

Sleep facilitates generalisation of accent adaptation to a new talker

Xin Xie, F. Sayako Earle & Emily B. Myers

To cite this article: Xin Xie, F. Sayako Earle & Emily B. Myers (2018) Sleep facilitates generalisation of accent adaptation to a new talker, *Language, Cognition and Neuroscience*, 33:2, 196-210, DOI: [10.1080/23273798.2017.1369551](https://doi.org/10.1080/23273798.2017.1369551)

To link to this article: <https://doi.org/10.1080/23273798.2017.1369551>

 View supplementary material 

 Published online: 29 Aug 2017.

 Submit your article to this journal 

 Article views: 108

 View related articles 

 View Crossmark data 

REGULAR ARTICLE



Sleep facilitates generalisation of accent adaptation to a new talker

Xin Xie ^a, F. Sayako Earle^b and Emily B. Myers^{c,d,e}

^aDepartment of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA; ^bDepartment of Communication Sciences & Disorders, University of Delaware, Newark, DE, USA; ^cDepartment of Speech, Language, and Hearing Sciences, University of Connecticut, Storrs, CT, USA; ^dDepartment of Psychology, University of Connecticut, Storrs, CT, USA; ^eHaskins Laboratories, New Haven, CT, USA

ABSTRACT

Lexically-guided phonetic retuning helps listeners adapt to the phonetic “fingerprint” of a talker. Previous findings show that listeners can generalise from one accented talker to another accented talker, but only for phonetically similar talkers. We tested whether sleep-mediated consolidation promotes generalisation across accented talkers who are *not* phonetically similar. Native-English participants were trained on a Mandarin-accented talker and tested on this talker and an untrained Mandarin talker. Experiment 1 showed adaptation for the trained talker and a weak transfer to the untrained talker. In Experiment 2, participants were trained and tested either in the morning (Same-Day group) or evening (Overnight group), and again after twelve hours. Both groups retained talker-specific learning over the 12-hour delay. Importantly, the Overnight group showed improvements for the untrained talker, whereas the Same-Day group’s performance on the untrained talker deteriorated. We suggest that sleep facilitated talker generalisation by helping listeners abstract away from specific acoustic properties of the trained talker.

ARTICLE HISTORY

Received 21 February 2017
Accepted 13 August 2017

KEYWORDS

Perceptual learning; sleep consolidation; foreign-accented speech; talker generalisation; memory

Human listeners comprehend speech by mapping the speech signal onto linguistically meaningful categories. The speech signal itself, however, is inherently noisy in that the mapping between the acoustic signal and phonetic categories is highly variable across speech contexts (listening conditions, speakers, accents, and so on). By examining the way listeners cope with unfamiliar variants within the native speech processing system, a wide array of studies has revealed that listeners rapidly overcome initial difficulty in processing non-standard speech signal via top-down-guided adaptation (e.g. Clarke & Garrett, 2004; Norris, McQueen, & Cutler, 2003; Trude & Brown-Schmidt, 2012). Indeed, rapid perceptual learning for speech has been documented for many types of speech variants, including acoustically-distorted speech (e.g. noise-vocoded, time-compressed) or naturally-produced speech containing non-canonical tokens (e.g. dialectal or foreign-accented). Meanwhile, our understanding of the mechanism by which learning can be generalised from specific contexts to new instances is limited. In particular, it continues to be unclear how the online speech processing system interfaces with memory systems to build robust representations that can be used for the long run, yet can also adapt to different perceptual criteria needed for different talkers and accents. In the current study, we address this issue by asking how listeners, having adapted to the

phonetic characteristics of a specific foreign-accented talker, retain this learning experience in memory and more importantly, how they generalise perceptual performance to other talkers. Specifically, we investigate the role of sleep in this process. In order to formulate a hypothesis regarding the involvement of memory systems in talker generalisation, it is first necessary to consider the conditions under which listeners are able to generalise perceptual learning from one talker to a new talker.

In considering adaptation to atypical speakers (in the form of unfamiliar idiolects, regional or foreign accents), evidence suggests that as listeners encounter unfamiliar speakers whose productions do not align well with canonical forms of speech and create auditory ambiguity, they use higher-level linguistic knowledge to resolve perceptual uncertainty and retune acoustic-to-category mappings accordingly (e.g. McQueen, Cutler, & Norris, 2006; Norris et al., 2003). For instance, after hearing an ambiguous sound midway between /d/ and /t/ embedded in words such as “kingdom”, the lexical information biases listeners to interpret the ambiguous signal as /d/ and results in increased tendency to interpret other similarly ambiguous sounds along a /d/-/t/ continuum as members of the /d/ category (Kraljic & Samuel, 2006). Such lexically-guided phonetic retuning is found to support adaptation to foreign-accented

CONTACT Xin Xie  xxie13@ur.rochester.edu

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/23273798.2017.1369551>

© 2017 Informa UK Limited, trading as Taylor & Francis Group

speech as well (Sumner, 2011; Xie, Theodore, & Myers, 2017). This type of talker-specific adaptation seems to be long-lasting to the extent that improvement in word recognition is observed up to one week after initial learning (Eisner & McQueen, 2006; Witteman, Bardhan, Weber, & McQueen, 2015).

Given that talkers vary so extensively in their production of speech sounds, understanding how listeners generalise learning experience to novel talkers is of great theoretical importance. Practically speaking, this would help us to devise training paradigms that maximise real-world benefits of perceptual adaptation, which inevitably involve dealing with multiple speakers. A few studies have investigated how perceptual learning effects generalise to novel talkers. Early investigations along this line have yielded mixed results. Specifically, the generalisability of learning seems to be dependent on the particular sound categories that are learned. For fricatives, listeners do not transfer what they have learned about one talkers' shifted category boundary to a different talker (Kraljic & Samuel, 2007). Furthermore, using a delayed test, Eisner and McQueen (2006) found that intervening speech stimuli from other talkers do not undermine perceptual learning results for a learned talker. These results suggest that listeners keep person-specific adjustments separate for fricatives. In comparison, phonetic adjustments generalised to different talkers for stop consonants (Kraljic & Samuel, 2005, 2007), leading to the suggestion that stop categories may not be as closely linked to talker identity as are fricative categories. Indeed, the productions of fricatives seem to be acoustically more variable across talkers than stops (e.g. Allen, Miller, & DeSteno, 2003; Newman, Clouse, & Burnham, 2001). It is possible that listeners have tacit knowledge of these distributions in prior native experience and thus demonstrate different generalisation behaviors (see Feldman, Griffiths, & Morgan, 2009; Kleinschmidt & Jaeger, 2015 for discussions on listeners' sensitivity to such distributions in language).

More recent evidence points to talker similarity as a factor that constrains generalisation effects in immediate tests following adaptation. Using a similar paradigm as Kraljic and Samuel (2007), Reinisch and Holt (2014) showed that listeners readily generalised the retuned phonetic boundary between /s/ and /f/ to a novel talker, when, and only when, the trained talker and the test talker had similar degrees of ambiguity in their productions of the learned fricative. In our own work (Xie & Myers, 2017), we have extended this result and found that listeners were not only sensitive to the degree of ambiguity in the speech sounds, but also to the relevant acoustic dimensions that are responsible for this

ambiguity. In particular, the degree of generalisation across talkers depends on the acoustic-phonetic distributions of each talker and the cross-talker alignment at this level. In this study, native-English listeners were exposed to a Mandarin-accented speaker whose productions of word-final /d/ sounds were perceptually ambiguous to the listeners. After a brief exposure phase, listeners were tested in their recognition of words produced by an unfamiliar Mandarin-accented speaker. Critically, several acoustic cues (for example, shorter durations of preceding vowels and longer bursts of a stop sound) are associated with a more /t/-like percept than with a /d/-like percept, and productions from the tested speakers varied in the distributional properties along these acoustic dimensions. Test results from multiple training-test talker pairs indicated that listeners showed the strongest generalisation when distributional properties of these acoustic cues were shared across talkers. Despite the fact that all speakers shared the same foreign accent and produced ambiguous /d/ sounds, listeners more readily generalised between talker pairs whose productions had the same mean distributions of several phonetically-relevant properties. In contrast, little evidence of generalisation was observed when the productions of speakers failed to show sufficient overlap in the acoustic-phonetic space. These results suggested that while perceptual adaptation occurs quickly, generalisation effects, when tested shortly after learning, were restricted to physically similar stimuli (i.e. acoustically-similar speakers).

This finding parallels empirical evidence from other aspects of human learning (e.g. novel word learning, Friedrich, Wilhelm, Born, & Friederici, 2015) showing that the initial encoding of learning materials tends to be context-specific, and online generalisation, if any, relies on low-level properties. What is yet unknown is whether listeners' ability to generalise from one talker to another continues to evolve in the offline stage following the initial training. In particular, we ask whether phonetic retuning can show evidence of generalisation after listeners have had the opportunity to consolidate the adaptation experience during sleep. Of note, talker differences represent lawful variability in the real world and listeners already have lifelong experience regarding how such variability is structured across speakers within their native language. In theory, listeners may be able to take advantage of this experience to hypothesise a set of systematic rules or mappings by which the signal differs from expectation (such as in foreign-accented speech). If listeners can abstract from the distributional patterns available in the input to a more general scheme for categorising sounds from a similarly accented talker, this will

facilitate subsequent online processing of other talkers with the same accent.

There is reason to suppose that sleep-mediated memory consolidation may help listeners to abstract away from the specific acoustic details in the training stimuli and facilitate generalisation to new instances or contexts. The *Complementary Learning Systems Account* (McClelland, McNaughton, & O'Reilly, 1995) predicts that during offline consolidation, there is integration of experience with prior knowledge, abstracting salient features away from the episodic details in the process (termed memory “triage” by Stickgold & Walker, 2013). Relevant examples for sleep-mediated abstraction include nap-dependent learning of grammatical rules (Hupbach, Gomez, Bootzin, & Nadel, 2009) and statistical learning of auditory tones (Durrant, Cairney, & Lewis, 2013; Durrant, Taylor, Cairney, & Lewis, 2011). In the domain of speech sound learning, there is emerging evidence that sleep facilitates generalisation of task performance to novel speech stimuli (Earle & Myers, 2015a; Fenn, Nusbaum, & Margoliash, 2003). For example, Fenn et al. (2003) trained listeners to map an open set of synthetic-speech tokens onto a native lexicon, with a new token used at every trial. In this sense, the mode of training targeted generalisation specifically. The authors found that those who were trained in the evening maintained their training-induced task performance over 24 h, whereas those trained in the morning appeared to degrade in performance over the course of the day, recovering their performance following sleep. In training listeners on a non-native contrast using a closed token set, Earle and Myers (2015a) found that listeners who were trained to identify a phonetic contrast from one non-native talker performed at chance on identification of the same phonetic contrast from an untrained talker when tested immediately after training. However, performance significantly improved after an overnight interval, suggesting that sleep facilitated talker generalisation. These reports motivate our prediction that sleep may facilitate the abstraction of relevant acoustic features for foreign-accented speech, resulting in a more generalised perceptual adaptation that is applicable across changing talkers of a foreign accent.

We explicitly test this hypothesis by examining the generalisation patterns across a talker pair who both use the same acoustic dimensions to cue a critical phonetic contrast, but differ in the exact distributions of acoustic values along each dimension. Based on previous work from our lab and others (Reinisch & Holt, 2014), this talker pair should show little generalisation from one talker to the other because of surface differences in the distribution of the phonetic tokens. Crucially, however,

both talkers utilise the same acoustic dimension to cue the phonetic contrast, offering an *opportunity* for listeners to generalise if they can learn this higher-order structure. To this end, the exposure-test talker pair was selected such that both talkers produced /d/ sounds that often are misheard as /t/ sounds by English listeners, and both used an acoustic cue (burst length) that was not typically used by native-English listeners for voicing contrasts (see description under Speech Materials in **Methods**). We adapted a perceptual learning paradigm that has been used in our previous work to examine talker accent learning. In Xie et al. (2017), we found talker-specific adaptation following brief exposure to phonetic variation in speech produced by an individual with a foreign accent. The same paradigm was applied to the two experiments in the current study. Experiment 1 was designed to replicate talker-specific adaptation and to investigate whether such adaptation is generalisable to an acoustically-dissimilar novel talker with the same accent when tested immediately post-exposure. Experiment 2 examined whether talker-specific adaptation was retained over 12 h and, crucially, tested the effect of sleep on talker generalisation. We hypothesised that talker generalisation would not be observed after a 12-hour delay that did not contain sleep, but *would* be observed after an overnight delay.

Experiment 1

Experiment 1 was designed to establish the baseline talker-adaptation effects to which sleep effects will be compared. We intended to replicate the talker-specific adaptation effect as previously reported in Xie et al. (2017). In particular, we examined whether adaptation to a foreign-accented talker (henceforth Trained Talker), whose speech was heard during a brief accent exposure, was generalisable to a novel talker (henceforth Generalisation Talker) in the immediate test following exposure. The experiment consisted of an exposure phase and a test phase (see Figure 1 for a schematic of exposure and test phases for Experiments 1 and 2). All participants completed an exposure phase during which they listened to speech materials from the Trained Talker and completed a lexical decision task. The exposure consisted of a list of words and non-words spoken by the Trained Talker. Critically, for half of the participants (/d/-exposure group), the list contained some multisyllabic words ending in /d/ (e.g. “overload”) which a listener could potentially use to adapt to the way word-final /d/ tokens are produced in Mandarin-accented English (see Speech Materials, below). The other half of the participants (Baseline group) heard the same list of words, except that the /d/-final critical items were replaced

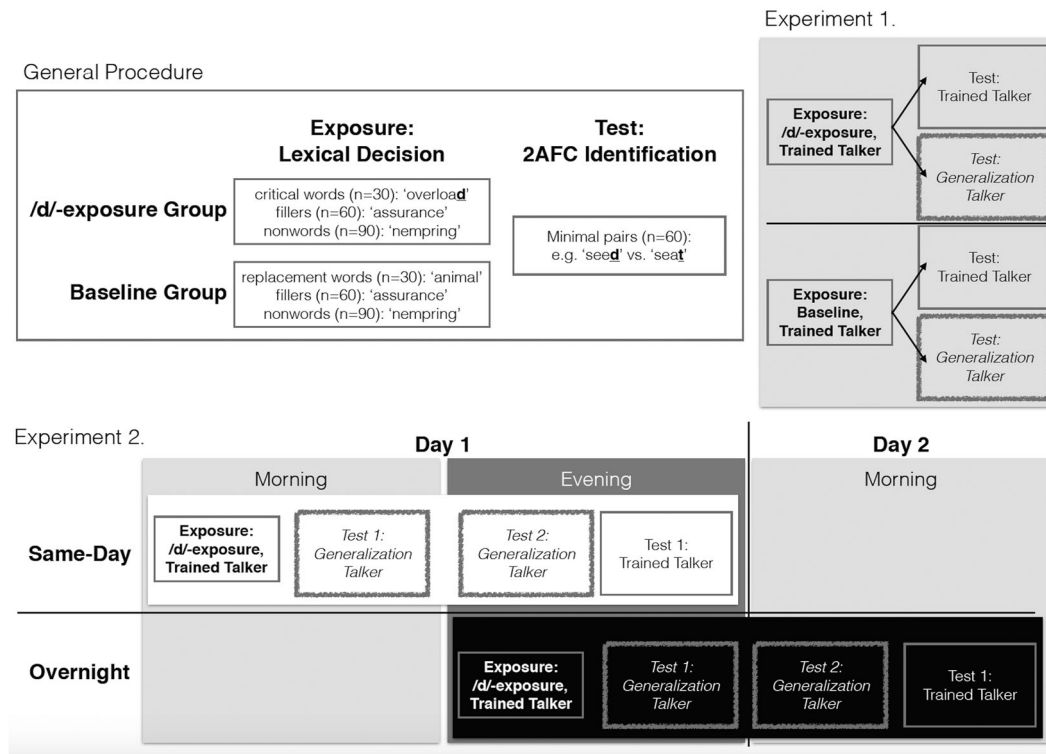


Figure 1. Schematic of testing procedures for Experiments 1 and 2.

with filler items. In addition, for both groups, there were no /t/ tokens or other stop sounds in word-final positions. Theoretically, the Baseline group has no evidence in the input that could be used to adapt to differences in the way that Mandarin-accented talkers realise word-final voiced stops.

To test for adaptation to the accented /d/ tokens, participants categorised the final consonants of minimal pairs of /d/- and /t/-final words (e.g. "seed" vs. "seat"). The two groups (/d/-exposure vs. Baseline) were further subdivided: half were tested with the Trained Talker to test for talker-specific adaptation and half with the Generalisation Talker to examine evidence for generalisation across talkers (Figure 1). Within each test talker condition, the critical manipulation was whether participants heard /d/-final words during exposure (/d/-exposure group) or not (Baseline group).

Note that instead of using typical signal detection measures that assess discrimination sensitivity (d') or response bias (c or β), we examined perceptual responses to the trained category /d/ and the untrained category /t/ separately for the following reasons. First, similar studies on the learning of talker-related production patterns (e.g. Kraljic & Samuel, 2007; Norris et al., 2003; Reinisch & Holt, 2014; Xie et al., 2017) have typically reported proportion or percent responses in categorisation and we want to keep our results comparable.

Second and more importantly, it is of theoretical interest to distinguish learning outcomes for the trained category, /d/, and the untrained category, /t/, and using d' measures would not reveal any differences between the two categories. Specifically, we predict that the presence vs. absence of /d/ exposure (i.e. the critical contrast between /d/-exposure group and Baseline group) will change listeners' perception of /d/ category, but would not necessarily alter how they categorise /t/ sounds. We make this distinction for two reasons. First, because /t/ tokens were absent in the exposure phase (for both groups), listeners did not have direct experience of /t/ sounds from the Trained Talker. It is possible that listeners treat /d/ and /t/ distributions as independent such that learning of the acoustic properties of /d/ does not affect listeners' hypotheses about how /t/ sound like in this unfamiliar non-native accent. Recent research has yielded results consistent with this possibility (Eisner, Melinger, & Weber, 2013). Second, Mandarin-accented /d/ and /t/ differ in terms of their acoustic-phonetic proximity to corresponding categories in native-accented English. While Mandarin voiced stops are perceptually ambiguous and often perceived as voiceless, voiceless tokens can usually be recognised with much greater accuracy (e.g. Xie & Fowler, 2013). That is to say, Mandarin-accented /d/ tokens lie in a perceptually more ambiguous region in the acoustic-phonetic space than /t/ tokens do. So even in the case where listeners

do not treat /d/ and /t/ distributions as independent and make adjustments for /t/ category in the absence of direct exposure to /t/ sounds, we still expect such adjustments to be minimal, given the fact that past work have observed salient phonetic retuning effects in the ambiguous region along a nonword-nonword continuum (e.g. /ada-/ata/) but little changes in the unambiguous regions (e.g. Reinisch & Holt, 2014). For these reasons, we assess perceptual responses to /d/- and /t/- words separately and expect to see signals of perceptual learning primarily for /d/ category.

Relative to baseline control participants, we expected increased /d/ responses in the categorisation test task among the /d/-exposure group for the Trained Talker's productions. If a similar pattern is observed for the productions of the Generalisation Talker as well, then it suggests that listeners generalise perceptual learning results to the novel talker. Conversely, an absence of a group difference (/d/-exposure group vs. Baseline group) in categorising /d/ sounds from the Generalisation Talker will suggest a lack of talker generalisation. Based on our previous investigations on talker generalisation, we hypothesised that immediate cross-talker generalisation is unlikely to appear for the Generalisation Talker, given that this talker has distinctive acoustic distributions in /d/ productions from the Trained Talker (see detailed description under Speech Materials in **Methods**).

Methods

Participants

Forty-eight monolingual English speakers with no hearing or visual problems participated in this experiment. Equal numbers of participants were randomly assigned into one of the four sub-groups: 2 Exposure Condition (/d/-exposure vs. Baseline) $\times$ 2 Test Talker (Trained Talker vs. Generalisation Talker). All participants were undergraduates at University of Connecticut. They gave informed consent according to the University of Connecticut Institutional Review Board and received course credits for their participation. All participants reported having minimal prior experience with Mandarin-accented English or the Mandarin language. We did not control for the time of day at which participants were tested. Thus, testing times varied throughout the day from morning to evening.

Speech materials

Exposure stimuli consisted of 30 multisyllabic critical words, 60 filler words and 90 nonwords. Critical words were 30 /d/-final words (e.g. *overload*) for the /d/-exposure group and these were replaced by 30 extra

filler words (e.g. *animal*) for the Baseline group. All critical words, filler words and nonwords had 3–4 syllables, with /d/ tokens appearing only in the critical words. In addition, participants heard no other alveolar stops or other voiced stops in the experiment; voiceless stops (/p/ or /k/) occurred only in word-initial position. Test stimuli were identical for both exposure groups, consisting of 60 monosyllabic minimal pairs ending in /d/ or /t/ (e.g. *seed* – *seat*). The test stimuli were organised 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.

The Trained Talker recorded both exposure and test stimuli. The Generalisation Talker recorded only test stimuli. Both talkers were L2 learners of English and acquired English in Mainland China, although the Trained talker was more proficient than the Generalisation talker. The two talkers used the same acoustic property to distinguish between /d/ and /t/ tokens, one that differs from the native American-English acoustic distributions (See Figure 2). Specifically, for word-final stop sounds, American-English talkers reliably show differences in the length of the preceding vowel (longer vowels for /d/ compared to /t/), whereas the two Mandarin-accented talkers do not show vowel duration differences before voiced vs. voiceless stops. Instead, both talkers cue the voicing contrast by manipulating the length of the final stop burst, with longer burst lengths for voiceless compared to voiced stops. Crucially, burst length is typically not used by native-English listeners to signal word-final stop voicing. Recordings were made in a sound-proof room using a microphone linked to a digital recorder, digitally sampled at 44.1 kHz and normalised for root mean square (RMS) amplitude to 70 dB SPL.

Procedure

Each participant completed an auditory lexical decision task during exposure, which was immediately followed by a two-alternative, forced-choice (2AFC) phonetic category identification task at test. Figure 1 illustrates the experimental procedures.

Lexical decision exposure phase. During the exposure phase, participants heard 180 words produced by the Trained Talker from either the /d/-exposure list or the Baseline list. Items were presented in a random order. Participants were instructed to decide whether each auditory stimulus was a real English word and to press a yes/no button as quickly as possible without sacrificing accuracy. Ten practice trials were given to the participants before the actual task to familiarise them with the task procedure. The practice items were not used

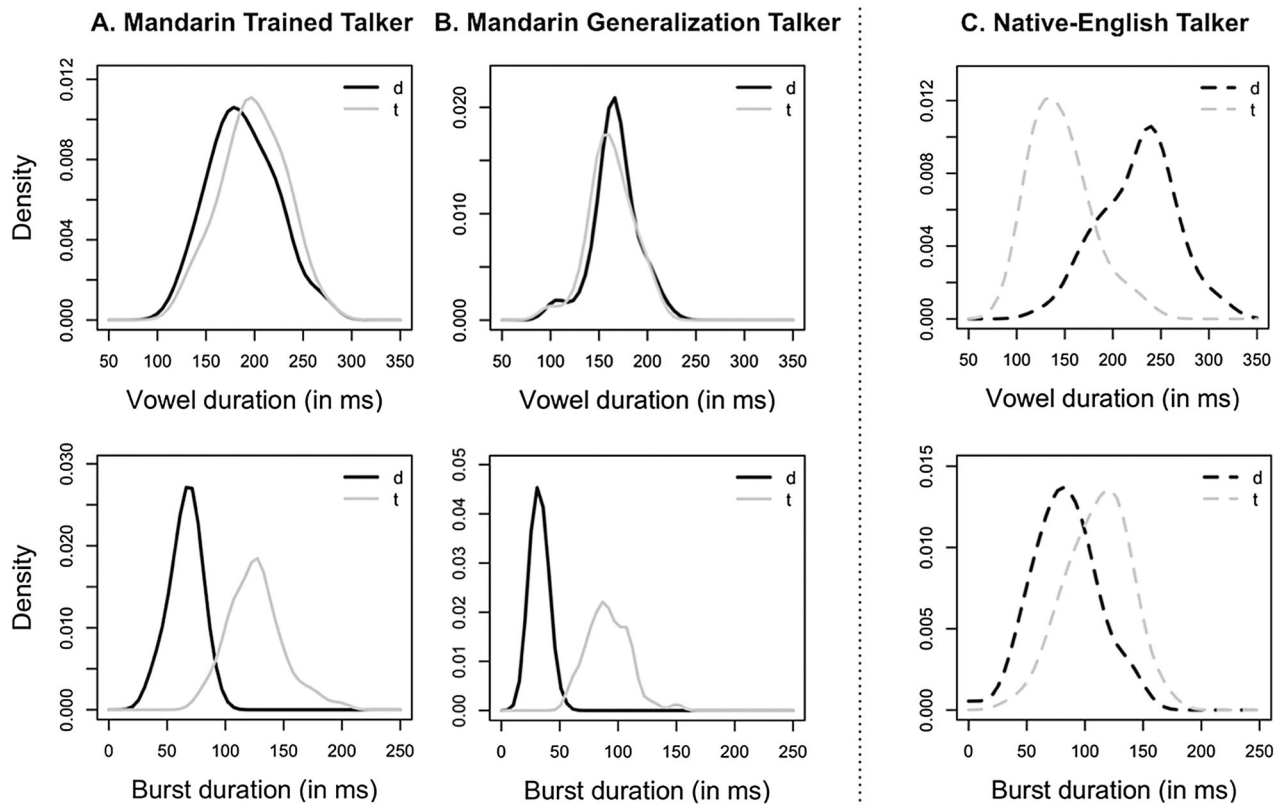


Figure 2. Probability density plots of acoustic measures (vowel duration and burst duration, respectively) across all /d/-final vs. /t/-final words used at test. Panel A: a Mandarin-accented speaker serving as the Trained Talker; Panel B: a Mandarin-accented speaker serving as the Generalisation Talker; Panel C: a native-English talker for comparison purposes. Note: the Native-English talker clearly marks voiced /d/ with longer vowels than /t/ tokens. In contrast, this difference is not observed with the two Mandarin talkers used in Experiments 1 and 2; their vowel durations are closer to vowel durations before /t/ than that before /d/ in native-accented English. On the other hand, both Mandarin talkers produced /d/s with shorter bursts than /t/s, yielding the burst length as a more informative cue for the voicing contrast in this accent, even though the exact distributions were not aligned between the two talkers.

in the actual exposure task. Each trial was preceded by a 1000 ms fixation cross at the center of the screen and was presented with an inter-onset interval of 3000 ms. No feedback was provided.

2AFC identification test phase. During the test phase, 120 test items were presented in two blocks, with an inter-trial interval of 2000 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. The experiment was presented using Eprime 2.0.10 running on a desktop computer. Audio stimuli were delivered via Sennheiser HD280 headphones at a comfortable listening level constant across participants.

Results

Response accuracies for the lexical decision exposure phase are presented in Table A1 (see Supplemental Materials). Accuracies for each type of words were

comparable between the /d/-exposure group and the Baseline group. Critical /d/-final words were largely judged to be real words by the /d/-exposure group ($M = .79$, $SD = .10$). To assess talker learning and generalisation effects, logistic mixed-effects models were used to analyze the probability of /d/ responses the 2AFC categorisation test task, separately for /d/ and /t/ words (Figure 3). We used the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015) to conduct the analysis. Exposure Condition was contrast coded, included as the fixed effect (/d/-exposure = 1 vs. Baseline = -1) and by-item and by-subject intercepts were included as random effects. For the dependent measures, /d/ responses were coded as 1 and /t/ responses were coded as 0.

Talker-specific adaptation

For the Trained Talker, the /d/-exposure group reported significantly more /d/ responses for words that were intended by the talker to be /d/-final ("seed") than the Baseline group ($\beta = .30$, $SE = .15$, $p < .05$), indicating evidence of perceptual



Figure 3. Mean percent /d/ responses for the 2AFC identification task in Experiment 1 as a function of Group (/d/-exposure vs. Baseline) for the two word types for each talker. Error bars indicate standard errors of the mean.

adaptation. Consistent with our prediction, the two groups showed no significant difference in their responses to /t/ words ($\beta = .04$, $SE = .16$, $p = .80$).

Talker generalisation

For the Generalisation Talker, there was no group difference in the responses either for /d/ words ($\beta = .19$, $SE = .19$, $p = .32$) or for /t/ words ($\beta = .0003$, $SE = .18$, $p = .99$).

Given the qualitatively different patterns between the two talkers for the trained category /d/, we examined whether an interaction between Test Talker and Exposure Condition was present in the responses for /d/ words, in order to statistically test for the difference between the Trained Talker and the Generalisation Talker in terms of the learning effect. Test Talker was contrast coded (Trained Talker: 1; Generalisation Talker: -1). Fixed effects included Test Talker, Exposure Condition and their interaction; random effects were by-item and by-subject intercepts. Results revealed a significant main effect of Test Talker ($\beta = .85$, $SE = .11$, $p < .0001$), with, as expected, more overall /d/ responses for /d/-final words for the Trained Talker than for the

Generalisation Talker. The learning outcome evident in the main effect of Exposure Condition was marginally significant ($\beta = .20$, $SE = .11$, $p = .056$), whereas the Exposure Condition-by-Test Talker interaction did not reach significance ($\beta = .03$, $SE = .11$, $p = .80$). Thus, consistent with what we see by visually inspecting Figure 3, the group patterns for /d/ words seem to be comparable between the Trained Talker and the Generalisation Talker, even though the learning effect for the latter was not statistically significant. On balance, exposure to the Trained Talker facilitated the perception of this talker's production of /d/ tokens, rendering the /d/-exposure group an advantage over the Baseline group, and such learning had a non-significant transfer effect onto the Generalisation talker, showing very weak generalisation in the test phase immediately following the initial accent exposure.

Discussion

Overall, the data show clear changes in categorisation judgments for the trained category /d/ and not for the

untrained category /t/. As noted in the Introduction, we suggest that the lack of adjustments for /t/ category were expected, due to the absence of /t/ tokens during exposure and the acoustic properties of Mandarin-accented English. We leave it for future studies to investigate whether exposure to exemplars from both contrastive categories would lead to a different behavioral pattern for /t/ sounds. We now turn to the interpretation of exposure-induced changes for the trained category /d/.

Increased /d/ responses for /d/-final words among the /d/-exposure group (relative to the Baseline group) for the Trained Talker replicated our previous finding on talker-specific adaptation. It is also consistent with past research (e.g. Kraljic & Samuel, 2006; Norris et al., 2003) that shows a shift of category boundary towards a trained category after lexically-guided phonetic retuning. Meanwhile, a similar trend was observed in the categorisation of /d/ sounds produced by the Generalisation Talker, although the group difference did not reach significance. Nevertheless, given the lack of a significant interaction between Exposure Condition and Talker, the results together pointed to weak evidence of generalisation to the novel talker when the two talkers were not sufficiently similar in the distributional properties of their acoustic-phonetic productions of the critical segment. This weak generalisation is consistent with similarity-modulated generalisation in the period immediately after the initial exposure (Reinisch & Holt, 2014; Xie & Myers, 2017). It is highly plausible that although the /d/-exposure group had not yet discovered a higher-order structure (using the burst length to distinguish /d/ from /t/) shared by the two talkers, token-based similarity among the words contributed to the modest generalisation. We return to this issue and offer our interpretations in the context of other evidence of immediate generalisation in the General Discussion. In the next experiment, we explored whether sleep-mediated consolidation increases generalisation of learned properties of the accent across talkers.

Experiment 2

In Experiment 2, we examined whether sleep promotes generalisation of adaptation from one Mandarin-accented talker (Trained Talker) to another talker (Generalisation Talker) with the same accent. The perceptual learning paradigm in Experiment 1 was adapted such that participants were tested twice for their performance with the Generalisation Talker: immediately after exposure and again after a 12-hour delay. All participants heard the critical /d/-final words during exposure, and as such, the exposure phase was identical to the “/d/-

exposure” group from Experiment 1 (refer to Figure 1 for task schedule). Critically, we compared the within-subject improvement in two groups: Overnight group and Same-Day group. For the Overnight group, the between-session interval was overnight (thus presumed to contain sleep for the majority of participants), whereas the Same-Day group returned after 12 h of daytime (and, for most participants, wake-state) activity. We examined whether there was increased generalisation to the Generalisation Talker in each group after 12 h, and more importantly, whether there was a difference between the Overnight and Same-Day groups in their generalisation pattern. As in Experiment 1, Generalisation would be manifested by an increase in /d/ responses for /d/-final words. A clear generalisation pattern is hypothesised to emerge only when the interval between exposure and test contains sleep (Overnight group). The Same-Day group served as a control for the time delay: if generalisation is promoted by any time delay, then we would see enlarged generalisation effects in both Overnight and Same-Day groups; if sleep is critical to promote generalised learning, then we would see a dissociation between the two groups in terms of the pattern and/or magnitude of improvement in their performance for the Generalisation Talker. In addition, we tested for the retention effect of talker-specific adaptation. Empirical evidence is limited in this regard, but extant studies have reported stable learning effects for an adapted talker over time in a 25 min, 12-h and one week interval, with no attenuation over time nor additional benefit from sleep (Eisner & McQueen, 2006; Kraljic & Samuel, 2005; Witteman et al., 2015). Thus, we expected talker-specific effects to be stable over the 12-hour interval. To examine this, we compared the performance of Overnight and Same-Day groups to the groups in Experiment 1, who were tested immediately after exposure. For clarity, we adopted a longer version of the group labels in the across-experiments comparisons, in order to highlight the critical differences between different groups of participants. *Immediate Test: /d/-exposure* and *Immediate Test: Baseline* refer to the two groups in Experiment 1 who were tested with no delay; *Same-Day: /d/-exposure* and *Overnight: /d/-exposure* refer to the two groups in Experiment 2 who had an identical exposure phase as the /d/-exposure group in Experiment 1 but were retested after a 12-h delay.

Methods

Participants, materials and procedure

Thirty-eight students from University of Connecticut participated in this experiment. All participants completed

the experiment in two sessions, with 12 h apart. The Overnight group ($n=20$) completed the exposure phase and their first test session between 8 and 9 PM and the second test session between 8 and 9 AM the next day. The Same-Day group ($n=18$) completed the first session between 8 and 9 AM and the second session between 8 and 9 PM on the same day.

The exposure and test word stimuli were identical to those used in Experiment 1 in the “/d/-exposure” group. During the first session, listeners were exposed to words and nonwords produced by the Trained Talker and were tested immediately on the minimal /d/-/t/ pairs produced by the Generalisation Talker. During the second session, there was no exposure phase and listeners were administered two blocked categorisation tests: Generalisation Talker first and then Trained Talker, to avoid any carryover effects from categorising the Trained Talker to Generalisation Talker, and thus to ensure that performance for the Generalisation Talker was the result of the initial exposure phase only.

Results

We conducted separate analyses on the categorisation results for the Trained Talker and the Generalisation Talker, in order to independently assess the maintenance of talker-specific adaptation and the development of generalisation effects over the 12 h.

Talker-specific adaptation

We first asked whether Same-Day and Overnight groups differed in the degree to which talker-specific adaptation was maintained over a 12 h interval that either did contain sleep (Overnight group) or did not (Same-Day group). Figure 4 shows the categorisation results for the Trained Talker tested during the second session. The performance of Same-Day and Overnight groups was plotted against the *Immediate Test: /d/-exposure* group to compare differences in immediate versus delayed testing. We first analyzed the performance of the two groups within this experiment. A mixed-effects model was used with Group (Overnight vs. Same-Day) as fixed effects. By-item and by-subject intercepts were included as random effects. Results indicated no significant difference in categorisation responses between the Same-Day group and the Overnight group for either /d/ words ($\beta = -.11$, $SE = .11$, $p = .31$) or /t/ words ($\beta = -.20$, $SE = .17$, $p = .24$). Thus, despite the fact that there were numerically more /d/ responses across word types for the Overnight group than the Same-Day group, the group difference did not reach significance.

Of interest is whether participants who experience a 12-hour delay between the initial exposure and test (Experiment 2) show comparable talker-specific learning compared to participants who are tested immediately after exposure (Experiment 1). Hence we contrasted the *Same-Day: /d/ exposure* and *Overnight: /d/ exposure* groups from Experiment 2 with the *Immediate Test: /d/-exposure* group in Experiment 1, in order to assess whether the learning effect that was present immediately after exposure (Experiment 1) was still present after a 12-h interval (Experiment 2). A similar mixed-effects model was fitted; treatment coding was used for the Group variable and the *Immediate Test: /d/-exposure* group served as the reference level. Results for /d/ words indicated that the Overnight group had comparable /d/ responses to the *Immediate Test: /d/-exposure* group ($\beta = -.14$, $SE = .24$, $p = .57$), and so did the Same-Day group ($\beta = -.37$, $SE = .25$, $p = .13$). The results for the untrained category /t/ were similar, with neither the Overnight group ($\beta = .49$, $SE = .36$, $p = .18$) nor the Same-Day group ($\beta = .09$, $SE = .37$, $p = .80$) differing significantly from the *Immediate Test: /d/-exposure* group. These results suggest that talker-specific adaptation is maintained similarly whether or not the twelve-hour interval contains sleep.

Talker generalisation

Figure 5 presents the categorisation results as a function of Group and Word Type over the two test sessions. We first tested whether groups differed immediately after training to rule out any effects of time of day on training by comparing the performance of the two groups at Test 1. As predicted, there was no group effect for /d/ words ($\beta = .08$, $SE = .14$, $p = .54$) or /t/ words ($\beta = .08$, $SE = .08$, $p = .33$), suggesting that listeners' performance on the Generalisation Talker in the immediate test were not different between the Same-Day and Overnight group, despite the training and test period being scheduled at different times of day.

The primary question of the present study is to investigate whether there was any change in listeners' ability to generalise to a new talker over time, and whether any such change was facilitated by sleep occurring between training and test. Notably, evidence of generalisation should manifest in increased accuracy in identifying the trained category /d/ (i.e. more /d/ responses for /d/-final words) for a novel talker. We addressed this question by examining within-subject changes in performance across time in the Same-Day and Overnight groups for the Generalisation Talker and by comparing the pattern across groups. Two mixed-effects models were used for /d/ and /t/ categories separately, with Group (Same-Day = 1 vs. Overnight = -1) and Test

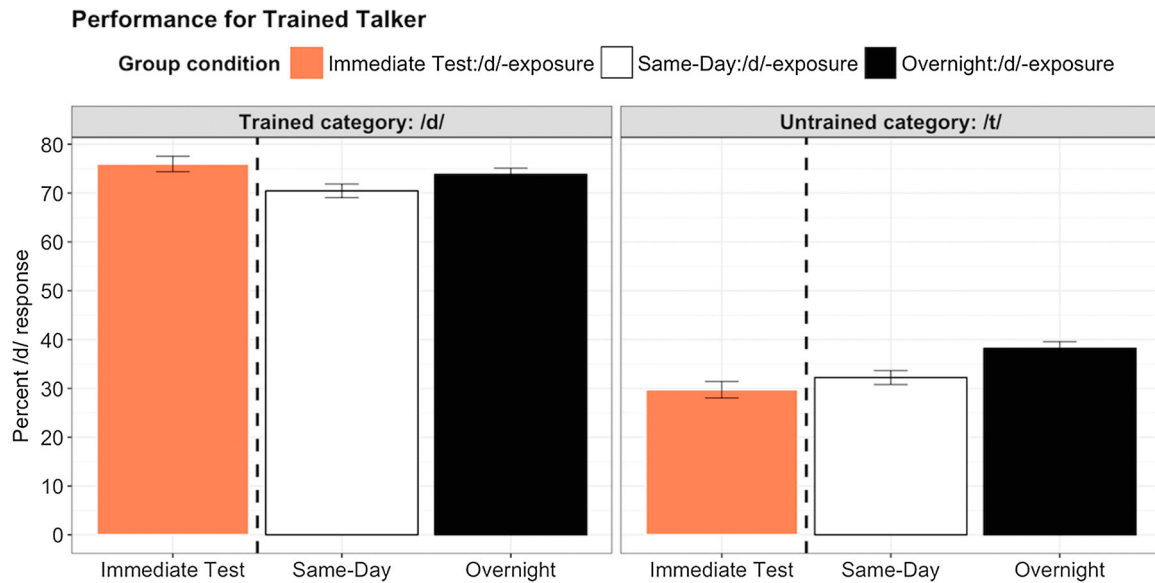


Figure 4. Experiment 2 results for the Trained Talker as a function of Group (Same-Day vs. Overnight) for the two word types, tested after a 12-h delay from the initial exposure. Results are plotted against the *Immediate Test: /d/-exposure* group participants from Experiment 1. Error bars represent standard errors of the mean.

Session (Test 1 = 1 vs. Test 2 = -1) and their interactions as the independent variables in each model. Random effects included by-subject intercepts and slopes for Test Session and by-item intercepts and slopes for Group. Results for the trained category /d/ revealed a critical interaction between Group and Test Session ($\beta = .12$, $SE = .05$, $p = .01$). Neither main effect reached statistical significance at the .05 level ($ps > .85$). Additional analyses on the two groups separately revealed no effect of Test Session among the Same-Day participants ($\beta = .11$, $SE = .07$, $p = .14$). However, the Overnight group showed significantly more /d/

responses at Test 2 than at Test 1 ($\beta = -.12$, $SE = .06$, $p = .05$), consistent with the prediction that sleep promoted the perceptual accuracy for the trained category in this group.

Results for the untrained category /t/ showed a main effect of Test Session ($\beta = .19$, $SE = .07$, $p < .01$), driven by overall fewer /d/ responses at Test 2 relative to Test 1 across the two groups. No other effects reached statistical significance at the .05 level ($ps > .19$). Although the Group-by-Test Session interaction was not significant, a visual inspection of Figure 5 indicates that the main effect of Test Session was primarily driven by the

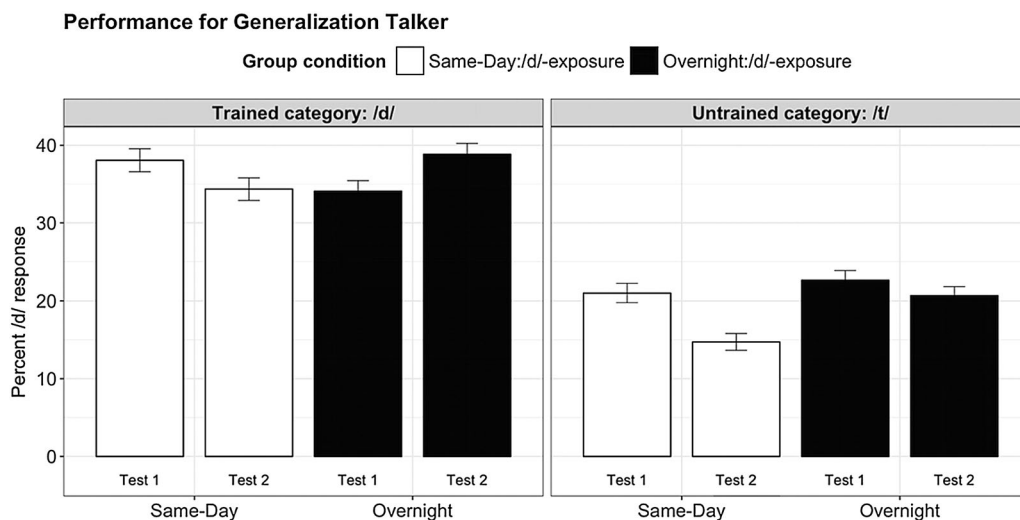


Figure 5. Experiment 2 results for the Generalisation Talker as a function of Group (Same-Day vs. Overnight) for the two word types at two test sessions: immediately after exposure (Test 1) and 12-h after exposure (Test 2). Error bars represent standard errors of the mean.

Same-Day group, who showed an overall increased bias in /t/ reports over time.

Discussion

In the current study, we tested whether listeners can maintain the information they had learned about one Mandarin-accented talker over a twelve-hour delay, and further, whether sleep occurring within that twelve-hour interval facilitates generalisation of talker-specific information to a new Mandarin-accented talker. First, we turn to the question of whether perceptual learning of a foreign accent is retained over time. Notably, a previous study of perceptual learning in the context of a single (native) talker's idiosyncratic speech has shown robust maintenance of such effects (Eisner & McQueen, 2006), regardless of whether sleep was involved in the retention period. Witteman et al. (2015) reported retention of learning effect up to one week from the initial training. Yet to our knowledge, no prior study has directly addressed whether and how offline consolidation affects the learning and maintenance of talker-dependent properties characteristic of a foreign accent. We were motivated in this investigation by reports that sleep-mediated memory consolidation is relevant for broad perceptual reorganisation (e.g. Fenn et al., 2003), and in establishing non-native speech categories (Earle & Myers, 2015a, 2015b).

Effects of time delay and sleep on talker-specific adaptation

Learning the acoustic properties of a foreign-accented talker requires encoding of this information, but also maintenance of these acoustic distributions over time. The current results suggest that the maintenance of talker-specific adaptation during a 12-h delay was stable over time, and was not affected by the activity that participants experienced (sleep or wake state) during the delay. Specifically, both the Same-Day group and the Overnight group had comparable performance to /d/-exposure group participants in Experiment 1, who were tested immediately after the initial exposure to the accented trained talker. Our results extended findings from Eisner and McQueen (2006) to a novel accent by showing no significant decline over time and no effect of experience during a 12-hour delay (wake state or sleep state) on the retention of talker-specific learning of accent information.

Effects of time delay and sleep on talker generalisation

Previous studies suggest that one effect of sleep-mediated consolidation may be the ability to generalise

from specific instances to the more abstract representations. Abstract representations reflect knowledge of learned materials at a level higher than the veridical encoding of trained stimuli, for instance, a statistical pattern that governs the sequences of auditory tones instead of specific tone sequences (Durrant et al., 2011, 2013) or a sound category instead of specific tokens (Earle & Myers, 2015a). This motivates the prediction that sleep may also facilitate the transfer of adaptation to accent-specific acoustic properties from one talker to a new talker with the same accent. Our results on talker generalisation across time and sleep revealed an interesting pattern. We first ruled out diurnal effects on perceiving the Generalisation Talker before the consolidation period, showing that the Same-Day and Overnight groups demonstrated equivalent performance on the Generalisation Talker when tested immediately after training. Differences between the Same-Day and Overnight groups only emerged following a 12-hour delay.

The Overnight participants showed significant improvement in their categorisation of /d/ tokens into the intended category after 12 h that included a night's sleep. In contrast, the Same-Day participants who had a waking interval of 12 h after the initial exposure were more likely to report /t/s when hearing both /d/- and /t/- words at Test 2 (relative to Test 1), if anything, showing unlearning rather than maintenance. While the cause of this change in response bias is not presently clear, it is possible that it reflects a contrastive effect: having contact with other native-accented speakers during the day who produce clear /d/ tokens has made the /d/ tokens produced by the Generalisation Talker even more /t/-like. In this case, the contrast between Same-Day and Overnight participants could be partially attributed to the presence of interference or not, with the interference coming from native-accented speech tokens that do not share critical properties of Mandarin-accented speech. Notably, the bias was observed for the Generalisation Talker, but not for the Trained Talker. We return to this point in the General Discussion and suggest that this is consistent with the *Complementary Learning Systems* account (McClelland et al., 1995; O'Reilly, Bhattacharyya, Howard, & Ketz, 2014).

In sum, transfer of learning from the Trained Talker to the Generalisation Talker did not develop automatically over a passage of time. Sleep appeared to be critical for participants to gain perceptual benefits from a previous exposure. This finding is novel, as, to our knowledge, no past work has specifically investigated how talker generalisation effects develop over time after listeners have adapted to an unfamiliar speaker. We now discuss the interpretations of our results regarding the

retention of talker-specific adaptation and more importantly, the generalisation learning across talkers in connection with the literature on the role of sleep on learning and memory in general.

General discussion

Our work joins an emerging literature on the role of sleep in the perceptual learning of speech and, to our knowledge, presents the first study on the effects of consolidation on generalisation of talker accent adaptation. Importantly, we found that while adult listeners rapidly adapt to non-canonical speech following a brief exposure, such as that produced by a foreign-accented speaker (e.g. Xie et al., 2017), perceptual learning is a dynamic process that continues to evolve in the post-exposure period. On the one hand, talker-specific learning is largely stable over a period of twelve hours; on the other hand, listeners' ability to generalise beyond the trained talker to a novel talker appears to be modulated by sleep-mediated consolidation processes. Compared to a waking interval, sleep promoted generalisation of learning effects to a novel talker. The differential cognitive benefits of memory consolidation on the trained talker versus a generalisation talker reveals critical insight about the underlying role of memory processing on learning.

Theoretically, these observations are consistent with the predictions of *Complementary Learning Systems* account (Kumaran & McClelland, 2012; McClelland et al., 1995; O'Reilly et al., 2014). The CLS account postulates two learning systems that are anatomically and functionally distinct: the hippocampus is argued to support strictly similarity-based generalisation which can occur *immediately* after learning, whereas neocortically-based generalisation emerges after a *delayed* period and is facilitated by memory consolidation that primarily occurs during sleep. As a general theoretical framework of learning, CLS has been applied to several aspects of cognition and language learning, achieving a particular success in predicting a facilitative role of memory consolidation in generalised learning (e.g. Davis & Gaskell, 2009; Tamminen, Davis, & Rastle, 2015). While there is limited evidence that consolidation supports generalisation of phonetic learning (e.g. Earle & Myers, 2015a, 2015b; Fenn et al., 2003), how the dual memory systems, as proposed by CLS, support speech sound learning remains significantly understudied. Below we illustrate how the CLS framework can conceptually accommodate the adaptation and generalisation phenomenon shown in the present study. Bringing together other existing empirical evidence in talker accent adaptation, we argue that there are two

forms of generalisation (immediate vs. consolidation-dependent) in the perceptual learning of talker-related properties that can be separately supported by the hippocampal vs. neocortical learning systems. We conclude by suggesting the implications for perspectives on talker adaptation, and in particular, the long-term maintenance and generalisation of perceptual adjustments made to accommodate non-standard speech.

Immediate generalisation in talker adaptation

Immediate generalisation is manifested in talker-specific adaptation. For the Trained Talker, listeners were only exposed to a limited number of speech tokens and yet were able to generalise to untrained words containing the trained category. Similar findings of rapid generalisation within the lexicon for a trained talker are widely reported (e.g. Eisner et al., 2012; McQueen et al., 2006; Xie et al., 2017). Such efficient generalisation is compatible with a hippocampus-mediated learning system. According to CLS, the hippocampus encodes perceptual items (e.g. a word "seed" produced by a particular talker) in two layers: as individual episodic representations and simultaneously as recurring componential features (e.g. a vowel length of 100 ms and a burst length of 60 ms). It supports generalisation via recurrent activation of multiple episodic traces (determined by their similarity to the current input) and related features (see Kumaran & McClelland, 2012 for details). Thus, to the extent that training exemplars of the sound category /d/ produced by the Trained talker are sufficiently similar in terms of the combinatorial features to the /d/ sounds in the test words (e.g. all drawn from a distribution that is centered around a vowel of 100 ms and a burst of 60 ms), it is expected that exposure-induced learning can generalise to the test phase, demonstrating talker-specific adaptation.

As described in the Introduction, there is evidence showing that phonetic retuning can sometimes generalise to a different talker whom listeners do not have direct experience with, although such generalisation appears to be modulated by the phonetic similarity between the talkers (Reinisch & Holt, 2014; Xie & Myers, 2017). Consistent with hippocampal learning, such similarity-graded generalisation across talkers occurs without any consolidation processes. A weak version of this generalisation is also evident in Experiment 1. In the Introduction, we provided the rationale for selecting these two particular talkers: speech from the two talkers to have shared structures at a higher-level (i.e. they both used burst length as the most informative cue to the voicing contrast /d/-/t/), although their productions exhibit distinctive acoustic distributions. That said, it could be

that some individual exemplars produced by the Generalisation Talker were similar to those produced by the Trained Talker at the acoustic feature level, and this token-based similarity resulted in the very modest transfer of learning to the Generalisation Talker in the immediate test phase.

In addition, we found talker-specific retention over a twelve-hour period, whether or not sleep occurred during the delay (see also Eisner & McQueen, 2006; Wittenman et al., 2015). Empirically, our results extend work on phonetic adaptation to idiosyncratic productions in one's native accent that showed equivalent maintenance of adaptation (Eisner & McQueen, 2006). It should be noted that there are both quantitative and qualitative differences between foreign-accented speech used in the current study and the native-accented speech used in past research. Quantitatively, phonetic deviations in foreign-accented speech are likely to be more drastic than those tested in idiosyncratic native-language adaptation (Eisner & McQueen, 2006; Kraljic & Samuel, 2005). Although we examined only one specific deviation, deviations from "typical" pronunciations will be found not only on the target phoneme, but pervasively across other phonetic contrasts. Qualitatively, the Mandarin-accented speech uses different acoustic-phonetic dimensions to mark the voicing contrast from those used in the native English accent (as shown in Figure 2). In this regard, our results are reminiscent of a pattern observed in performance on a trained talker following categorisation training on a nonnative contrast (Earle & Myers, 2015a). Although the amount of training required for second language category learning is noticeably more extensive than the brief exposure applied in the current study, talker-specific learning is nevertheless retained over time in both scenarios.

Conceivably, this talker-specific retention reflects a maintenance of episodic memory traces in the hippocampus. However, since the hippocampus is limited in capacity, the encoded items would be ultimately overwhelmed by new perceptual experiences (for instance, new instances of /d/s and /t/s spoken by talkers in the listening environment). If our application of the CLS account to the problem of talker adaptation is correct, then without additional learning opportunities, talker-specific adaptation effects (as well as any immediate generalisation that is based on veridical similarity between a trained talker and a generalisation talker) would be susceptible to loss when episodic memory of learned items is no longer strong. While future experiments are needed to evaluate this possibility, results from Experiment 2 reveal a second mechanism that may substantiate generalisation for the long term.

Consolidation-dependent generalisation in talker adaptation

We observed clear sleep effects on talker generalisation: task performance on an unfamiliar talker improved following a 12-h interval containing sleep (Overnight group), while performance deteriorated following a comparable period of wake state (Same-Day group). There is a resemblance between the current results on foreign-accent adaptation and previous work by Fenn et al. (2003) on adjusting to synthetic speech: specifically, the Same-Day group showed a response pattern that is opposite to the effect of perceptual recalibration of phonetic boundaries, consistent with the pattern observed in Fenn et al. (2003). This pattern is also similar to that observed in discrimination performance following categorisation training on a nonnative contrast, in which an Evening-trained group improved, while a Morning-trained group deteriorated, in performance over 24 h (Earle & Myers, 2015b).

How does memory consolidation support generalisation across talkers? According to CLS, the neocortical system complements hippocampal learning by storing overlapping representations of perceptual items and thereby providing abstract knowledge for the long run. The neocortex learns by slowly adjusting connection weights among neurons, allowing a very small impact of a single item on these adjustments. This slow adjustment of neural connections ensures that the resulting cortical representations capture consistent patterns shared by an ensemble of learned materials, rather than item-specific characteristics (McClelland et al., 1995). As such, the neocortex can represent environmental inputs in abstract forms and thus support generalisation that is not dependent on the veridical similarity among stimulus, but on higher-order shared structures.

As noted above, the Trained Talker and the Generalisation Talker both deviated from native-accented talkers in that they used burst length as a marker for stop voicing rather than vowel length. We thus hypothesised that if listeners recognise the shared use of the same acoustic dimension, then learning of the trained talker may allow them to generalise beyond the specific tokens (or the talker). Crucially, this type of cross-talkers generalisation requires listeners to discover a common pattern in the production of the two talkers despite that their speech tokens were not veridically similar overall. Informed by the memory literature (see Diekelmann & Born, 2010 for a review), we predicted that sleep-dependent offline consolidation is critical for the passing of information from the hippocampus to the neocortex and therefore facilitates the formation of a more abstracted representation, or a "generalised

schema", for the Mandarin-accented talkers. The behavioural performance of the Overnight group was consistent with this prediction.

In contrast, the Same-Day group exhibited a different pattern that is compatible with hippocampal learning. After a day spent in the environment where the ambient language was dominantly native-accented English, the Same-Day participants appeared to have slid back into their original native-like "template", reporting the Generalisation Talker's tokens as even more /t/-like than they did at the first test session. Notably, similar interference effects were not observed for the trained talker in that performance was retained to the same level after a 12-h delay. It is plausible that the weak generalisation to the Generalisation Talker in the immediate test, which is contingent on veridical similarity to a subset of exemplars from the Trained Talker, is more vulnerable to interference from other native-accented speakers and thus appears to be less robust than phonetic adjustments for the Trained Talker.

To summarise, sleep may facilitate perceptual encounters with unfamiliar talkers in two ways: in storing salient acoustic-phonetic features abstracted away from the training experience, and in protecting listeners from "interference" coming from the speech communication outside the laboratory. A testable hypothesis for future work is that such interference could result from hearing any type of speech with distinctive acoustic regularities for the trained category than Mandarin-accented English has. If this is the case, then adding such interference before the sleep period of the Overnight group may similarly undermine the weak generalisation effect observed in the immediate test. In addition, future research will be necessary to determine the conditions under which sleep is or is not facilitatory. What remains to be explored is whether sleep has a specific role in generalisation *per se* (that is, consistent with the proposed role for sleep in abstracting away from specific instances to more general schemes), or instead whether the more relevant dimension is the resistance to external interference (from other speech input) on the learned representations.

The present study extends existing work on talker accent adaptation by examining perceptual generalisation beyond a single learning session. We suggest that the CLS account could be a powerful tool that account for different types of generalisation in a unifying account. Taken together, the striking parallels between the role of sleep in adapting to non-native acoustic variants in the context of accented English and the role of sleep in acquiring new non-native contrasts highlight the importance of considering offline consolidation processes and long-term memory effects in the investigation of speech

perceptual learning outcomes. We believe it is an important step towards a comprehensive understanding of how the speech perceptual system maintains a balance between flexibility and stability in the service of robust speech communication.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by NIH [grant R01 DC013064] to EBM. The content is the responsibility of the authors and does not necessarily represent official views of the National Institute on Deafness and Other Communication Disorders [NIDCD].

ORCID

Xin Xie  <http://orcid.org/0000-0002-7021-5053>

References

- Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, 113(1), 544–552. doi:10.1121/1.1528172
- 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 <https://CRAN.R-project.org/package=lme4>
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, 116, 3647–3658. doi:10.1121/1.1815131
- 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 B: Biological Sciences*, 364(1536), 3773–3800. doi:10.1098/rstb.2009.0111
- Dielkmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11, 114–126. doi:10.1038/nrn2762
- Durrant, S. J., Cairney, S. A., & Lewis, P. A. (2013). Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cerebral Cortex*, 23, 2467–2478. doi:10.1093/cercor/bhs244
- Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, 49, 1322–1331. doi:10.1016/j.neuropsychologia.2011.02.015
- Earle, F. S., & Myers, E. B. (2015a). Overnight consolidation promotes generalization across talkers in the identification of nonnative speech sounds. *The Journal of the Acoustical Society of America*, 137, EL91–EL97. doi:10.1121/1.4903918
- Earle, F. S., & Myers, E. B. (2015b). Sleep and native language interference affect non-native speech sound learning. *Journal of Experimental Psychology: Human Perception and Performance*, 41, 1680–1695. doi:10.1037/xhp0000113

- Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, 119, 1950–1953. doi: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. doi:10.3389/fpsyg.2013.00148
- Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, 425, 614–616. doi:10.1038/nature01951
- Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, 116(4), 752. doi:10.1037/a0017196
- Friedrich, M., Wilhelm, I., Born, J., & Friederici, A. D. (2015). Generalization of word meanings during infant sleep. *Nature Communications*, 6, 1–9. doi:10.1038/ncomms7004
- Hupbach, A., Gomez, R. L., Bootzin, R. R., & Nadel, L. (2009). Nap-dependent learning in infants. *Developmental Science*, 12, 1007–1012. doi:10.1111/j.1467-7687.2009.00837.x
- 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(2), 148. doi:10.1037/a0038695
- Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51, 141–178. doi: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. doi:10.3758/BF03193841
- Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1), 1–15. doi:10.1016/j.jml.2006.07.010
- Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological review*, 119(3), 573. doi:10.1037/a0028681
- 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. doi:10.1037/0033-295X.102.3.419
- McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30, 1113–1126. doi:10.1037/0033-295X.102.3.419
- Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America*, 109(3), 1181–1196. doi:10.1121/1.1348009
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204–238. doi:10.1016/S0010-0285(03)00006-9
- O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary learning systems. *Cognitive Science*, 38(6), 1229–1248. doi:10.1111/j.1551-6709.2011.01214.x
- 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. doi:10.1037/a0034409
- Stickgold, R., & Walker, M. P. (2013). Sleep-dependent memory triage: Evolving generalization through selective processing. *Nature Neuroscience*, 16, 139–145. doi:10.1038/nn.3303
- Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, 119, 131–136. doi:10.1016/j.cognition.2010.10.018
- Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive psychology*, 79, 1–39. doi:10.1016/j.cogpsych.2015.03.003
- Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, 27, 979–1001. doi:10.1080/01690965.2011.597153
- Witteman, M. J., Bardhan, N. P., Weber, A., & McQueen, J. M. (2015). Automaticity and stability of adaptation to a foreign-accented speaker. *Language and Speech*, 58, 168–189. doi:10.1177/0023830914528102
- 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. doi:10.1016/j.wocn.2013.06.003
- Xie, X., & Myers, E. B. (2017). Learning a talker or learning an accent: Acoustic similarity constrains generalisation of foreign accent adaptation to new talkers. *Journal of Memory and Language*, 97, 30–46. doi:10.1016/j.jml.2017.07.005
- Xie, X., Theodore, R. M., & Myers, E. B. (2017). More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 43, 206–217. doi:10.1037/xhp0000285