants in the discrimination task. Participants were trained/assessed on a subset of 10 tokens (either /ɖ̪ug/ and /d̪ug/ OR /ɖ̪ig/ and /d̪ig/) for the identification task.

For the identification task, we employed two novel visual objects ("fribbles"; Stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University, http://www.tarrlab.org/), one for each word within the minimal pair on which participants were trained. Stimuli were presented such that each place of articulation (dental or retroflex) corresponded to a different fribble. The pairing between the minimal pair words and the two fribbles was counterbalanced across participants.

E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) was used for stimulus presentation and recording participant response. Participants heard auditory stimuli through SONY MDR-7506 Hi-Fi digital Sound Monitor headphones, at an average listening level of 75 dB SPL (range: 44 – 80dB SPL).

Task schedule. Participants were randomly assigned to morning and evening groups, with morning participants receiving training between 8 a.m. and 10 a.m. and Evening participants receiving training between 6 p.m. and 9 p.m. Participants returned to the lab approximately 12 and 24 hr after training to assess maintenance of learning (See Figure 1). Identification (ID) of the learned contrast was assessed after training and at the two follow-up sessions. Discrimination ability (AX task) was measured at four time points: immediately before training (baseline), immediately following training during session one (Posttest 1), and again in sessions two and three (Posttests 2 and 3).

Identification training and test. Participants were trained to perceive the contrast in one vowel context: half of the sample were trained to identify /ɖ̪ug/ with /d̪ug/, and the other half were trained on /ɖ̪ig/ with /d̪ig/. During an initial familiarization sequence, each fribble was presented in the center of the screen while the participant heard "this is a . . ." with the corresponding token repeated five times. The training itself consisted of 200 trials of a self-

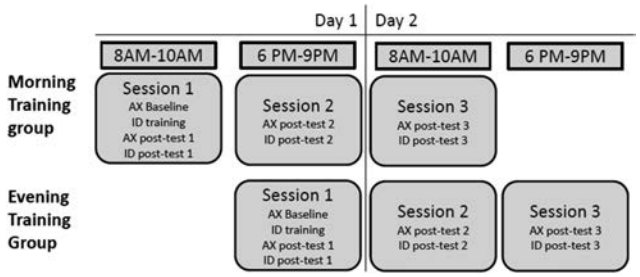

Figure 1. Overview of timing in the experimental protocol for Experiment 1.

paced, forced-choice identification task with a 3-min break after the first 100 trials. Within each trial, two fribbles remained visible on the screen while the participant heard "this is a . . ." followed by a dental or retroflex token. Participants indicated their choice with a mouse click, and written feedback (*correct* or *incorrect*) was given immediately following the response for every trial. During each identification posttest, 40 trials of identification without feedback were administered.

Discrimination test.

The discrimination task followed an AX design, with an interstimulus interval of one second between tokens. At each of the four time points, participants completed a total of 128 trials, such that half of the word pairs contained /u/ and half contained /i/. Note that for every participant, one vowel was the trained vowel, and the other was untrained, with the trained vowel counterbalanced across participants. Within each vowel set, 32 of the trials contained a pair of the "same" words and 32 contained "different" words. Same trials used two acoustically distinct exemplars of /d̪_g/ and /d̪_g/; /d̪_g/ and /d̪_g/ such that the measure tapped an individual's recognition of the speech sound *category* rather than allowing participants to use low-level acoustic information (e.g., pitch) to discriminate tokens, and every 'same' trial was acoustically unique. Similarly, each 'different' trial contained either a unique pairing or a unique ordering of the dental and retroflex exemplars, such that no two different trials were identical. Participants were instructed to decide if the sound at the beginning of each word was the same type of speech sound, or belonged to different types of speech sounds. Participants completed eight practice trials with feedback prior to each assessment.

To ensure that only participants who were actively engaged in the task for the duration of the session were included, participants whose scores on either the identification or discrimination posttest were at or below chance (a d' value of 0) were excluded. Data from three participants were excluded on this criterion. Data from the remaining 51 participants (n = 26/morning; n = 25/evening) are included in the following analyses.

Results

Preliminary analyses and data preparation. Percent accuracy in identification and discrimination were converted to d' scores (MacMillan & Creelman, 2004). See Table 1 for mean percent accuracy and response bias. In order to rule out any pretraining differences in discrimination ability, we ran a 2 × 2 mixed models analysis of variance (ANOVA) with vowel context (trained or untrained) as the within-subjects measure and group as the fixed factor on the baseline discrimination scores. There were no main effects of group or vowel context, $F(1, 49) = .32$, $p = .425$, $\eta^2 = .013$; $F(1, 49) = .26$, $p = .610$, $\eta^2 = .005$, respectively, and no interaction between group and vowel context, $F(1, 49) = .19$, $p = .665$, $\eta^2 < .004$. This suggests that discrimination ability across vowel context and group were comparable prior to training.

A baseline measure of identification performance was not obtained, because the decision over arbitrary token-label pairings would have been random prior to receiving instruction in the token-label assignments. Therefore, to ensure that participants performed above chance following training, we performed a one-sample t test on the identification posttest immediately after training (ID Posttest 1). ID Posttest 1 scores differed significantly from 0 ($t_{50} = 7.13$, $p < .000$, 95% CI: [1.65; 2.94]). Furthermore, in

Table 1
Mean Accuracy and Response Bias by Vowel Context by Group for Experiment 1

	Discrimination performance				Identification performance
	Trained vowel context		Untrained vowel context		Trained vowel context
Time	Accuracy (% Correct)	Response bias (% False alarm)	Accuracy (% Correct)	Response bias (% False alarm)	Accuracy (% Correct)
			Morning training group		
Baseline	.64 (.10)	.47 (.19)	.65 (.11)	.43 (.21)	
Posttest 1	.70 (.10)	.35 (.17)	.65 (.11)	.44 (.23)	.73 (.18)
Posttest 2	.70 (.13)	.33 (.17)	.67 (.10)	.39 (.19)	.74 (.20)
Posttest 3	.67 (.12)	.40 (.18)	.66 (.11)	.43 (.20)	.76 (.22)
			Evening training group		
Baseline	.63 (.11)	.40 (.16)	.64 (.08)	.44 (.17)	
Posttest 1	.68 (.11)	.38 (.20)	.66 (.11)	.44 (.22)	.75 (.16)
Posttest 2	.71 (.13)	.39 (.24)	.67 (.10)	.48 (.22)	.80 (.17)
Posttest 3	.72 (.12)	.39 (.24)	.66 (.11)	.47 (.20)	.80 (.17)

Note. "% False alarm" is the percentage of trials incorrectly identified as "different" when the tokens belong to the same category. Standard deviations of the mean are indicated in parentheses.

order to ensure that both groups achieved comparable levels of performance on the identification task, an independent samples *t* test by group on the ID Posttest 1 scores was performed. Differences in group performance immediately following training were not statistically significant ($t_{49} = -.31$, $p = .757$, 95% CI: [−1.54; 1.13]; confidence intervals were adjusted for family wise error rate [FWER] using Holms-Bonferroni correction at $p < .05$). This suggests that both groups improved on the identification task as a result of training, and that the degree of improvement was comparable across groups. Learning rate, as measured by average accuracy per 50 trials during the training phase, is depicted in Figure 2a.

Identification. To determine if there were any changes in identification performance over the 24-hr experiment period in the absence of further training, we ran a 2 × 3 mixed-model repeated measures ANOVA with group (morning or evening training) as the between-subjects factor, and two levels of time (ID Posttest 1, ID Posttest 2, ID Posttest 3) as the within-subjects factor was performed. Identification performance remained relatively stable over the 24-hr period for both groups (no main effects or interactions: Group = $F[1, 49] = .06$, $p = .801$, $\eta^2 = .001$; Time = $F[2, 98] = 1.95$, $p = .148$, $\eta^2 = .038$; Time × Group: $F[2, 98] = .51$, $p = .605$, $\eta^2 = .010$; see Figure 3).

Training-related changes in discrimination performance. Discrimination performance improved in both groups following training, even though this task was not explicitly trained. Comparable gains across groups were confirmed via a 2 × 2 × 2 mixed models ANOVA on just the time points immediately before and after training, with group as the between-subjects factor (morning or evening training), and two levels of time (baseline and Posttest 1) and vowel context (trained/untrained vowel: whether the vowel context was explicitly trained [trained] or not [untrained]) as within-subjects factors; see Figure 4). Participants in both groups improved from pretest to posttest, primarily on the trained vowel context (significant main effect of Time = $F[1, 49] = 15.50$, $p < .001$, $\eta^2 = .24$; interaction between time and vowel context: $F[1, 49] = 1.30$, $p = .010$, $\eta^2 = .13$). No other main effects or interactions emerged (main effect of Group = $F[1, 49] = .59$, $p = .585$, $\eta^2 = .01$; Vowel Context × Group: $F[1, 49] = .44$, $p = .508$, $\eta^2 = .01$; Time × Group: $F[1, 49] < .00$, $p = .996$, $\eta^2 < .00$; Time × Vowel Context × Group: $F[1, 49] = 2.63$, $p = .111$, $\eta^2 = .05$). The factors driving the Time × Vowel Context interaction were explored by performing two paired samples *t* tests comparing baseline and Posttest 1 scores for each vowel context, collapsed across groups. For the trained vowel context, Posttest 1 score was significantly higher than at baseline ($t_{50} = -5.68$, $p < .001$, 95% CI: [−0.58; −0.28]), whereas for the untrained vowel context, the difference was not statistically significant ($t_{50} = -0.11$, $p = .301$, 95% CI: [−0.32; 0.10]). Taken together, this suggests that both groups improved in discrimination performance in the trained vowel context, but not the untrained vowel context, through identification training. The magnitude of gain furthermore appears to be comparable between groups.

Sleep-mediated changes in discrimination maintenance. The influence of time of training relative to sleep on changes in discrimination ability over 24 hr was investigated using a 2 × 3 × 2 mixed-models ANOVA with group as the single between-subjects factor, and three levels of time (Posttest1, Posttest2, and Posttest3) and vowel context as within-subjects factors. There was

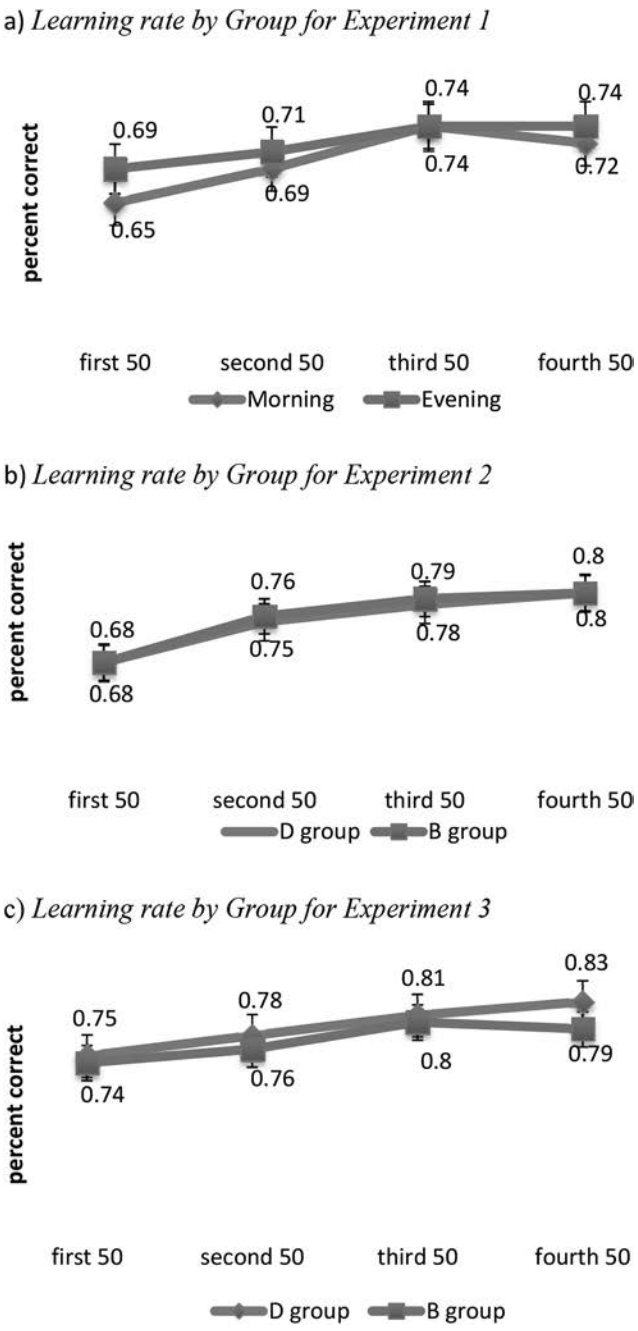

Figure 2. Learning rate by Group by Experiment. Group average response correct is plotted per 50 trials of identification training (trials with feedback). Error bars denote standard errors of the mean. See the online article for the color version of this figure.

a significant main effect of vowel context, $F(1, 98) = 5.79$, $p = .017$, $\eta^2 < .11$, and a significant three-way interaction between group, time, and vowel context, $F(2, 98) = 3.77$, $p = .027$, $\eta^2 < .07$.

Visual inspection of the means suggested that this interaction likely resulted from significant differences between groups and time in the trained contrast but not in the utrained contrast. This was confirmed

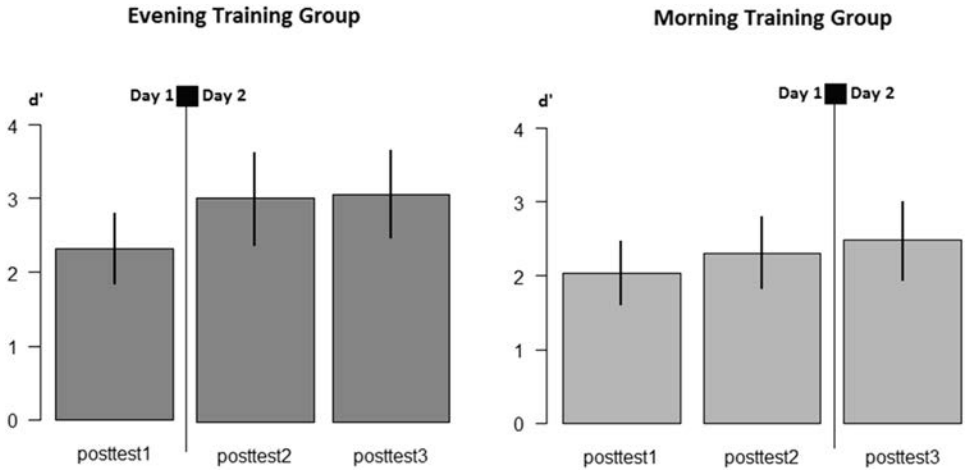

Figure 3. Profile of changes in identification performance by training group for Experiment 1. Error bars indicate standard error of the mean. See the online article for the color version of this figure.

by performing two repeated measures ANOVAs on each vowel context (trained/untrained) separately. In the trained vowel condition, we observed a significant interaction between time and group, $F(2, 98) = 4.52$, $p = .013$, $\eta^2 = .09$, but neither group nor time main effects, $F(1, 49) = .03$, $p = .857$, $\eta^2 < .01$; $F(2, 98) = .48$, $p = .618$, $\eta^2 = .01$, respectively. In the untrained vowel condition, we observed no significant effects or interactions (Time: $F[2, 98] = .37$, $p = .69$, $\eta^2 = .01$; Group: $F[1, 49] = .02$, $p = .886$, $\eta^2 < .01$; Time × Group $F[1, 49] = .54$, $p = .466$, $\eta^2 = .011$).

Visual inspection of the pattern within the trained contrasts suggests that groups differ in the direction of change over time, with the morning group losing sensitivity and the evening group gaining sensitivity. This was confirmed by a 2 × 2 mixed-models ANOVA with group as the fixed factor and two levels of time (Posttest 1 and Posttest 3) as the within-subjects factor. A significant interaction between time and group, $F(1, 49) = 11.66$, $p = .001$, $\eta^2 = .09$, but no main effects of time or group, $F(1, 49) = .29$, $p = .590$, $\eta^2 = .01$; $F(1, 49) = .19$, $p = .668$, $\eta^2 < .01$;

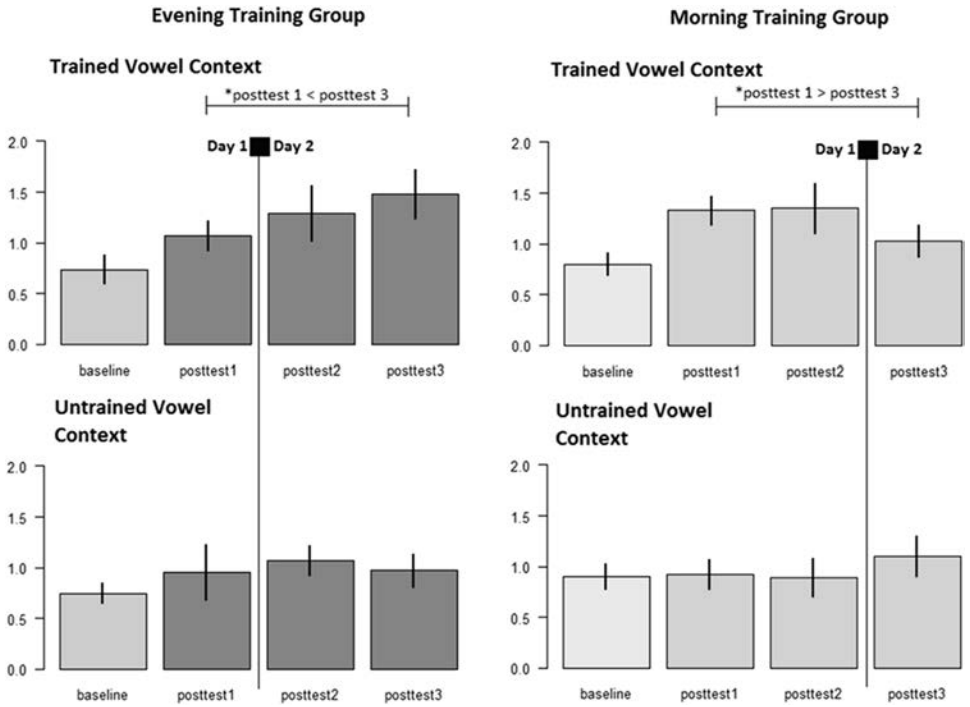

Figure 4. Profile of changes in discrimination performance by training group and by vowel context (trained or untrained) for Experiment 1. Error bars indicate standard error of the mean. * Indicates statistical significance at alpha = .05. See the online article for the color version of this figure.

respectively, were found. Paired samples *t* tests between Posttests 1 and 3 separately by group indicated that the morning group exhibited significantly lower scores at Posttest 3 than immediately after training ($t_{25} = 2.61$, $p = .015$, 95% CI: [0.06; 0.54]). For the evening group, Posttest 3 scores were significantly higher than immediately after training ($t_{24} = -2.34$, $p = .028$, 95% CI: [−0.78; −0.05]).

Discussion

Results of Experiment 1 support the view that sleep plays a role in enhancing discrimination of a trained non-native contrast. In contrast, the identification data shows a gradual (non-significant) increase in mean performance per session. We reserve the discussion on the pattern of changes to identification performance over 24 hr until the discussion section of Experiment 2.

Of interest, only individuals trained in the evening demonstrated significant improvement in discrimination following the overnight interval. No improvement in performance on an untrained vowel context was seen. The finding that the morning group shows no sleep-mediated improvement suggests that the effects of sleep may depend in part on the duration or quality of posttraining wake state activity before sleep. While the specific effects on performance were different in their case, Fenn et al. (2003) similarly described different consequences of the overnight interval on performance for participants trained in the morning versus evening. This postsleep discrepancy in performance between groups may reflect differences in the quality of non-native phonetic representations that emerged overnight, though why this might be is unclear. One possibility is that differences in circadian rhythms contribute to diurnal differences in the learning that is taking place in the morning versus evening. A second possibility points to the amount of incidental exposure to native language sounds before sleep. That is, the morning group is likely to be exposed to more English between training and sleep than the evening group.

Several accounts of non-native speech sound learning in adulthood suggest that the presence of similar sounds in one's native language interferes with the learning of the non-native sounds (Best, 1995; Flege, 1995). However, these accounts focus on the difficulty in distinguishing the non-native tokens from the existing representation of native speech sounds. Results of Experiment 1 raise the possibility that this difficulty may be compounded by active interference from exposure to native language tokens subsequent to training. Given that the dental and retroflex sounds perceptually resemble the English /d/ sound, exposure to alveolar /d/ may prevent the perceptual enhancement of the learned contrast overnight.

Similar interference effects have been previously reported in the procedural learning literature. For example, Walker, Brakefield, Hobson, and Stickgold (2003) trained three groups of participants on a motor (finger tapping) sequence. The first group only learned one sequence and was retested after 24 hr. The second group learned a second sequence immediately after the first, and was also retested after 24 hr. The third group learned a second sequence immediately after the first, and was retested immediately after training. Although the first group showed an increase in speed and accuracy on the target (first) sequence, the second group only showed a performance increase on the second sequence. Performance immediately after training in the third group however indicated that the two sequences were comparably learned. Taken together, the authors interpreted that while the learning of the second sequence does not impede the learning of the first sequence initially, the learning of the second sequence interfered with the latent consolidation of the first. Similarly, our Morning group shows stable performance at the Session 2 posttest, suggesting that the decline in discrimination performance does not occur until sleep during the overnight interval. Experiment 2 tests this interpretation directly.

Experiment 2

To isolate the effect of native language interference and to control for the potential confound introduced from diurnal effects, we trained all participants in the evening. The only substantial difference from Experiment 1 was that, immediately following training and Posttest 1, participants were randomly assigned to one of two interference conditions and exposed to a train of native-language syllables beginning with /d/ (D group) or /b/ (B group). We predicted that passive exposure to /d/s immediately following training would prevent sleep-mediated improvement on discrimination of the dental-retroflex contrast, whereas exposure to /b/s would not.

Method

Participants. Sixty-eight (10 male, 58 female) participants were recruited from the University of Connecticut community and were given course credit for participation. All participants gave informed consent in accordance with the University of Connecticut institutional review board guidelines. This experiment was advertised to monolingual speakers of American English only; upon enrollment, data from 11 participants were excluded on the basis of participants' reports that they were bilingual, or had grown up in a multilingual household. Two participants who were enrolled and met our criteria did not finish the study. For Experiment 2, we introduced an exit survey that asked participants to report on the approximate number of hours that they slept during the overnight between-session interval during the 24-hr experiment period. Three students reported having slept less than four hours during the 24-hr experiment period, and were excluded from the analyses in case fatigue played a role in performance. Seven additional participants were excluded due to the same posttest performance criterion from Experiment 1. Forty-five (4 male, 41 female; $n =$ 22/B group; $n = 23$/D group) participants met all criteria and are included in the analyses.

Stimuli. The materials and methods for the training session and the two posttest sessions are identical to those used in Experiment 1, following the protocol schedule of the evening group (see Figure 5). In addition, digitally recorded, naturally spoken speech tokens produced by native speakers of English were used as "interference" tokens. Each condition (B or D) employed 300 acoustically unique tokens, consisting of 5 exemplars each of /dV/ or /bV/ tokens occurring in six vowel contexts (/æ/, /ɑ/, /ɛ/, /i/, /u/, /o^{w}/) produced by 10 native speakers of English (five female, five male). The tokens were presented in random order at an interstimulus interval of 300 ms through five cycles, such that the

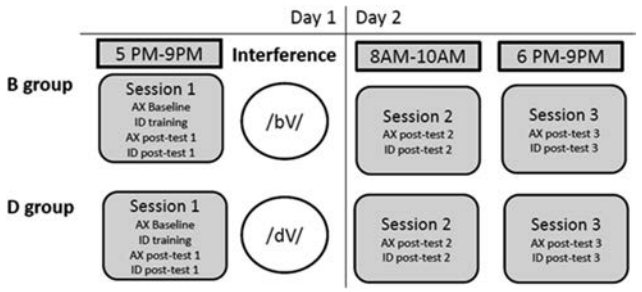

Figure 5. Overview of timing in the experimental protocol for Experiment 2.

stimulus train in each interference condition included 1,500 tokens lasting approximately 15 min. Immediately following the posttest in Session 1, participants were randomly assigned to the B or D group, and given a choice of either working on a Sudoku puzzle or drawing while they were passively exposed to the interference stimuli train respective to their group assignment through SONY MDR-7506 Hi-Fi digital Sound Monitor headphones.

Results

Preliminary analyses and data preparation. Percent accuracy in identification and discrimination were converted to d' scores (MacMillan & Creelman, 2004); mean percent accuracy and response bias are reported in Table 2. We first determined the comparability of groups and of the two vowel contexts by running a 2×2 repeated measures ANOVA on the baseline discrimination performance of the vowel context (trained or untrained) with group as the fixed factor. There were no main effects of group nor vowel context, $F(1, 43) = .19$, $p = .67$, $\eta^2 = .558$; $F(1, 43) = .28$, $p = .602$, $\eta^2 = .006$, respectively, and no interaction between group and vowel context, $F(1, 43) = .02$, $p = .897$, $\eta^2 < .001$. This suggests that discrimination ability across vowel context and group were comparable prior to training.

As in Experiment 1, a one-sample t test on ID posttest 1 d' scores across both groups (B and D) was performed to ensure that performance on the trained task was above chance following training. Session 1 identification scores differed significantly from 0 ($t_{44} = 35.90$, $p < .001$, 95% CI: [0.71; 0.80]). To ensure that both groups achieved comparable levels of performance on the identification task, we performed an independent samples t test by group on the ID Posttest 1 scores. We found that group performances did not differ significantly ($t_{43} = .66$, $p = .616$, 95% CI: [-.06; .11]. This suggests that groups improved on the identification task as a result of training and that the degree of improvement was comparable across groups. Learning rate during the training phase for each group is depicted in Figure 2b.

Identification. To determine if there were any changes to Identification performance over the 24-hr experiment period, we conducted a 2×3 mixed models ANOVA was conducted with group (B or D) as the fixed factor and three levels of time (ID Posttest 1, ID Posttest 2, ID Posttest 3) as the within-subjects factor (see Figure 6). There was no main effect of group, $F(1, 43) = .09$, $p = .761$, $\eta^2 = .002$, and no interaction between time and group, $F(2, 86) = .03$, $p = .966$, $\eta^2 = .001$. In contrast to Experiment 1, there was a main effect of time, $F(2, 86) = 6.26$, $p = .003$, $\eta^2 = .127$. We further determined which sessions were driving the time main effect by running three paired samples t tests (ID Posttest 1 – ID Posttest 2; ID Posttest 2 – ID Posttest 3; ID Posttest 1 – ID Posttest 3) collapsed across groups with the Holms-Bonferroni correction applied to calculate confidence intervals. Session 3 performance was significantly higher than Posttest 1 ($t_{44} = -3.00$, $p = .004$, 95% CI: [−.44; .012]), and there was a trend (after correction) toward a higher performance on Posttest 2 than on Posttest 1 ($t_{44} = -2.06$, $p = .046$, 95% CI: [−.01; 0]) and no significant difference between Posttests 2 and 3 ($t_{44} = -1.51$, $p = .138$, 95% CI: [−.06; .01]).

Comparison of identification data to Experiment 1. On the basis of within-experiment analyses of the identification data, it appears that Experiments 1 and 2 diverge in patterns of change over time. However, visual inspection of the identification patterns

Table 2
Mean Accuracy and Response Bias by Vowel Context by Group for Experiment 2

	Discrimination performance				Identification performance
	Trained vowel context		Untrained vowel context		Trained vowel context
Time	Accuracy (% Correct)	Response bias (% False alarm)	Accuracy (% Correct)	Response bias (% False alarm)	Accuracy (% Correct)
			D group		
Baseline	.64 (.08)	.42 (14)	.63 (.07)	.41 (14)	
Posttest 1	.69 (.08)	.38 (.17)	.65 (.08)	.40 (.19)	.76 (.20)
Posttest 2	.69 (.10)	.36 (.13)	.65 (.08)	.41 (.23)	.83 (.11)
Posttest 3	.70 (.13)	.35 (.19)	.67 (.09)	.38 (.18)	.91 (.19)
			B group		
Baseline	.60 (.09)	.47 (.15)	.56 (.14)	.46 (.18)	
Posttest 1	.65 (.13)	.33 (.16)	.62 (.11)	.42 (.17)	.80 (.08)
Posttest 2	.69 (.17)	.33 (.21)	.62 (.14)	.43 (.20)	.86 (.10)
Posttest 3	.69 (.18)	.35 (.23)	.65 (.11)	.39 (.19)	.83 (.14)

Note. "% False alarm" is the percentage of trials incorrectly identified as "different" when the tokens belong to the same category. Standard deviations of the mean are indicated in parentheses.

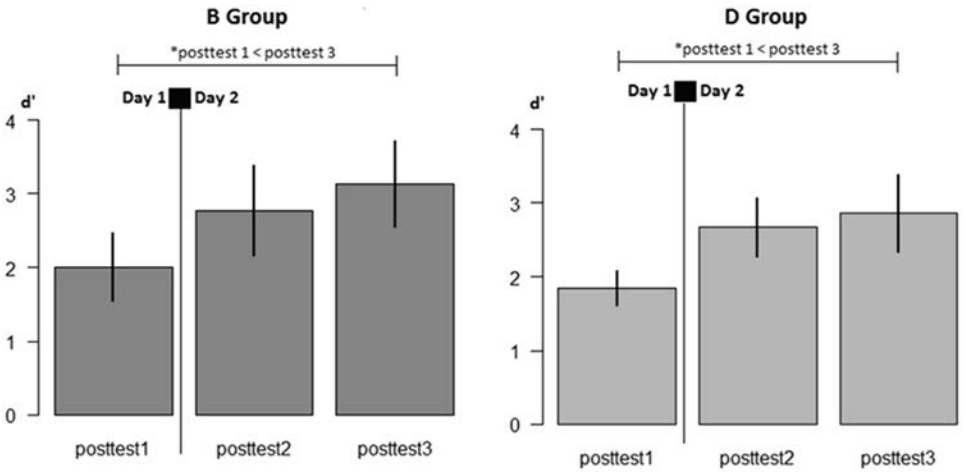

Figure 6. Profile of changes in identification performance by training group for Experiment 2. Error bars indicate standard error of the mean. * Indicates statistical significance at alpha = .05. See the online article for the color version of this figure.

for Experiments 1 and 2 suggests that the evening group's identification pattern of performance appear to be similar to the two groups in Experiment 2. Therefore, we ran an additional 4 × 3 mixed-model ANOVA with time (three levels) as the within-subjects factor and group (Morning, Evening, B, D) as the fixed factors. There was a significant main effect of time, $F(2, 184) = 8.439$, $p < .001$, $\eta^2 = .084$, but no main effect of group, $F(2, 92) = .421$, $p = .657$, $\eta^2 = .009$, nor an interaction between time and group, $F(4, 184) = .166$, $p = .955$, $\eta^2 = .004$. To further investigate the time main effect, we ran two paired samples *t* tests collapsed across groups comparing Posttests 1 and 2 and Posttests 2 and 3. Holms-Bonferroni correction was applied in calculating CIs. We found that Posttest 2 scores were significantly higher than Posttest 1 scores ($t_{95} = -3.04$, $p = .003$, 95% CI: [−1.11; −.15]), but that the difference between Sessions 2 and 3 were not statistically significant ($t_{95} = -1.04$, $p = .302$, 95% CI: [–.56; .17]). This suggests that the Time main effect is driven by the changes between Posttests 1 and 2. Furthermore, the lack of an effect or an interaction involving group suggests that the patterns observed in Experiment 1 are not dissimilar to Experiment 2. The lack of a main effect of time in Experiment 1 may therefore have been due to greater within-group variability in identification scores relative to Experiment 2 (see Tables 1 and 2 for standard deviations of percent accuracy in identification).

Training-related changes in discrimination performance. As in Experiment 1, we first determined that identification training resulted in a comparable gain in discrimination performance in both Groups by running an initial 2 × 2 × 2 mixed-model ANOVA with group (B or D), and 2 levels of time (baseline and Posttest 1) and vowel context as within-subjects factors (see Figure 7). There was a significant main effect of time, $F(1, 43) = 4.93$, $p < .034$, $\eta^2 = .10$, but no main effect of group, $F(1, 43) = .04$, $p = .837$, $\eta^2 < .01$, nor any interactions involving group (Vowel Context × Group: $F[1, 43] = .40$, $p = .533$, $\eta^2 = .01$; Time × Group: $F[1, 43] = 1.51$, $p = .226$, $\eta^2 = .034$; Time × Vowel Context × Group: $F[1, 43] = .34$, $p = .566$, $\eta^2 = .01$). Posttest 1 scores were significantly higher than baseline ($t_{44} = -2.18$, $p = .035$, 95% CI: [–.46; −.02]). Taken together, this suggests that both groups improved in discrimination performance through identification training, and that the magnitude of gain was comparable across groups.

Interference-related changes in discrimination performance. To determine if the type of interference condition affected changes to performance subsequent to training, a 2 × 3 × 2 mixed-model ANOVA was performed with group as the single between-subjects factor, and three levels of ime (Posttest1, Posttest2, and Posttest3) and vowel context as within-subjects factors. There was a significant main effect of vowel context, $F(1, 43) = 13.44$, $p = .001$, $\eta^2 < .24$, and significant interactions between time and group, $F(2, 86) = 3.14$, $p = .048$, $\eta^2 < .07$, time and vowel context, $F(2, 86) = 3.85$, $p = .025$, $\eta^2 < .08$, and a trend toward an interaction between vowel context and group, $F(2, 86) = 3.90$, $p = .055$, $\eta^2 < .08$. There were no other main effects or interactions (Time: $F[2, 86] = 1.11$, $p = .336$, $\eta^2 = .025$; Group: $F[2, 86] = .83$, $p = .367$, $\eta^2 = .019$; Group × Time × Vowel Context: $F[2, 86] = .45$, $p = .641$, $\eta^2 < .01$).

In order to determine the nature of the interactions between time and group, vowel context and time, and the trending interaction between vowel context and group, we ran two additional (2 × 3) mixed-model ANOVAs with group as the fixed factor and time as the within-subject factor. Even though we did not observe a three-way interaction, we chose to conduct these ANOVAs separately for each vowel context because of the two interactions involving vowel context and the vowel context main effect. In the trained vowel context, there was a significant main effect of time, $F(2, 42) = 3.23$, $p = .049$, $\eta^2 = .133$, and a significant interaction between time and group, $F(2, 42) = 4.51$, $p = .038$, $\eta^2 = .177$. In the untrained vowel context, there was no main effect of time, $F(2, 42) = 1.36$, $p = .269$, $\eta^2 = .061$, and no interaction between time and group, $F(2, 42) = 1.97$, $p = .153$, $\eta^2 = .086$.

Visual inspection of the Trained Vowel means suggested that the interaction between Group and Time was largely due to improvements over time in the B Group, while the D Group appeared to maintain performance over the 24-hr interval. This was con-

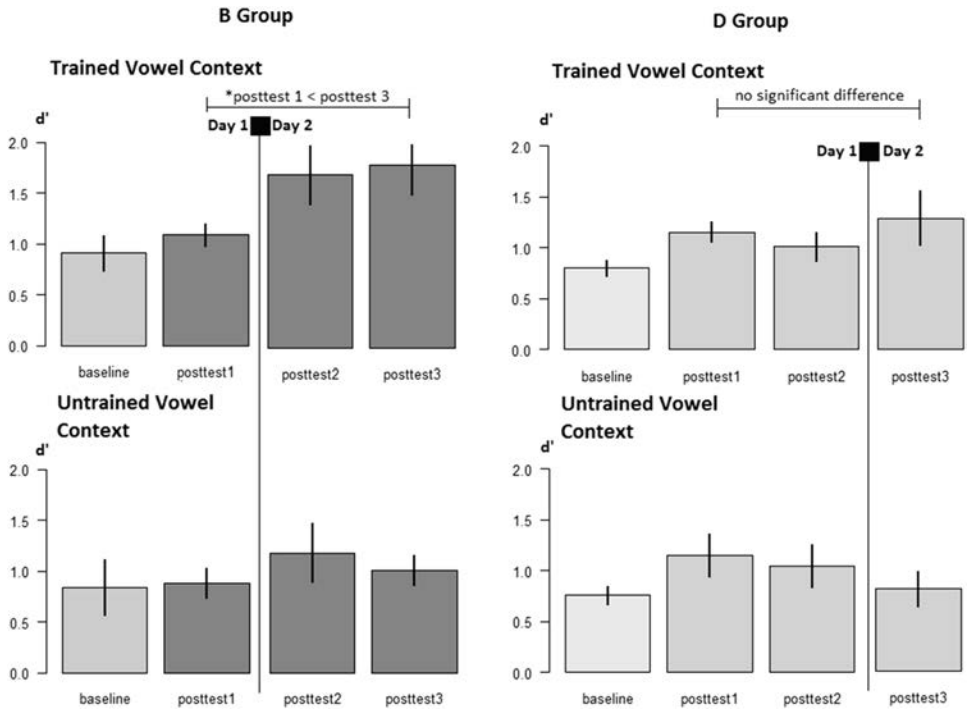

Figure 7. Profile of changes in discrimination performance by interference group and by vowel context (trained or untrained) for Experiment 2. Error bars indicate standard error of the mean. * Indicates statistical significance at alpha = .05. See the online article for the color version of this figure.

firmed by performing two separate repeated measures ANOVAs by group, with three levels of time as the single within-subjects factor. In the D group, there was no main effect of time, $F(2, 42) = .92$, $p = .405$, $\eta^2 = .042$, whereas in the B group, there was a significant main effect of time, $F(2, 44) = 3.50$, $p = .039$, $\eta^2 = .137$. The direction of the time main effect in the B group was explored by running three paired samples t tests (Posttest 1 – Posttest 2; Posttest 2 – Posttest 3; Posttest 1 – Posttest 3). Discrimination performance is significantly higher in Posttest 2 and 3 compared with Session 1 ($t_{22} = -2.52$, $p = .020$, 95% CI [−1.29; –.013]; $t_{22} = -2.76$, $p = .012$, 95% CI [−1.16; –.07], respectively). The difference between Posttest 2 and 3 is not statistically significant ($t_{22} = .64$, $p = .797$, 95% CI [–.83; .64]), suggesting that the Time main effect is driven by the gain in performance overnight (between Sessions 1 and 2) in the B group that is maintained until Session 3, roughly 24 hr following training.

In summary, the differences between D and B exposure groups emerge primarily because of differences in performance on discrimination of the trained contrast, with the B group showing improvements in performance following the overnight interval which are maintained at the 24-hr retest, a similar pattern to those observed in the Evening group in Experiment 1. In contrast, the D group shows no such changes in performance after training.

Discussion

Results of Experiment 2 suggest that posttraining linguistic exposure affects performance outcome on perceptual discrimination 24 hours following training. Specifically, those who are exposed to tokens that are dissimilar to the trained non-native contrast (/bV/) appear to improve in discrimination performance following sleep, in a pattern similar to the Evening group in Experiment 1 (see Figures 4 and 7). In contrast, those exposed to tokens that are similar to the trained non-native contrast (/dV/) do not improve performance following sleep in the time period subsequent to training. As previously mentioned, this interference effect resembles other work in the procedural learning literature that shows an attenuated retention of learning when individuals are exposed to conflicting information between learning and sleep (Walker, Brakefield, Hobson, & Stickgold, 2003; Goedert & Willingham, 2002). Furthermore, this interference effect appears not to affect identification performance, which lends support to our speculation that these two tasks are aided by information encoded by two distinct memory systems that are differentially susceptible to latent effects of interference.

At face value, these identification results appear in conflict with those in Experiment 1. However, a direct comparison between experiments in the patterns of changes over time suggests that these patterns of improvement are not significantly different. The lack of a significant effect of time in Experiment 1 therefore may have been due to greater within-group variability in performance in the morning and evening groups.

A question still remains as to precisely when the effect of this linguistic interference emerges in discrimination performance. In Experiment 1, we observed that the morning group's discrimination performance remained stable at Session 2, followed by a performance decline subsequent to a period of sleep. Thus, if the

amount of linguistic exposure (rather than the length of time between training and sleep) is to explain the overnight decline in the morning group, we must further establish that a decline in discrimination ability subsequent to interference is not observed prior to sleep. Crucially, we ask whether interference from listening to /d/ tokens has an immediate effect, or whether sleep is required in order for those tokens to interfere with the established memory trace.

Experiment 3

To determine if the effect of linguistic interference on the trained non-native contrast emerges immediately after exposure or only after sleep, we replicated Experiment 2 with two alterations in the experiment design. First, participants in Experiment 3 were exposed to the interference tokens between training and assessment. Second, though we were motivated in Experiment 2 to replicate the pattern over 24 hr in Experiment 1, our question in Experiment 3 concerns the time frame bound by the posttraining interference and the postsleep assessment. Thus, unlike the first two experiments, Experiment 3 was conducted in two sessions: a p.m. Training + Interference Session, and one a.m. reassessment session (see Figure 8 for schedule of protocol in Experiment 3).

Method

Participants. Thirty-nine (18 male, 21 female) participants were recruited from the University of Connecticut community, and were given course credit for participation. All participants gave informed consent in accordance with the University of Connecticut institutional review board guidelines. This experiment was advertised to monolingual speakers of American English, with a history of typical language and reading development only. Upon enrollment, data from two participants were excluded on the basis that their self-report indicated that they are bilingual. Data from three additional participants were excluded due to noncompliance with the experimental task. One participant who was enrolled and met our criteria did not finish the study, and data from one participant was lost due to equipment malfunction. Thirty-two (16 male, 16 female; n = 16/B group, n = 16/D group) participants met all criteria and finished the study; the data from these 32 are included in our following analyses.

Stimuli. The materials and methods for the training, interference, and the reassessments are identical to those used in Experiment 1 and 2, following the protocol schedule outlined in Figure 8. To reiterate, the critical difference concerned the timing of the interference block, which preceded the posttest assessment on Day 1.

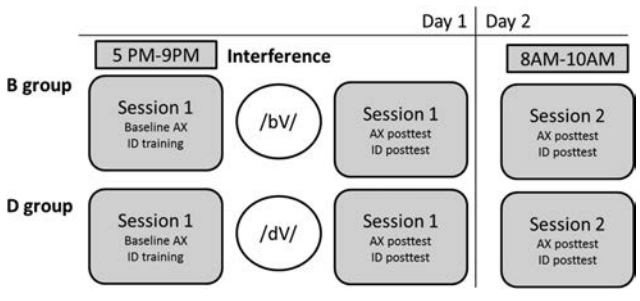

Figure 8. Overview of timing in the experimental protocol for Experiment 3.

Results

Preliminary analyses and data preparation. Percent accuracy in identification and discrimination were converted to d' scores (MacMillan & Creelman, 2004); mean percent accuracy and response bias are reported in Table 3. To determine the comparability of groups and of the two vowel contexts, we ran an initial 2 × 2 repeated measures ANOVA on the baseline discrimination performance of the vowel context (trained or untrained) with group as the fixed factor. There were no main effects of group nor vowel context, $F(1, 32) = .920$, $p = .345$, $\eta^2 = .028$; $F(1, 32) = .987$, $p = .328$, $\eta^2 = .030$, respectively, and no interaction between group and vowel context, $F(1, 32) = .001$, $p = .982$, $\eta^2 < .001$. Therefore, differences in discrimination ability across vowel context and group were not significant prior to training.

As in Experiments 1 and 2, a one-sample t test on ID Posttest 1 d' scores across both groups (B and D) was performed to ensure that participants were performing above chance following training. Session 1 identification scores differed significantly from 0 ($t_{33} = 8.408$, $p < .001$, 95% CI: [2.21; 3.65]). To ensure that both groups achieved comparable levels of performance on the identification task immediately following training, we performed an independent samples t test by group on the ID Posttest 1 scores. We found that group performances did not differ significantly ($t_{32} = 1.119$, $p = .272$, 95% CI: [−2.18; 0.63]). This suggests that groups improved on the identification task as a result of training, and that the degree of improvement was comparable across groups. Learning rate by group during the training phase is depicted in Figure 2c.

Identification. To determine if there were any changes to ID performance over the 12-hr experiment period, a 2 × 2 mixed-models ANOVA was conducted with Group (B or D) as the fixed factor, and two levels of Time (ID Posttest 1, ID Posttest 2) as the within-subjects factor (see Figure 9). There was no main effect of group, $F(1, 32) = 3.105$, $p = .088$, $\eta^2 = .088$, but we did observe an interaction between time and group, $F(1, 32) = 4.244$, $p = .048$, $\eta^2 = .117$, and a trend toward a main effect of time, $F(1, 32) = 4.040$, $p = .053$, $\eta^2 = .112$. We further determined the source of the interaction by conducting two paired samples t tests on performance at each session for each group separately, using Holms-Bonferroni correction for the calculation of confidence intervals. We found that for the D group, the difference in performance across Sessions 1 and 2 was not significant ($t_{15} = .040$, $p = .968$, 95% CI: [.60; .62]. For the B group, Session 2 performance was significantly higher than Posttest 1 ($t_{15} = -2.598$, $p = .019$, 95% CI: [−1.82; −.05]).

Changes in discrimination performance over time. As the interference block occurred between training and posttest for Experiment 3, we could not be certain as to when we should expect behavior to diverge between groups (immediately after training or not until after sleep). Therefore, we ran an initial omnibus 2 × 3 × 2 mixed models ANOVA on the discrimination d' scores with group (B or D) as the fixed factor and time (three levels) and vowel context (trained or untrained) as the within-subjects factors. There was a significant main effect of time, $F(2, 64) = 9.387$, $p < .001$, $\eta^2 = .227$a significant interaction between time and group, $F(2,$

Table 3
Mean Accuracy and Response Bias by Vowel Context by Group for Experiment 3

	Discrimination performance				Discrimination performance
	Trained vowel context		Trained vowel context		Trained vowel context
Time	Accuracy (% Correct)	Response Bias (% False alarm)	Accuracy (% Correct)	Response Bias (% False alarm)	Accuracy (% Correct)
			D group		
Baseline	.63 (.10)	.23 (16)	.61 (.10)	.26 (16)	
Posttest 1	.73 (.08)	.25 (.10)	.64 (.12)	.22 (.16)	.80 (.12)
Posttest 2	.70 (.08)	.26 (.10)	.61 (.10)	.19 (.12)	.88 (.10)
			B group		
Baseline	.60 (.13)	.40 (.20)	.62 (.12)	.36 (.24)	
Posttest 1	.69 (.31)	.33 (.20)	.65 (.14)	.22 (.28)	.85 (.23)
Posttest 2	.70 (.17)	.20 (.23)	.65 (.14)	.28 (.24)	.88 (.17)

Note. "% False alarm" is the percentage of trials incorrectly identified as "different" when the tokens belong to the same category. Standard deviations of the mean are indicated in parentheses.

64) = 6.664, $p = .002$, $\eta^2 = .172$, and an interaction between time and vowel context, $F(2, 64) = 3.547$, $p = .035$, $\eta^2 = .100$. There were no other main effects or interactions (vowel: $F[1, 32] = 2.020$, $p = .165$, $\eta^2 = .059$; group: $F[1, 30] = .337$, $p = .566$, $\eta^2 = .010$; Vowel × Group: $F[1, 32] = .348$, $p = .559$, $\eta^2 = .011$; Group × Time × Vowel Context: $F[2, 64] = .988$, $p = .378$, $\eta^2 = .030$).

Because of the interaction between time and vowel context, we conducted two additional 2 × 3 mixed-models ANOVAs for each vowel context. For the trained vowel, there was a significant main effect of time, $F(2, 64) = 15.554$, $p < .001$, $\eta^2 = .327$, and a significant interaction between time and group, $F(2, 64) = 8.202$, $p = .001$, $\eta^2 = .204$. There was no group main effect, $F(1, 32) = .808$, $p = .375$, $\eta^2 = .025$. For the untrained vowel, there were no significant main effects nor interactions (time: $F[2, 64] = .842$, $p = .436$, $\eta^2 = .023$; group: $F[1, 32] = .034$, $p = .855$, $\eta^2 = .001$; Time × Group: $F[2, 64] = 1.289$, $p = .283$, $\eta^2 = .039$). Therefore, the Time × Group interaction in the omnibus ANOVA appears to be driven by the Time × Group interaction in the trained vowel context.

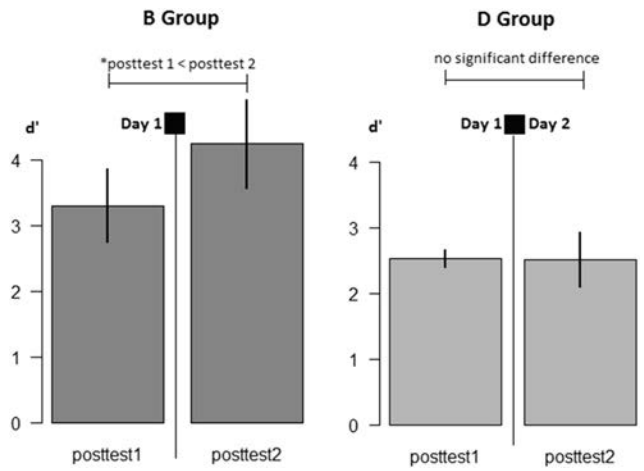

Figure 9. Profile of changes in identification performance by training group for Experiment 3. Error bars indicate standard error of the mean. * Indicates statistical significance at alpha = .05. See the online article for the color version of this figure.

To investigate the source of the Time × Group interaction in the trained vowel context, we ran two 2 × 2 mixed-models ANOVAs on the baseline and Posttest 1, and Posttest 1 and Posttest 2 scores. For baseline and Posttest 1, there was a significant main effect of time, $F(1, 32) = 31.109$, $p < .001$, $\eta^2 = .493$, but no main effect of group, $F(1, 32) = .696$, $p = .410$, $\eta^2 = .021$, nor an interaction between time and group, $F(1, 32) = .022$, $p = .737$, $\eta^2 = .004$. The direction of the Time main effect was determined by running a single paired samples *t* test on the baseline and Posttest 1 scores, collapsed across groups due to the lack of group main effect. Posttest 1 was significantly higher than baseline performance ($t_{33} = -5.654$, $p < .001$; 95% CI: [−.796; −.375]).

For the 2 × 2 mixed models ANOVA run on Posttests 1 and 2, there was a significant interaction between time and group, $F(1, 32) = 9.438$, $p = .004$, $\eta^2 = .228$, but no main effects (Time: $F(1, 32) = 2.218$, $p = .146$, $\eta^2 = .065$; group: $F(1, 32) = 2.153$, $p = .152$, $\eta^2 = .063$). We determined the source of the Time × Group interaction by running two separated paired samples *t* tests for each group, with Holms-Bonferroni correction applied for the calculation of CIs. For the D group, Posttest 2 scores were significantly lower than Posttest 1 scores ($t_{15} = 2.956$, $p = .009$; 95% CI: [−.891; −.209]). For the B group, Posttest 2 scores were significantly higher than Posttest 1 scores ($t_{15} = -3.933$, $p = .001$; 95% CI: [−.956; −.286]). Therefore, differences in discrimination accuracy across the two groups appear to emerge only after the overnight interval and not immediately following the interference block.

Discussion

Results from Experiment 3 replicate our discrimination findings in Experiment 2. First, both groups appear to achieve comparable gains in performance between baseline and Posttest 1 (see Figure 10). Overnight, their behaviors appear to diverge: The B group improves in performance, whereas the D group appears to decline, following sleep. This supports our interpretation that the effect of

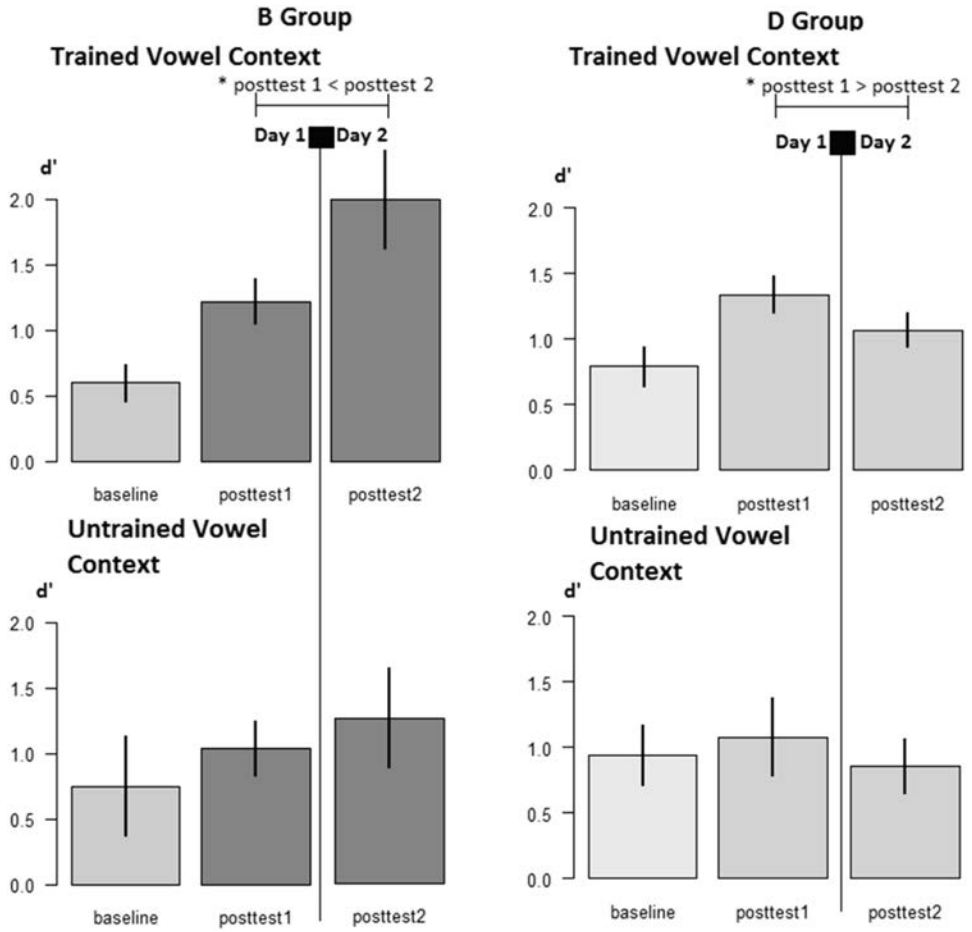

Figure 10. Profile of changes in discrimination performance by interference group and by vowel context (trained or untrained) for Experiment 9. Error bars indicate standard error of the mean. * Indicates statistical significance at alpha = .05. See the online article for the color version of this figure.

interference on discrimination performance is a latent phenomenon that does not emerge until posttraining sleep has taken place.

Identification performance diverges from the patterns observed in the Evening group in Experiment 1 and the B and D groups in Experiment 2. Specifically, the B group improved in identification performance overnight (similar to evening-trained groups in Experiments 1 and 2), whereas the D group in Experiment 3 did not. In other words, there was an effect of interference on identification performance in Experiment 3 that was not observed in Experiment 2. In previous literature, it has been demonstrated that learning interfering information immediately after training on the target information affects declarative recall during reassessment 12 hr posttraining (Ellenbogen, Payne & Stickgold, 2006). Thus, the interference effect observed in Experiment 3 is perhaps less surprising than the lack of interference effects observed for identification performance in Experiment 2.

The difference between Experiments 2 and 3 are primarily in the ordering of the interference block in relation to posttraining assessment. In Experiment 2, the interference block occurred after the Session 1 posttest, whereas in Experiment 3, the interference block occurred right before the Session 1 posttest. One thing to consider is that in Experiment 3, Posttest 1 followed an intervening period of some other activity. As such, it may have been memory reactivation prior to sleep that destabilized the declarative trace, as to make the trace susceptible to proactive interference from the interference block (see Dudai, 2004, for review). As a result, it seems, subsequent sleep had a stabilizing, but not enhancing, effect on identification performance. It should be noted that, statistically, the morning group's pattern does not differ from the evening-trained groups in Experiments 1 and 2 despite the appearance of relatively stable behavior over time (see Figure 3). As the morning group also experienced memory reactivation in Posttest 2 prior to sleep, this speculation regarding the effect of memory reactivation warrants further investigation.

General Discussion

The acquisition of non-native sounds poses a challenge for adult language learners. A lifetime of exposure to native language speech shapes a listener's sensitivity, and produces a perceptual system that struggles to distinguish non-native speech sounds that fall within a native category. One account posits that perceptual space around native speech categories is warped such that non-native tokens that are proximal in acoustic-phonetic space are

assimilated into that category (see Kuhl & Iverson, 1995). The current investigation highlights a different barrier to learning: native language interference prior to sleep-mediated consolidation. This work joins a growing literature implicating the role of sleep in consolidation of linguistic information. Previous work has examined sleep effects for lexical and grammatical learning (Dumay & Gaskell, 2007; Gomez, 2011), and for the perceptual learning of speech in one's native language, such as in adjusting speech sound boundaries to adjust for nonstandard speech tokens (Fenn et al., 2003, 2013).

Taken together with Experiment 1, results of Experiment 2 and 3 suggest that, although sleep affects listeners' ability to discriminate trained non-native sounds, this effect is mediated by the amount of exposure to a similar native-language sound (i.e., /d/) between training and sleep. In Experiment 2, listeners who heard a train of native /d/ sounds (which perceptually resemble /d̪/ and /d̪/) did not improve performance following sleep (D group), whereas listeners who heard the perceptually distinct tokens significantly improved following the overnight interval (B group), patterning similarly to the evening group from Experiment 1 (see Figures 4, 7, and 10). These results suggest that the decline in performance in the morning group in Experiment 1 following the overnight between-session interval is explained, at least in part, by the incidental exposure to the English /d/ prior to sleep.

Analogous to the auditory skill learning literature, we propose that the function of sleep in discrimination performance is to improve a listener's ability to automatically direct attention toward the acoustic cues in the signal that will aid him/her in distinguishing the non-native contrast. It has been suggested that learning to discriminate non-native tokens requires not a change in the sensitivity of the perceptual system, per se, but rather a change in how attention is allocated to portions of the signal that are relevant for the new sounds (e.g., Francis, Baldwin & Nusbaum, 2000; Francis & Nusbaum, 2002). This allocation of attention, considered as an auditory skill, is implicitly acquired within our training protocol. It has been suggested that procedural learning is, in the absence of interfering information, enhanced as the result of synaptic strengthening during REM sleep (Walker et al., 2003; Diekelmann & Born, 2007). The consequence of synaptic strengthening to perceptual learning may be in enhancing the automaticity with which attention is directed selectively to domain-specific features (Atienza et al., 2004).

Similar domain-specific interference effects have been reported in visual perceptual learning, particularly in cases in which interference stimuli overlap in retinotropic location to training stimuli (Yotsumoto et al., 2009; Seitz et al., 2005). A similar mechanism is proposed to be at work here between the non-native contrast and the /d/ tokens that overlap in acoustic-phonetic features. During the inference block, attention is repeatedly pulled to features relevant to the English /d/ category (rather than those for the Hindi contrast). Interfering stimuli may either destabilize the path of activation to the trained stimuli prior to sleep, or alternatively, sleep may strengthen the experience of attentional allocation to the learned tokens and the interference tokens indiscriminately, reinforcing connections between both learned and interference tokens and therefore decreasing the salience of the trained items upon waking.

The patterns of behavioral change we observed in identification performance differ from that in discrimination performance. In summary, five out of six groups appeared to improve in identification performance as a function of time (with the caveat that the pattern in the Morning group appears comparatively subtle, despite having a performance profile that is statistically comparable to the other groups'). In the introduction, identification performance was predicted to benefit from two separate sleep-mediated consolidation effects. First, as a declarative task, systems consolidation is thought to facilitate generalization to a different talker. This prediction was supported by our previous work (Earle & Myers, 2015) and was therefore not tested in the current set of studies. The second prediction was that the implicitly acquired auditory skill (modulation of attention) would enhance performance in both discrimination and identification tasks, provided that the training tokens are identical to those used in training. In most cases, identification performance did improve; however, it was not susceptible to the effects of passive interference in the same way that was observed in discrimination performance. There are at least two potential explanations for this. First, for the purposes of completing the identification task, the acquired ability to selectively attend to relevant stimuli may have been anchored to the visual stimulus, such that the skill was made accessible post-interference by the cue of the visual object. Second, the same sleep-mediated processes involved in increasing synaptic strength in local sensory cortices may also apply to the network connections involved in episodic recall. For example, it has been found that theta activity during REM increases not just after procedural learning, but after word-pair learning as well (Fogel, Smith & Cote, 2007). Therefore, though the precise mechanism is not yet understood, such evidence suggests that REM, and its association with latent synaptic consolidation, may also benefit performance on declarative tasks.

Only in one group, the D group in Experiment 3, exhibited what may be interpreted as a latent interference effect in identification performance. Specifically, the D group showed a pattern of stability, rather than improvement, following sleep, despite the D and B groups demonstrating comparable performance immediately following the interference block. This pattern was unexpected, and the possible explanations are speculative. However, a reasonable assumption is that the intervening time period between learning and assessment in Experiment 3 somehow made the D group susceptible to the effects of interference in the identification task. Therefore, by manipulating the ordering of tasks, we may have inadvertently changed the conditions under which the phonetic tokens were encoded. In the cases in which assessment immediately followed the training, the assessment phase may have been encoded as a continuation of the training event. In contrast, by inserting an approximately 15-min delay between training and assessment, those in Experiment 3 may have recruited the earlier (relatively stabilized) episodic trace, such that the assessment phase was encoded as a separate event involving the reactivation of the training episode. Episodic memory has been hypothesized to undergo a relatively short period of vulnerability upon reactivation, such that every instance of recall introduces an opportunity to corrupt or degrade the integrity of the original trace (see Dudai, 2004, for review). Upon reactivation, the trace may have been made susceptible to proactive interference by the preceding interference tokens, such that the reconsolidation of the token-label mapping during the assessment event were corrupted by the preceding bombardment of /dV/ stimuli. Again, this explanation is speculative, and more research is necessary to understand the

differences in timing of interference stimuli to identification performance.

Notably, our current results are inconsistent with our previous study (Earle & Myers, 2015) in that, in the previous study, we did not observe the changes to task performance in either task when the training tokens were used in assessment. Differences between the data for the current study and Earle and Myers (2015) may be attributable to the variability in the stimulus set in the previous investigation. In the previous work, the discrimination task contained three generalization conditions, with only (40) trials per condition. In other words, only 40 trials assessed discrimination of the trained tokens while an additional 120 trials assessed discrimination of unfamiliar tokens. Thus, low-level auditory input was not a reliable source of information; consequently, the input may have been too variable for participants to come up with an effective strategy for attending to relevant cues in the auditory signal. In the current investigation, we limited our generalization condition to just one (untrained vowel), and increased the number of discrimination trials in each condition, in order to facilitate improvement in perceptual tasks on the trained tokens.

The current findings provide no clear evidence of generalization of discrimination performance to an untrained vowel context (see Figures 4, 7, and 10). We have outlined in the introduction our reasons for suspecting that sleep-mediated generalization effects may be more salient in identification over discrimination performance. While decreased variability in the training set may have improved discrimination performance on the trained tokens, generalization to new phonological contexts may require more variability in the training set. Generalization to a new vowel context involves extraction of acoustic cues that distinguish the contrast, yet these acoustic cues may vary significantly across phonological contexts (see Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967, although see Stevens & Blumstein, 1982 for evidence that invariant acoustic cues to this distinction may be accessible in the signal). As such, it may be that training in one vowel context provides insufficient variability to enable listeners to generalize to new vowel contexts (Pisoni, 1992).

As a general caveat to this discussion, it is likely too simplistic to consider either perceptual task as being purely procedural or declarative—rather, task demands may manipulate the weights placed on different sources of information encoded by the two memory systems in parallel. For example, although our previous work (Earle & Myers, 2015) indicated that sleep facilitates generalization only in the identification task, we might suppose that eventually, either through time or exposure to phonetic variation, abstract information may increase its influence on discrimination performance of novel speech tokens as well.

In considering baseline performance and learning trajectories across experiments, it may be worthwhile to note that perceptual learning of non-native speech appears highly variable. Possible directions for future investigation are to determine specific sources of variability in non-native speech learning, such as quality/duration of sleep and susceptibility to interference, and contributions of individual differences such as language ability.

Conclusion

Our findings suggest that the successful discrimination of a new speech sound contrast, at least in the initial 24 hr, may depend on the amount of exposure to interfering stimuli prior to sleep. This may have broader implications for perceptual learning research in which training protocols span multiple days, or in studies of individual differences contributing to success in learning novel speech sounds.

References

Antoniou, M., Best, C. T., & Tyler, M. D. (2013). Focusing the lens of language experience: Perception of Ma'di stops by Greek and English bilinguals and monolinguals. *The Journal of the Acoustical Society of America, 133,* 2397–2411. http://dx.doi.org/10.1121/1.4792358

Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics, 40,* 582–594. http://dx.doi.org/10.1016/j.wocn.2012.05.005

Atienza, M., Cantero, J. L., & Stickgold, R. (2004). Posttraining sleep enhances automaticity in perceptual discrimination. *Journal of Cognitive Neuroscience, 16,* 53–64. http://dx.doi.org/10.1162/089892904322755557

Best, C. T. (1995). Learning to perceive the sound pattern of English. *Advances in Infancy Research, 9,* 217–217.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America, 109,* 775–794. http://dx.doi.org/10.1121/1.1332378

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer [Computer program]. Version 5.4.12, retrieved 10 July 2015 from http://www.praat.org/

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition, 97*(3), B45–B54. http://dx.doi.org/10.1016/j.cognition.2005.02.002

Brawn, T. P., Nusbaum, H. C., & Margoliash, D. (2010). Sleep-dependent consolidation of auditory discrimination learning in adult starlings. *The Journal of Neuroscience, 30,* 609–613. http://dx.doi.org/10.1523/JNEUROSCI.4237-09.2010

Clay, F., Bowers, J. S., Davis, C. J., & Hanley, D. A. (2007). Teaching adults new words: The role of practice and consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 970–976. http://dx.doi.org/10.1037/0278-7393.33.5.970

Davis, M. H., Di Betta, A. M., Macdonald, M. J., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21,* 803–820. http://dx.doi.org/10.1162/jocn.2009.21059

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 364,* 3773–3800. http://dx.doi.org/10.1098/rstb.2009.0111

Diekelmann, S., & Born, J. (2007). One memory, two ways to consolidate? *Nature Neuroscience, 10,* 1085–1086. http://dx.doi.org/10.1038/nn0907-1085

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11,* 114–126.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology, 55,* 51–86. http://dx.doi.org/10.1146/annurev.psych.55.090902.142050

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18,* 35–39. http://dx.doi.org/10.1111/j.1467-9280.2007.01845.x

Dumay, N., Gaskell, M. G., & Feng, X. (2004). A day in the life of a spoken word. In *Proceedings of the twenty-sixth annual conference of the cognitive science society* (pp. 339–344). Mahwah, NJ: Lawrence Erlbaum.

Earle, F. S., & Myers, E. B. (2014). Building phonetic categories: An argument for the role of sleep. *Frontiers in Psychology, 5,* 1192. http://dx.doi.org/10.3389/fpsyg.2014.01192

Earle, F. S., & Myers, E. B. (2015). Overnight consolidation promotes generalization across talkers in the identification of nonnative speech sounds. *The Journal of the Acoustical Society of America, 137*(1), EL91–EL97. http://dx.doi.org/10.1121/1.4903918

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

Ellenbogen, J. M., Payne, J. D., & Stickgold, R. (2006). The role of sleep in declarative memory consolidation: Passive, permissive, active or none? *Current Opinion in Neurobiology, 16,* 716–722. http://dx.doi.org/10.1016/j.conb.2006.10.006

Fenn, K. M., & Hambrick, D. Z. (2013). What drives sleep-dependent memory consolidation: Greater gain or less loss? *Psychonomic Bulletin & Review, 20,* 501–506. http://dx.doi.org/10.3758/s13423-012-0366-z

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature, 425,* 614–616. http://dx.doi.org/10.1038/nature01951

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–272). Timonium, MD: York.

Fogel, S. M., Smith, C. T., & Cote, K. A. (2007). Dissociable learning-dependent changes in REM and non-REM sleep in declarative and procedural memory systems. *Behavioural Brain Research, 180,* 48–61. http://dx.doi.org/10.1016/j.bbr.2007.02.037

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

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89,* 105–132. http://dx.doi.org/10.1016/S0010-0277(03)00070-2

Goedert, K. M., & Willingham, D. B. (2002). Patterns of interference in sequence learning and prism adaptation inconsistent with the consolidation hypothesis. *Learning & Memory, 9,* 279–292. http://dx.doi.org/10.1101/lm.50102

Gómez, R. L. (2011). Memory, sleep and generalization in language acquisition. *Experience, Variation and Generalization*. Learning a First Language, *7,* 261.

Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience, 10,* 100–107. http://dx.doi.org/10.1038/nn1825

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the" perceptual magnet effect. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Baltimore, MD: York Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461. http://dx.doi.org/10.1037/h0020279

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English/r/and/l/by Japanese bilinguals. *Applied Psycholinguistics, 2,* 369–390. http://dx.doi.org/10.1017/S0142716400009796

Macmillan, N. A., & Creelman, C. D. (2004). Detection theory: A user's guide. Aarhus, Denmark: Psychology press.

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences, 11,* 442–450.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience, 2,* 89–108.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419–457. http://dx.doi.org/10.1037/0033-295X.102.3.419

Pisoni, D. B. (1992). Talker normalization in speech perception. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, speech production, and linguistic structure* (pp. 143–151). Tokyo, Japan: OHM.

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews, 93,* 681–766. http://dx.doi.org/10.1152/physrev.00032.2012

Roth, D. A. E., Kishon-Rabin, L., Hildesheimer, M., & Karni, A. (2005). A latent consolidation phase in auditory identification learning: Time in the awake state is sufficient. *Learning & Memory, 12,* 159–164. http://dx.doi.org/10.1101/87505

Schwartz, S., Maquet, P., & Frith, C. (2002). Neural correlates of perceptual learning: A functional MRI study of visual texture discrimination. *Proceedings of the National Academy of Sciences, USA of the United States of America, 99,* 17137–17142. http://dx.doi.org/10.1073/pnas.242414599

Seitz, A. R., Yamagishi, N., Werner, B., Goda, N., Kawato, M., & Watanabe, T. (2005). Task-specific disruption of perceptual learning. *Proceedings of the National Academy of Sciences, USA of the United States of America, 102,* 14895–14900. http://dx.doi.org/10.1073/pnas.0505765102

Stevens, K., & Blumstein, S. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. l–38). Hillsdale, NJ: Lawrence Erlbaum.

Strange, W. (1995). Book review: The development of speech perception: The transition from speech sounds to spoken words by J. C. Goodman and H. C. Nusbaum. *Language and Speech, 38,* 217–222.

Swan, K., & Myers, E. (2013). Category labels induce boundary-dependent perceptual warping in learned speech categories. *Second Language Research, 29,* 391–411.

Tamminen, J., Davis, M. H., Merkx, M., & Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition, 125,* 107–112. http://dx.doi.org/10.1016/j.cognition.2012.06.014

Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature, 425,* 616–620. http://dx.doi.org/10.1038/nature01930

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24,* 672–683. http://dx.doi.org/10.1037/0012-1649.24.5.672

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science, 265,* 676–679. http://dx.doi.org/10.1126/science.8036517

Yotsumoto, Y., Chang, L. H., Watanabe, T., & Sasaki, Y. (2009). Interference and feature specificity in visual perceptual learning. *Vision Research, 49,* 2611–2623. http://dx.doi.org/10.1016/j.visres.2009.08.001

Received December 1, 2014
Revision received June 18, 2015
Accepted June 23, 2015 ■

See page 1708 for a correction to this article.