Pamela Fuhrmeister

Interference in memory consolidation of non-native speech sounds

Abstract: For several decades, researchers have been investigating the challenges of and constraints on learning the speech sound inventory of a second language in adulthood. Commonalities that have emerged from these findings include the immense individual variability reported in non-native speech sound learning studies and the rare attainment of native-like proficiency in perception or production of second language speech sounds. While numerous studies have shed light on various aspects of this challenging process, many questions about the extent and nature of these difficulties remain. A nascent line of research suggests that some of the difficulty in non-native speech sound learning could be attributed to various sources of interference that disrupt the memory consolidation process, thus interfering with the retention of learned phonetic information. It is well documented in the broader learning literature that interference from competing stimuli or subsequently learned skills can disrupt memory consolidation processes. However, this phenomenon has received little attention in the speech literature, and the potential sources of interference in the speech domain have yet to be identified. In this review, I discuss how integrating theories of memory consolidation with non-native speech sound learning models can more accurately capture patterns of learning observed in the non-native speech sound learning literature, specifically patterns showing failures of memory consolidation due to interference.

Keywords: sleep, second language learning, adults, memory consolidation, non-native speech sounds

1. Introduction

Adult second language learners face many challenges, especially when attempting to master the speech sounds of a non-native language. Although a second language learner must gain proficiency in a number of linguistic domains (e.g., morphology, syntax, semantics), acquiring perceptual sensitivity to the speech sounds of another language is an important step in language acquisition. Indeed, several studies support the notion that speech perception abilities can facilitate higher levels of language learning

(e.g., lexical acquisition). For example, native speech perceptual abilities in infancy have been shown to predict language development in early childhood (Tsao et al., 2004; Kuhl et al., 2008), and in adulthood, superior perceptual discrimination of non-native speech contrasts can facilitate learning of lexical items that contain those contrasts (Silbert et al., 2015). Thus, mastery of the speech sound inventory of a language may be a crucial step in language acquisition more generally. Unfortunately, however, most studies of second language learners report rather poor perceptual abilities for difficult phonetic contrasts (e.g., Bradlow et al., 1999; Flege, 2003), even when learning commenced in childhood (Pallier et al., 1997). As such, several types of training paradigms have been devised in an attempt to optimize non-native speech sound learning, specifically for difficult speech sound contrasts. Paradigms that have been employed in training studies differ in terms of whether learning takes place in an implicit (Lim & Holt, 2011; Vlahou et al., 2012; Wade & Holt, 2005) or explicit manner (e.g., Earle et al., 2017; Earle & Myers 2015a, 2015b), or whether participants were trained on tokens with limited variability or high variability (e.g., tokens produced by multiple talkers or occurring in multiple phonological contexts, Logan et al., 1991; Lively et al., 1993, 1994; Bradlow et al., 1997, 1999). Other paradigms utilized either natural speech or exaggerated versions of speech to make differences between stimuli more salient (McCandliss et al., 2002; Golestani & Zatorre, 2004; Swan & Myers, 2013). Although each of these training paradigms has indeed demonstrated learning, the majority of these studies report only moderate success, and most second language learners ultimately fail to attain native-like perception or production of non-native speech sounds (e.g., Bradlow et al., 1999; Piske et al., 2001; MacKay et al., 2001).

Most studies probing plasticity in the speech system focus on the initial processes of *learning* speech sounds; however, language acquisition involves more than the initial encoding of stimuli. In order for an individual to develop and maintain language proficiency, critical aspects of a language, such as phonemes and lexical items, need to be consolidated into long-term memory for later retrieval. Stable long-term memory representations may then facilitate retention and generalization of learned speech sounds. Until recently, the role of memory in speech sound acquisition had been underexplored, and therefore, many questions about the

memory functions underlying this process remain. Due to the general lack of success in non-native speech sound learning, it is becoming more apparent that consistent failures of memory consolidation may underlie the challenge of learning non-native speech sounds, and this merits further exploration. More specifically, training conditions that facilitate stronger initial learning of speech sounds and limit exposure to interference after training and before consolidation takes place may be key factors in promoting long-term changes in the speech system. The current chapter begins with a review of models of non-native speech sound learning, followed by a discussion of memory consolidation theories and experimental evidence for those. In light of this evidence, I conclude with an alternative interpretation of some work on non-native speech sound learning and discuss how considering strength of learning and interference in memory consolidation may have explanatory power for some of the challenges reported in these studies.

2. Constraints on non-native speech sound learning: Perceptual similarity of first language speech sounds

By the first few months of life, infants demonstrate perceptual sensitivity to many different speech sounds that are found in world languages (Eimas et al., 1971). However, by the end of the first year of life, infants lose the ability to discriminate certain phonetic contrasts that do not occur in the ambient language (e.g., Werker & Tees, 1984; Kuhl et al., 2006). While this warping of perceptual space appears to facilitate language acquisition in early childhood (Tsao et al., 2004; Kuhl et al., 2008), the process of perceptual reorganization can greatly constrain the acquisition of non-native speech categories later in life. In particular, perceptual similarity of native and non-native speech sounds can cause non-native speech sounds to be more difficult to perceive as distinct categories (e.g., Best et al., 2001). For example, native English speakers often struggle to perceptually distinguish dental and retroflex voiced stop consonants found in Hindi. The perceptual similarity of these sounds to the English alveolar /d/ category and their similarity to each other make these exceptionally challenging for learners to disambiguate. Indeed, several models of non-native speech sound learning account for such difficulties by considering the relationship

between native and non-native speech sounds. For example, the perceptual assimilation model (e.g., Best, 1994; Best & Tyler, 2007) and the native language magnet model (Kuhl, 1994; Kuhl et al., 2008) both posit that non-native speech sounds that are perceptually similar to native phonemes will be harder to perceive than perceptually dissimilar sounds. However, the two models differ on which dimensions of the speech signal are considered important for perception. The perceptual assimilation model focuses on naive listeners' perception of non-native speech sounds as a result of articulatory similarity between native and non-native speech sounds (Best, 1994; Best & Tyler, 2007). Specifically, this model predicts that naive listeners will assimilate unfamiliar non-native speech sounds to the perceptual category in the native language that is produced by the most similar articulatory gesture (Best & Tyler, 2007). For example, native English speakers often map dental and retroflex voiced stop consonants found in Hindi onto the alveolar /d/ sound found in English. As a result, certain speech sounds are more difficult to perceive, while speech sounds without a similar native language category are perceived more easily (Best et al., 2001). For instance, English contains no speech categories similar to click sounds found in Zulu, and native English speakers typically discriminate these sounds accurately (Best et al., 1988). In contrast to articulatory gestures, the critical dimension of the native language magnet model is acoustic space. This model concentrates on the developmental processes underlying the acquisition of native speech sounds (Kuhl, 1994; Kuhl et al., 2008) and postulates that infants take advantage of statistical learning in order to acquire the speech sound categories of their native language. After sufficient exposure, infants' perceptual space becomes "warped," and prototypes for native language speech categories emerge. These prototypes act as magnets that attract perceptually similar speech sounds. Analogous to the perceptual assimilation model, this magnet effect would result in some non-native speech sounds (i.e., perceptually similar sounds) being more difficult to learn than others (perceptually dissimilar sounds). While similar in some ways to the perceptual assimilation model and the native language magnet model, Flege's (1995) speech learning model takes a slightly different approach. First, this model focuses on experienced adult second language learners and proposes that difficulties in second language speech production stem from perceptual obstacles. In

other words, one cannot produce what one cannot perceive. An additional component of this model predicts that non-native speech sound learning becomes more difficult over the lifespan; however, it deviates from traditional critical period hypotheses, as it predicts a gradual decrease in non-native speech production abilities over the lifespan, rather than an abrupt decline after puberty. While this model emphasizes adult second language learners' speech production abilities, it is similar to the models described above in that it attributes difficulties with the second language sound system to similarities with the native language.

Attention to dimension models add to these models by clarifying the processes required for acquiring non-native speech categories. Attention to dimension models propose that a speaker of a language has learned to direct attention to relevant parts of the acoustic signal for his or her native language and that learning new speech categories requires a learner to attend to previously unattended dimensions of the signal (Francis & Nusbaum, 2002). Specifically, native speakers of a language have learned to direct attention to relevant acoustic cues in the speech signal, and they have simultaneously learned to ignore other cues. Learning to reweight acoustic cues or to direct attention to different parts of the speech signal presents a challenge for learners. This model differs from the others described above in that it explains how a listener's perceptual space changes as a result of experience with the native language (i.e., via selective attention to meaningful features).

While each of these models focuses on distinct phases of the acquisition process (e.g., infants, naive listeners, or experienced second language learners) and accounts for challenges in different ways (e.g., development, articulatory representations, production, or dimensions of the acoustic signal), some similarities emerge from a close comparison of them. For example, these models do not necessarily attribute poor second language perception and production abilities to a critical period or a loss of neural plasticity; rather, difficulties are attributed to prior experience with the first language and how that experience shapes perception and production of new speech categories. Clearly, the native language exerts constraints on perception and production of second language speech sounds, and these constraints are especially strong when speech sounds in the native and non-native languages are close in perceptual space. These models certainly

capture a great deal of the non-native speech sound learning process, especially the relative difficulty of different speech sounds as a result of native language background. However, many non-native speech training studies observe considerable variability even among individuals of the same language background (e.g., Golestani & Zatorre, 2004, 2009; Myers & Swan, 2012; Yi et al., 2014), which suggests that factors beyond the native language constrain this process.

3. Memory consolidation in speech and language learning

To account for individual variability in speech sound learning, it may be helpful to consider how models of memory consolidation predict retention, forgetting, or elaboration of learned perceptual information. Most models of memory consolidation posit that two separate memory systems work in tandem to consolidate different types of learning. Specifically, the fast, hippocampus-mediated and often sleep-dependent system serves to consolidate declarative or explicit memories in order to integrate them into existing networks of knowledge, while the slow, hippocampus-independent system associated with implicit or non-declarative learning induces local changes to neuronal circuitry as a result of continued experience and does not typically rely on sleep (e.g., complementary learning systems, McClelland et al., 1995; McClelland, 1998; see also Marshall & Born, 2007; Dudai, 2004). With extensive exposure over longer periods of time, the slow, procedural learning system is able to discover regularities in the input from the environment. As a result, this memory system is able to sort information into categories, and this knowledge about categories can aid generalization to new contexts or exemplars (see McClelland et al., 1995 for discussion). In connectionist models, rapid sequential learning (i.e., learning one task immediately after another), leads to what is called catastrophic interference (McCloskey & Cohen, 1989). In other words, networks must completely forget or overwrite the information they have learned in order to accommodate new information. According to McClelland (1998), catastrophic interference underscores the need for the rapid, hippocampus-mediated consolidation system: the hippocampus acts as a temporary memory store and allows for memory traces to be selectively consolidated into long-term memory and integrated into existing

networks of information. Crucially, this process of selective consolidation via the hippocampus does not result in information being overwritten.

A rich literature on memory consolidation suggests that sleep can improve performance on a task even in the absence of further practice through a period of off-line consolidation (Stickgold et al., 2000; see Marshall & Born, 2007 for review). Sleep has been shown to facilitate abstraction of information from episodic traces, and this integration of information into existing knowledge networks allows learning to generalize to novel contexts (Davis et al., 2009). Sleep-mediated consolidation has also been shown to help protect newly learned information from interference (e.g., Ellenbogen et al., 2006, 2009; Drosopoulos et al., 2007) or recover learning that decayed throughout the course of a day (Fenn et al., 2003). Several studies have additionally investigated the contributions of sleep-dependent memory consolidation to word learning (e.g., Tamminen & Gaskell, 2013; Dumay & Gaskell, 2007; Davis et al., 2009). Consistent with McClelland's complementary learning systems account, a study by Davis et al. (2009) found that newly learned words become integrated into the existing lexicon after a period of sleep. In this study, participants learned novel words that overlapped with existing lexical items by several phonemes (e.g., *cathedruke-cathedral*), and these novel words only showed evidence of lexical competition with existing words (e.g., *cathedral*) on a lexical decision task after a period of sleep had occurred. Additionally, brain activation measured by functional magnetic resonance imaging (fMRI) in this study revealed activation of the hippocampus in response to the novel lexical items before sleep and cortical activation after sleep. This suggests that the newly learned lexical items were temporarily stored in the hippocampus and became re-represented in cortical areas after sleep, which is in line with predictions from McClelland's complementary learning systems account. In addition to offline gains in the absence of further practice, sleep between two periods of practice in word learning can benefit long-term retention of novel words (Mazza et al., 2016), and even daytime naps or short periods of sleep can improve or stabilize learning of tasks, including word learning (Lahl et al., 2008; Heim et al., 2017). The extant literature shows clear benefits of sleep for word learning; however, of interest to the current chapter is whether similar findings are also observed in speech learning.

Indeed, some studies have found benefits of sleep for speech or auditory learning tasks (see Earle & Myers, 2014 for review). For example, Fenn et al. (2003) carried out an experiment in which participants learned to understand computer-synthesized versions of native language words. In this study, participants who slept between training and test improved in the absence of further practice. Moreover, a degradation of performance was observed for participants who were trained in the morning, but this loss was recovered following sleep. Thus, sleep may help recover performance that has degraded throughout the course of a day, as well as facilitate generalization to novel contexts. In fact, a later study by the same group suggests that sleep promoted generalization of information learned with a large amount of variability, but not when learning took place with a closed set of tokens (Fenn et al., 2013). In similar fashion, Xie et al. (2017) tested listeners on generalization of learning of one Mandarin-accented talker to a novel Mandarin-accented talker. Like the Fenn et al. (2003) study, Xie and colleagues (2017) found that performance on the untrained talker degraded throughout the day for a group trained in the morning. Thus, sleep was necessary for generalization to a novel talker and was not simply observed with the passage of time. Although these studies have found performance gains following sleep, other work on sleep in auditory learning and perceptual learning of speech have not observed any added benefits of sleep. For example, Roth and colleagues (2005) found that sleep was not necessary for improvement on a speech in noise task, but rather, the passage of time was sufficient to induce performance gains. Similarly, Eisner and McQueen (2006) found no additional benefit of sleep for perceptual learning of speech. In a lexically-guided perceptual learning task, participants were exposed to lexical items containing non-canonical productions (an ambiguous fricative between /f/ and /s/) of certain speech sounds embedded in a lexical context that served to disambiguate the speech sound. After exposure, participants were asked to categorize tokens on a non-word continuum from /f/-/s/. In this paradigm, a shift in the category boundary, (i.e., categorization of more ambiguous tokens in a non-word context consistent with the lexical bias in the exposure condition) indicates perceptual learning. Indeed, participants in this study showed a category boundary shift consistent with their exposure condition both immediately after training and 12 hours later. Interestingly,

a group that slept during the 12-hour interval showed no greater learning effect than a group that remained awake during the day throughout the post-training interval.

Several differences could explain the discrepancies in the findings of these studies. First, different learning systems may underlie the various tasks employed in these studies. For example, the Eisner and McQueen (2006) study measured adaptation to episodic representations of talker-specific idiosyncrasies in speech production (that is, how that talker produced a particular speech sound in a non-standard way). Recent evidence suggests that adaptation to talker-specific, episodic information (i.e., details of a specific talker's voice) occurs rapidly and is stable over time, while tasks involving more abstract representations may emerge only after consolidation (Brown & Gaskell, 2014). This could explain why both groups in the study by Eisner and McQueen (2006) remained stable over time, regardless of sleep. The fact that participants in the study by Fenn and colleagues (2003) were trained and tested on completely different words may explain why these participants benefitted from sleep. Abstract representations facilitate generalization to novel contexts (in this case, generalizing knowledge of synthetic speech sounds to new lexical contexts), and because abstract representations manifest only after offline consolidation, improvement after sleep in this context is not surprising. However, studies by Earle and Myers (2015a) and Earle et al. (2017) have found overnight improvement on tasks even when no generalization was needed (i.e., the same tokens were used in training and testing).

Sleep may be especially advantageous when learning involves the formation of new representations, rather than adapting existing representations to accommodate atypical exemplars. For instance, several recent studies have found benefits of sleep in non-native speech sound learning. Earle and Myers (2015b) trained participants to learn the Hindi dental/retroflex contrast on a closed set of tokens but found that sleep enabled generalization to stimuli produced by novel talkers. Earle et al. (2017) even found that duration of sleep predicted overnight improvement on the tasks used to assess non-native speech sound learning. Specifically, they found that the amount of slow wave sleep predicted overnight gains on identification of the speech sounds, while total sleep duration predicted participants' ability to discriminate the speech sounds. Importantly, this study suggests

that individual differences in sleep duration may account for some of the individual variability typically observed in training studies. On the other hand, an additional study by this group found some surprising limits to the benefits of sleep consolidation (Earle & Myers, 2015a). This study likewise trained participants on the Hindi dental/retroflex contrast and observed improvement after sleep, but this advantage only held if participants had been trained in the evening. This study consisted of two groups of participants who were trained on the contrast. One group was trained in the morning hours and one in the evening hours. Each group returned approximately 12 and 24 hours later for reassessment. Surprisingly, only the evening-trained participants showed improvement following an overnight interval. The authors reasoned that this discrepancy could be a result of interference from native language exposure. Specifically, participants trained in the morning had a day's worth of input from their native language prior to sleep, while the evening-trained group presumably had much less. Subsequent experiments in this study indicated that the lack of overnight improvement seen for the morning trained group stemmed from exposure to perceptually similar native language speech sounds.

Other recent studies have similarly found that exposure to certain stimuli or engagement in certain tasks can interfere with learning or consolidation of newly acquired skills or representations. For example, alternating perceptual training with speech production practice has been shown to attenuate or interfere with learning. In a study by Baese-Berk and Samuel (2016), native Spanish-speaking participants were trained to learn a difficult, non-native Basque speech contrast in the laboratory. Training consisted of an ABX discrimination task in which participants heard three sounds and were asked to indicate whether the third was more similar to the first or second sound presented. For this study, some participants were asked to repeat the third sound presented on each trial out loud before indicating their decision. Surprisingly, the group that repeated the trained sound showed very little learning of the contrast on a perceptual post-training assessment. Furthermore, production of any sounds, even sounds that were dissimilar to the trained contrast, seemed to attenuate learning. This suggests that the cause of this interference was not solely a result of exposure to the participants' own poor productions of the non-native contrast. Rather, it raises questions about the mechanisms underlying

speech perception and production in the development of speech sound representations.

Exposure to phonological variability before or after training may additionally attenuate learning and disrupt consolidation. For example, a study by Fuhrmeister and Myers (2017) examined whether native English-speaking participants trained on a non-native, Hindi dental/retroflex contrast would benefit from additional exposure to the contrast in a different vowel context. In this study, one group of participants heard the contrast in only one vowel context in minimal pair non-words (/ḍug/ and /ḍug/) throughout training and testing. Another group heard the contrast in two different vowel contexts in assessments (/ḍug/ and /ḍug/ vs. /ḍig/ and /ḍig/; assessments consisted of a pretest, immediate posttest, and a delayed posttest), but they had identical training to the other group (i.e., /ḍug/ and /ḍug/ only). Notably, the participants who were exposed to the Hindi sounds in two vowel contexts performed significantly worse than the group exposed to one vowel context only on the tasks involving the contrast in the trained vowel context, despite having more total exposure to the contrast. Additionally, participants exposed to the contrast in two vowel contexts showed no evidence of overnight improvement on the stimuli in the trained vowel context, while those who heard the sounds in only one vowel context did improve after an overnight interval. These findings suggest that exposure to novel speech sounds in different vowel contexts may interfere with learning or consolidation of the contrast in a trained vowel context, even if that extra exposure is limited (i.e., at test only). It is also possible that the learning of the trained vowel context was less stable for participants exposed to two different vowel contexts, which may have prevented further improvement as a result of sleep. As can be seen, memory consolidation influences speech learning in the following ways:

• Sleep helps consolidate newly formed representations of both natural and synthetic speech sounds as indicated by performance improvement (Earle & Myers, 2015a; Earle et al., 2017) and generalization (Earle & Myers, 2015b; Fenn et al., 2003, 2013).

• Not all types of perceptual speech learning tasks show improvement after sleep (Roth et al., 2005; Eisner & McQueen, 2006), and this may depend on task difficulty, whether new representations are being

formed, or whether existing representations are being expanded to accommodate new exemplars.

- Exposure to certain stimuli (e.g., perceptually similar native language speech sounds) following training may interfere with learning of novel speech sounds (Earle & Myers, 2015a).
- Training conditions (e.g., exposure to phonological variability, Fuhrmeister & Myers, 2017; production of speech sounds or words, Baese-Berk & Samuel, 2016) may destabilize or attenuate learning, which may affect the consolidation process.

Although sleep appears to facilitate memory consolidation of speech in a variety of ways, many questions remain. For example, it is unclear what types of stimuli might interfere with non-native speech sound learning or under what conditions interference effects could be avoided. However, it may be possible to carry over insights from other domains in order to inform these questions and make predictions about speech learning.

4. Failures of consolidation: Interference effects in learning

In order to fully understand how new memories are formed, it is important to examine cases in which consolidation fails. Over a century ago, Müller and Pilzecker (1900) proposed that memories exist in an initially labile state, in which they are subject to interference from subsequently learned tasks. In their studies, they tested explicit recall of strings of unrelated digits and observed that their participants were not able to recall one list as well when tested 24 hours later if they had learned a subsequent list immediately following practice on the first list. Walker's (2005) model for procedural memory consolidation similarly assumes that newly acquired memory traces are fragile and must undergo a process of stabilization before becoming resistant to interference. This model is comprised of three main stages: acquisition, consolidation-based stabilization, and consolidation-based enhancement. Walker (2005) argues that the stabilization stage depends on the passage of time only, while the enhancement stage relies on sleep. Two dissociable systems have been proposed in the memory literature, which presents a challenge for extending theories of memory consolidation to other domains. The declarative memory system underlies explicit learning of facts or episodes (sometimes referred to as the

memory for "what"), while the procedural memory system serves memory of implicitly acquired actions or procedures (the memory of "how") (e.g., Squire, 2004). Although Walker's model was originally intended for procedural memory and Müller and Pilzecker's account for declarative memory (though their account predates this term), some findings suggest procedural and declarative memory systems may not be as dissociable as once thought (Poldrack et al., 2001). In addition, studies including both declarative and non-declarative tasks lend support to the notion that the consolidation of a newly acquired skill or memory can be disrupted if an interfering task or stimulus is introduced before the memory has stabilized. Furthermore, it remains unclear whether speech category formation can be neatly classified as either procedural or declarative learning, and this process may be different at different points throughout the lifespan or under different learning conditions. Therefore, for the remainder of this review, I will draw on the literature of learning and memory processes for both declarative and procedural tasks and will reflect on the importance of a stabilization period following initial encoding in order to help new memory traces become resistant to interference. This section reviews a series of studies that have examined interference effects in several domains of learning, including the time course required for stabilization and consolidation of memory traces and the strength of initial learning or encoding. The goal of drawing on this literature is to make predictions about how speech sound learning may be facilitated by mitigating interference or adhering to a training paradigm or schedule that is more conducive to consolidation and long-term retention.

A seminal study in the motor learning domain demonstrated interference with a behavioural task, in which participants learned to move a two-hinged handle to a target while compensating for perturbation (Brashers-Krug et al., 1996). Participants who learned to compensate for perturbation experienced a disruption of consolidation if, immediately following training, they were trained to compensate for perturbation in the opposite direction. Another group of participants completed identical training on the first task, but their second task consisted of moving the handle to a target in the absence of any perturbation. These participants showed no interference effect from the second task. An additional group completed the two training sessions with perturbation in opposing

directions but waited four hours between the two training episodes. This group also showed performance improvement on the first task, indicating that a period of four hours was sufficient to stabilize learning of the first task, making it immune to disruption from a second task. Similarly, a study by Walker et al. (2003) demonstrated interference effects in a finger tapping sequence task unless six hours had passed between the two training sessions. In addition to behavioural tasks, the application of transcranial magnetic stimulation (TMS) as a source of interference has been tested (Muellbacher et al., 2002). In this motor learning study, participants practiced a finger movement sequence and were assessed on their improvement in speed. When TMS was administered immediately following training, participants showed no retention of the behavioural gains observed during the learning phase. However, if a period of six hours had lapsed before TMS was applied, no interference was observed. These studies support the consolidation hypothesis and Walker's consolidation model by demonstrating that the passage of time is necessary to stabilize newly encoded motor memories. Furthermore, the type of task that follows learning may dictate whether learning on the first task is disrupted. These findings may be able to make important predictions in the speech domain. For example, if learners train on non-native speech sounds and are exposed to speech sounds in their native language before the stabilization period has concluded, consolidation of the non-native sounds may be obstructed. Similarly, it may also be the case in the speech domain that not all stimuli or tasks interfere equally. If that is indeed the case, it will be important to identify which types of stimuli or tasks (e.g., native language exposure, Earle & Myers, 2015a; speech production, Baese-Berk & Samuel, 2016) are able to interfere with speech sound learning.

Similar task and timing effects to those found in motor learning have been observed in visual perceptual learning tasks. For example, in a visual hyperacuity task, participants saw two presentations of three dots arranged vertically on a screen and were asked to indicate whether the middle dot was offset in the either the first or second group of dots presented (Seitz et al., 2005). Following training, participants completed training on another task: in one task, the presentation of the dots was the same except the offset was presented in the opposite direction, and other tasks varied the spatial location and the orientation of the dots (i.e., the

dots were presented horizontally). Crucially, only the participants who were trained with the opposite offset direction experienced interference. Visual perceptual learning of stimuli that were presented at different spatial locations or in different orientations did not interfere with initial encoding of the task, which the authors attributed to the retinotopic specificity of spatial location and orientation. Another critical finding in this paper was that participants who waited one hour before training on the opposite direction did not demonstrate any attenuation of learning on the first direction. An additional visual perceptual learning study using a line orientation detection task found a period of 3.5 hours to be sufficient to eliminate retrograde interference from a second visual task (Shibata et al., 2017). These findings from visual learning studies provide further support that fragile memory traces remain susceptible to interference until a period of stabilization has passed. Like the motor learning study by Brashers-Krug et al. (1996), the study by Seitz et al. (2005) shows that not all tasks have the potential to interfere with consolidation of a previously learned task. Evidence from these two domains, namely vision and motor learning, suggests that domain-general processes may underlie consolidation of learning and may therefore be applicable to the speech domain.

In further support of the consolidation hypothesis, one study investigating interference from consecutive tasks in patients with amnesia found surprisingly similar stabilization effects, despite the fact that declarative memory consolidation deficits are a hallmark of amnesia. Dewar et al. (2009) had individuals with amnesia learn word lists, and these participants showed a graded advantage in recall after the presentation of interfering stimuli at different time points. Participants experienced a delay between the initial learning session and the presentation of interfering stimuli, and longer delays facilitated recall of the original word lists more effectively than shorter delays. This suggests that even individuals with amnesia who have declarative memory consolidation deficits can benefit from a stabilization period following learning.

If non-native speech sound learning processes parallel those of visual, motor, and word learning, the stabilization phase prior to consolidation may be crucial to learning and retention of non-native speech sounds. If the stabilization phase is disrupted from exposure to conflicting stimuli or

practice on an interfering task, this may hinder consolidation and retention of novel speech sounds.

Although many studies have found robust support for the consolidation hypothesis, results from other studies challenge its reliability. For example, Goedert and Willingham (2002) trained participants on two implicit motor tasks with the goal of testing whether these memories undergo consolidation and become resistant to interference from learning a similar task. The researchers first utilized a serial reaction time task in this study, in which participants saw a sequence of circles appear in boxes on a screen and pressed a button corresponding to each box after a circle was presented. In this paradigm, participants are unaware that the sequence is not random but consists of an underlying pattern; therefore, learning is implicit. The second task used in this study was a task in which participants learned a new visuomotor mapping. Participants were instructed to point at a target on a screen while wearing prism glasses that displaced their vision. Training for each task followed a traditional interference paradigm (train on task A, train on task B, test on task A), and participants were trained on different sequences and visual displacements for task B at varying intervals following training on task A. Unlike several previous studies, this study did not find evidence that the motor memories had been consolidated and become resistant to interference, as even 24 hours was not sufficient to protect against interference from task B. A study employing similar visuomotor tasks by Caithness et al. (2004) additionally found that memories for one task were susceptible to interference from another task even 24 hours later. Walker et al. (2003) demonstrated similar effects in a finger tapping task. Interestingly, participants who trained on task A, waited 24 hours, and performed task A again before learning task B did not retain their learning of task A. Walker and colleagues (2003) posited that reactivation of consolidated memories can shift them into labile states, causing them to become susceptible to interference once again. Caithness et al. (2004) speculated that performance on task B in their study may have been sufficient to reactivate memory traces of task A, which allowed task B to interfere with task A. Goedert and Willingham (2002) largely attributed this ostensible lack of consolidation to task differences or neural structures underlying the specific task used in their study.

An open question is what these memories require in order to be transferred into a stable state and resist subsequent interference. It is possible that certain tasks or types of learning undergo a different consolidation process than others or are not consolidated at all. It appears that non-native speech sound learning can indeed undergo consolidation as evidenced by improvement in the absence of further practice (Earle & Myers, 2015a, 2015b; Earle et al., 2017), and it may be the case that a stabilization or consolidation period would protect newly formed phonetic category representations from interference.

5. Stability and strength of learning

The studies reviewed in the last section, which encompass both procedural and declarative tasks, provide important evidence for consolidation theories proposed by Müller and Pilzecker (1900) and Walker (2005): most newly acquired memory traces need to undergo a period of stabilization in order to become resistant to interference. In some cases, the presence of interfering stimuli during the stabilization phase may be strong enough to disrupt the consolidation process entirely. An important question to address in speech learning studies will be how to minimize interference during the stabilization phase or to identify training conditions in which information may be consolidated in spite of interference.

In addition to the passage of time during a stabilization phase, strength and stability of initial learning may be an important factor in determining whether information is consolidated. Ebbinghaus (1885) first proposed that increasing the repetition of practice trials in a task may lead to better retention of the information 24 hours later. In addition to the early findings by Ebbinghaus, several recent studies in the visual domain lend support to this idea. For example, a study by Hauptmann et al. (2005) found that participants who practiced a visual task until performance reached asymptote improved after a period of sleep, while those who did not practice to this criterion failed to show overnight improvement. Tucker and Fishbein (2008) trained participants on a series of declarative memory tasks and had some take a nap following training. Interestingly, only the high-performing participants in training benefitted from sleep, suggesting that stronger learning can facilitate overnight improvement. In a study by

Shibata and colleagues (2017), participants who overlearned (continued to practice after the point of mastery) a visual perceptual task did not experience interference from a second task, suggesting that hyper-stable learning can accelerate or even obviate the need for a stabilization phase following learning. Taken together, these studies suggest that benefits of consolidation may depend on how strongly information is initially learned.

Conversely, a few studies have found that sleep preferentially enhances recall of weakly learned information or performance on more difficult tasks. For example, Drosopoulos and colleagues (2007) had participants memorize word pairings and manipulated how strongly the pairings were learned. Of the participants who only weakly learned the pairings, participants who slept forgot significantly fewer word pairings when tested two days later as compared to a wake group. However, sleep and wake groups were comparable if the information was strongly learned during training. Although sleep did seem to benefit the group that did not learn the information as rigorously to begin with, these findings also support the benefits of strong initial encoding. Even though the wake group did not sleep immediately after learning, they performed equivalently to the sleep group. In addition, it seems difficult to rule out ceiling effects in this study, as the participants in the strong encoding group performed at over 95 % accuracy. Similarly, a study using a procedural motor learning task found superior benefits of sleep for the most difficult task during training, as measured by an increase in speed (Kuriyama et al., 2004). Analogous to the Drosopoulos et al. (2007) paper, the participants who learned the easier tasks still outperformed the group that learned the more difficult task, although the benefits from sleep were not as drastic.

As shown above, the mixed evidence presented here implies a complicated relationship between strength of learning and consolidation. Some studies suggest that strongly encoded information is advantageous for consolidation, while others show stronger sleep-related benefits for weakly learned information. Critically, in the studies showing benefits of sleep consolidation for weakly learned information, participants who trained on easier tasks or trained on the same tasks to a higher criterion demonstrated superior overall performance, which should be considered along with the superior benefits of sleep for weakly learned information. Additionally, ceiling effects arguably cannot be completely ruled out in these studies. It

is also possible that the benefits of sleep are observed in a u-shaped trajectory because the qualitative changes associated with sleep-mediated consolidation do not always manifest behaviourally. For example, newly learned information may need to reach a minimum level of stability in order to trigger consolidation, and sleep may show the strongest influence on these memories as far as behavioural changes can be observed. However, sleep has been shown to induce qualitative changes to memories, such as the ability to generalize to new contexts (e.g., Fenn et al., 2013), increased automaticity of a task as measured by electrophysiological components (Atienza et al., 2004), and differential functional activation patterns in response to stimuli after sleep (Davis et al., 2009). For example, the study by Atienza et al. (2004) trained participants to discriminate auditory tone patterns. Some participants slept after training, while another group was sleep deprived. They found improved behavioural performance for both groups, regardless of sleep; however, an electrophysiological component that responds to the involuntary switching of attention was elicited only in the participants who slept after training. Using fMRI, Davis et al. (2009) found changes in functional brain activation following sleep in participants who learned new words. Specifically, they found activation in the hippocampus before sleep consolidation, but after sleep, activation was observed in cortical areas. This indicates that the memory traces of the new words underwent qualitative changes in how they were represented in the brain. Therefore, sleep may indeed benefit learning or qualitatively reorganize information, even if these changes are not always evident in behavioural performance. All things considered, the benefits of strong initial learning seem clear: stronger encoding or overlearning typically results in better overall behavioural performance, and it can protect against interference from subsequent learning and potentially bestow benefits equivalent to those of sleep for long-term retention. This may be an important consideration for learning situations in which sleep following training is not possible.

6. Elucidating findings from non-native speech sound learning studies in the context of interference and stability

Concepts such as stability of learning and interference in memory consolidation may offer a more comprehensive account of some findings

from non-native speech sound learning. That fact that parallels emerge from several domains of learning (e.g., visual, motor, and word learning) may indicate that domain-general encoding, stabilization, and consolidation processes underlie many different types of learning, including speech sound learning. In fact, several studies reviewed above can be viewed through the lens of interference theories. For example, the finding by Earle and Myers (2015a), that native language exposure interfered with consolidation of a non-native, Hindi contrast, could be explained both by non-native speech sound learning theories and interference theories; however, a more comprehensive explanation could be arrived at by considering these theories together. Although theories of non-native speech sound learning (such as those reviewed above) differ on certain details and areas of focus, most attribute difficulties in non-native speech sound learning to perceptual similarity of native language speech sounds. In line with these theories, the stability and robustness of native-language phonetic categories may greatly enhance the difficulty of learning perceptually similar non-native categories. Additionally, both Müller and Pilzecker's (1900) consolidation hypothesis and Walker's (2005) procedural memory consolidation model postulate a necessary stabilization phase after learning takes place. If native language exposure immediately follows training on non-native speech sounds before the new speech category representations have had time to stabilize, native language input would interfere with these memory traces and impede or prevent consolidation from taking place. Results from Earle & Myers (2015a), in which native language exposure disrupted consolidation of a non-native phonetic contrast, diverge from the synthetic speech study by Fenn and colleagues (2003), in which sleep was able to recover information that was degraded (or possibly interfered with) throughout the course of a day. However, Ebbinghaus (1885) and other studies reviewed above support the view that strength and stability of learning is crucial to consolidation, and this notion may account for the discrepancy observed in these studies. Learning novel acoustic mappings to existing speech categories (that is, learning how the unusual synthetic speech signal maps to well-developed English phonology), as was done in the Fenn et al. (2003) paper, is arguably less difficult than establishing entirely new

perceptual categories. It is reasonable to speculate that both the time course of learning and consolidation and the strength of initial learning work in tandem to selectively consolidate memories. Learning may have been more stable for the synthetic speech task in Fenn et al. (2003) than the non-native speech sounds learned in Earle and Myers (2015a), which would explain why sleep-mediated consolidation was able to recover learning of synthetic speech that had decayed throughout the day but not the developing representations of non-native speech sounds.

Strength and stability of learning may further elucidate studies finding disruptions of learning or consolidation as a result of speech production (Baese-Berk & Samuel, 2016) or phonological variability (Fuhrmeister & Myers, 2017). Neither of these studies was designed according to the typical interference paradigm (learn task A, learn task B, test on task A); however, it appears that speech production and exposure to phonological variability resulted in representations that were less stable and less able to benefit from consolidation. Ultimately, the precise cause for attenuated perceptual learning following speech production remains unclear. Motor theories of speech perception posit that articulatory gestures underlie perceptual representations of speech (see Galantucci et al., 2006 for review). According to this view, it is possible that activating motor representations interferes with developing representations of speech categories. An additional possibility is that engaging native language phonological categories in any modality diminishes the strength and stability with which the novel categories are learned. Exposure to phonological variability may similarly reduce stability of learning: according to attention to dimension models of non-native speech sound learning, learners must direct their attention to relevant acoustic cues, which are, in many cases, different from the relevant cues for the first language (Francis & Nusbaum, 2002). Presentation of novel speech sounds in different phonological contexts may not allow the learner to quickly discover the acoustic cues that are necessary to distinguish different speech categories, as formant transitions sometimes change based on the vowel that follows a consonant. If the learner receives conflicting information for different phonological contexts, this would likely result in learning that is less stable, which may not benefit from consolidation, at least in the short term.

7. Promoting consolidation of non-native speech sounds

With models of interference and stability in memory consolidation in mind, it may be possible to improve training programs in order to support consolidation of novel phonetic categories. First, it is necessary to determine what types of stimuli can interfere with or destabilize this process. Specific tasks or stimuli that have the potential to interfere with non-native speech sound learning have not been extensively investigated. Nevertheless, by examining the extant literature, it appears that native language phonology is one source of interference (Earle & Myers, 2015a). Future perceptual training paradigms may induce more robust learning if they attempt to minimize exposure to native language phonology until a sufficient stabilization period has passed, especially if training takes place earlier in the day. If it is not possible to minimize native language exposure, learners may benefit from longer or more intensive training, in order to strengthen or stabilize learning. As seen in the study by Shibata and colleagues (2017), hyper-stable learning was resistant to subsequent interference, and non-native speech sound learning may show a similar pattern. Although some similarities between speech sound learning and visual or motor learning exist, this may be an area where speech diverges from other domains. Specifically, there are few everyday activities that come in conflict with the visual and motor tasks utilized in the studies reviewed above—for instance low-level line orientation detection tasks or compensation for perturbation in motor learning. On the other hand, is difficult to avoid speech in the real world, which presents challenges for experimental design. For example, participants who learn non-native speech categories in the laboratory most likely have immediate access to interfering or conflicting stimuli, such as their own speech production or acoustic speech input from listening to other talkers. Unless this is experimentally controlled (i.e., participants stay in the laboratory for an extended period of quiet time following training), it is difficult to account for the events that happen after a learning session. Due to practical limitations, this has yet to be explored.

 As discussed above, work by Baese-Berk and Samuel (2016) suggests articulation or production of speech sounds also seems to interfere with developing perceptual representations of speech. Interestingly however,

work by Bradlow and colleagues (1997) suggests that production accuracy can be enhanced by perceptual training alone, and this effect is stable over time (Bradlow et al., 1999). A similar study by Neufeld (1979) corroborates these results: participants who received perceptual training only were later able to produce words in a second language without a detectable non-native accent. It may be the case that perceptual training alone can induce concomitant improvements in speech production. Because relatively few studies have examined the relationship between speech perception and production and its influence non-native speech sound learning, future research will ultimately be needed to determine at what point in the learning trajectory and in what capacity production of new speech sounds is beneficial to the learner. With the available evidence, however, it seems that minimizing speech production in training, at least in the early stages of learning, may result in optimal outcomes for both perception and production of second language speech sounds.

Additionally, differing acoustic cues (e.g., the different formant trajectories associated with a dental stop in the context of an /i/ compared to an /u/ vowel) may be detrimental to stable learning and consolidation of non-native speech sounds. Thus, it may be advisable to limit phonological variability in training or during the stabilization phase following training. While several studies have found advantages of high-variability training procedures, these studies have typically taken place over the course of several weeks (e.g., Logan et al., 1991; Lively et al., 1993, 1994; Bradlow et al., 1997, 1999). In that case, the slow, procedural learning system may have had ample time to discover the regularities in the input (McClelland et al., 1995). This would also explain the enhanced generalization abilities as a result of this type of training; participants may have developed more robust abstract representations of the phonological categories, allowing for generalization to novel talkers or phonological contexts. While this may be the case, the efficacy of this training paradigm may be limited to certain situations. In particular, intensive training over long periods of time may not always be possible. For example, some second language classes meet only once per week, and this frequency may not be sufficient for learners to take advantage of high-variability training. As seen in Fuhrmeister and Myers (2017), even minimal exposure to phonological variability during a testing phase only (i.e., training was identical) attenuated learning of a

non-native contrast in the trained vowel context. Additionally, no over-night improvement was observed for those participants. This suggests that exposure to variability may not be optimal in certain cases, especially when training sessions are sparse. Especially in such situations, it is important that learners can consolidate newly acquired information to begin developing representations for non-native phonological categories. A more efficient training method may involve evening training sessions that occur close to sleep (to minimize native language exposure afterwards) that include only limited variability in the stimulus presentation. In fact, Earle and Myers (2015b) found generalization of non-native speech sounds to a novel talker following an interval of sleep, even though their training tokens consisted of sounds spoken by a single talker and presented in a single vowel context. Thus, sleep consolidation processes facilitated generalization to the sounds spoken by a new talker. Based on the evidence presented, striking a delicate balance between stimulus variability and proximity of training to sleep may promote consolidation of the trained information to long-term memory, which is essential to building novel phonological categories. Even so, second language learners experience different cues in the real world, and they need to be able to integrate information across them. Ultimately, future research will need to determine which factors promote abstraction over different acoustic cues.

   Although we have some evidence as to what types of stimuli have the potential to interfere with non-native speech sound learning, many questions remain open. For example, visual and motor learning studies have often found that not all tasks interfere equally. In the study discussed above by Seitz and colleagues (2005), visual stimuli that differed in spatial location and orientation did not interfere with a perceptual learning task. Because primary visual cortex is retinotopically organized, Seitz et al. (2005) reasoned that different neurons were responding to visual stimuli in different spatial locations and orientations, whereas the same neurons responded to the task in which only the direction of offset differed and location and orientation remained the same. These same neurons had to be overwritten in order to learn the second task. It is unknown whether any potential speech analogs for such a task exist. Because primary auditory cortex is tonotopically organized, however, it is possible that training on similar stimuli at a different frequency (e.g., varying talker gender) would

not interfere with training on the first frequency range. However, speech may differ from other types of learning due to the complexity of the signal. In addition, findings from motor learning suggest that training on an opposite force, as in Brashers-Krug et al. (1996), interferes with learning of the initial direction, and word learning studies using an A-B A-C paradigm indicate that a new pairing (C) with an original stimulus (A) interferes with recall of the original pairing (A-B). Speech correlates of such findings are less obvious, and future research will need to address these questions.

Next, the time course of consolidation and susceptibility to interference in non-native speech sound learning should be considered. If the consolidation hypothesis or Walker's (2005) model can be applied to speech learning, it is essential to determine under what conditions a stabilization phase is necessary, and when so, how long this stabilization period needs to be in order to protect memory traces from subsequent interference. In the visual, motor, and word learning studies reviewed above, several time frames ranging from a few minutes to several hours have proven successful in protecting information against interference; however, some studies found interference even after 24 hours had passed between training sessions on each task. This suggests that domain or task differences may be responsible for some of the varied results obtained in these studies. Based on the results of Earle and Myers (2015a), it seems clear that non-native speech learning would benefit from a stabilization period; however, this has yet to be explored in the speech domain. Ultimately, future research will need to elucidate the time course of stabilization and consolidation in the speech domain.

8. Interference and second language speech learning in naturalistic contexts

While this review has primarily focused on memory consolidation and interference in the context of laboratory learning of non-native speech sounds, these concepts may be relevant to naturalistic learning environments, as well. For example, an individual's amount of first language use (among other factors) has been found to predict speech production accuracy in the second language (Flege et al., 1997; Piske et al., 2001). One possible explanation for these results is that habitual

interference from the first language obstructs developing second language speech category representations, especially if the stabilization phase is consistently disrupted. Further support for this idea comes from studies measuring language proficiency following immersion programs. Immersion programs have largely been successful for second language acquisition, including acquisition of speech sounds (e.g., Anderson, 2004; Cheour et al., 2002; Freed et al., 2004). Immersion settings present few opportunities for interference from the native language, which may allow memory traces of second language speech sounds to stabilize and more efficiently be consolidated into long-term memory. In addition, the procedural memory system can likely develop more robust category representations with the time and amount of exposure afforded by the immersion setting to discover regularities in the second language speech system. In fact, the study by Freed et al. (2004) compared native English-speaking college students learning French in a classroom setting in the home country, an intensive summer immersion program in the home country, and a semester-long study abroad program in France. While students in the study abroad and summer immersion programs outperformed the classroom learners after the period of study, students in the study abroad program reported much more first language use (English) outside the classroom than the summer immersion group. Consistent with first language use accounts, students in the study abroad group made fewer gains in proficiency than the students in the immersion program. Walker et al.'s (2003) findings on reconsolidation may similarly explain why less frequent first language use contributes to better speech perception and production in the second language. In this study, they found that practice on a task that had already been consolidated through sleep could reactivate the memory trace of that task, causing it to return to a labile state. If participants learned a second task after reactivating memory traces of the first task, interference from the second task on the first was observed. If upon reactivation, memories return to a labile state subject to interference, using the second language may reactivate memories and transfer them to this labile state. If frequent and intermittent use of the first language interferes with the reactivated second language memory traces, it may have the power to interfere with reconsolidation of the second language memory traces, especially if first language memory traces are much more robust. Thus, studies of

interference in memory consolidation may have explanatory power even in naturalistic language learning settings.

9. Speech sound learning across the lifespan

A common empirical finding in studies of second language acquisition is that second language speech sound learning typically decreases with age. Many have attributed this to a putative critical period (e.g., Granena & Long, 2012); however, others have found a more linear decline in second language speech production abilities throughout the lifespan that would be less indicative of a critical period (e.g., Flege et al., 1995). Although a complete discussion of this issue is outside the scope of this chapter, some speculations can be made when considering both models of memory consolidation and non-native speech sound learning. Assuming no strictly defined critical period exists but non-native speech sound learning becomes increasingly more difficult throughout the course of the lifespan, ideas put forward in the learning and memory literature may be applicable. For example, the same system may underlie acquisition of speech sounds in infancy, childhood, and adulthood; however, adults cannot approach the speech sound learning task in the same way as an infant or child because their prior experience and interactions with the environment are vastly different (see Best & Tyler, 2007 for discussion). Infants do not have well-established first language speech categories, while adults have developed quite robust speech categories in the native language after years or decades of exposure. In fact, Burnham et al. (1991) found that speakers of a language become more categorical in their perception of native language speech sounds over the life span, and Baker et al. (2008) observed that children are less likely than adults to assimilate second language speech sounds to first language categories. This may imply that increasingly more stable first language categories are less malleable than less-stable categories, such as those in childhood. This could have several implications for non-native speech sound learning. As predicted by non-native speech perception and learning models, the native language exerts a powerful influence on the perception and ability to learn non-native speech categories. Additionally, theories of learning and memory (e.g., Ebbinghaus, 1885) and findings in the visual domain by Shibata et al. (2017) suggest that stable or strongly

learned information can cause proactive interference. In this way, theories from both of these domains could be taken together to imply that native language speech categories interfere with non-native speech sound learning in a graded manner: stronger native language categories may increase the difficulty in developing new categories. In addition, children may be less susceptible to interference in certain types of learning prior to puberty. In a study by Dorfberger and colleagues (2007), children before and after the onset of adolescence learned a procedural motor task. Younger children showed no advantage in learning or retaining the sequence; however, when they were trained on an additional, opposing sequence, 9- and 12-year-olds demonstrated no evidence of interference of the second sequence on the first. Seventeen-year-olds, on the other hand, did show an interference effect. This finding lends support to critical period hypotheses but in a different way than they are traditionally depicted. Proponents of a critical period typically focus on child advantages in learning, but this finding suggests adults may be as good as children at learning certain tasks; however, their consolidation processes may differ. Some studies investigating second language speech learning in children and adults have found superior perceptual learning in adults initially, but after some time children not only caught up to the adults in performance but actually surpassed them (Snow & Hoefnagel-Höhle, 1978). This finding may be consistent with the Dorfberger et al. (2007) study: interference from the first language may not have influenced children's learning of novel speech sounds, while it may have obstructed adults' learning over time.

Simultaneous and sequential bilinguals may additionally offer some insight into memory consolidation of speech sounds over the lifespan. For example, children who begin learning a second language in early childhood seldom have a detectable non-native accent in either language; however, late-onset bilinguals often do (e.g., Flege et al., 1995). Thus, it appears that infants and young children can learn the speech sound inventory of two languages simultaneously without interference from either language. Assuming the slow, procedural memory system subserves this process (McClelland et al., 1995; Dudai, 2004), this system would discover regularities in both sound systems simultaneously, even if speech categories in both languages are close in perceptual space. For example, an infant learning Spanish and English would learn that the distribution of the

bilabial voiced stop consonant clusters around a lower voice onset time than the same category in English. The lack of experience with a specific language's speech sound inventory may additionally facilitate the acquisition of speech categories in two languages simultaneously.

Furthermore, evidence from bilinguals who learned languages sequentially may shed light on this process. For example, a study by Antoniou et al. (2012) found that early sequential bilinguals who were dominant in their second language were influenced by their dominant (second) language in their perception of consonants in both languages. This parallels previously discussed findings showing effects of first language use on second language speech perception and production: less frequent first language use (or more frequent second language use) may be a strong factor in the development of second language speech categories. In other words, while important, age of initial acquisition of a language is not necessarily the determining factor in how well the speech system of that language will be learned. Ultimately, future research will be needed to elucidate relative influences of age of acquisition and amount of language use on the development of second language speech category representations.

10. Conclusion

Although a comprehensive account of the non-native speech sound learning process has yet to be established, many insights can be gained from considering findings from domain-general learning and memory studies. In particular, factors such as interference during a critical stabilization phase following learning and the strength of initial encoding may be important considerations when designing training paradigms for learning novel speech categories or when interpreting findings from this field. Whether models of memory consolidation can be applied to speech sound learning without modification remains unclear; however, they make concrete predictions for future studies and appear to have a great deal of explanatory power within the current literature.

Acknowledgements

References

Anderson, R.T. (2004). Phonological acquisition in preschoolers learning a second language via immersion: A longitudinal study. *Clinical Linguistics & Phonetics*, *18*(3), 183–210.

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*(4), 582–594.

Atienza, M., Cantero, J.L., & Stickgold, R. (2004). Posttraining sleep enhances automaticity in perceptual discrimination. *Journal of Cognitive Neuroscience*, *16*(1), 53–64.

Baese-Berk, M.M., & Samuel, A.G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, *89*, 23–36.

Baker, W., Trofimovich, P., Flege, J.E., Mack, M., & Halter, R. (2008). Child—adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech*, *51*(4), 317–342.

Best, C.T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *Haskins Laboratories Status Report on Speech Research*. SR-107/108, 1–30.

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 Acoustical Society of America*, *109*(2), 775–794.

Best, C.T., McRoberts, G.W., & Sithole, N.M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360.

Best, C.T., & Tyler, M.D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M.J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of*

language experience in speech perception and production. Amsterdam: John Benjamins, pp. 13–34.

Bradlow, A.R., Akahane-Yamada, R., Pisoni, D.B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long- term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977–985.

Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of Acoustical Society of America*, *101*(4), 2299–2310.

Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature*, *382*(6588), 252–255.

Brown, H., & Gaskell, M.G. (2014). The time-course of talker-specificity and lexical competition effects during word learning. *Language, Cognition and Neuroscience*, *29*(9), 1163–1179.

Burnham, D.K., Earnshaw, L.J., & Clark, J.E. (1991). Development of categorical identification of native and non-native bilabial stops: infants, children and adults. *Journal of Child Language*, *18*(2), 231–260.

Caithness, G., Osu, R., Bays, P., Chase, H., Klassen, J., Kawato, M., Wolpert, D.M. & Flanagan, J.R. (2004). Failure to consolidate the consolidation theory of learning for sensorimotor adaptation tasks. *Journal of Neuroscience*, *24*(40), 8662–8671.

Cheour, M., Shestakova, A., Alku, P., Ceponiene, R., & Näätänen, R. (2002). Mismatch negativity shows that 3–6-year-old children can learn to discriminate non-native speech sounds within two months. *Neuroscience Letters*, *325*(3), 187–190.

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*(4), 803-820.

Dewar, M., Garcia, Y.F., Cowan, N., & Sala, S.D. (2009). Delaying interference enhances memory consolidation in amnesic patients. *Neuropsychology*, *23*(5), 627–634.

Dorfberger, S., Adi-Japha, E., &Karni, A. (2007). Reduced susceptibility to interference in the consolidation of motor memory before adolescence. *PLoS One*, *2*(2), e240.

Drosopoulos, S., Windau, E., Wagner, U., & Born, J. (2007). Sleep enforces the temporal order in memory. *PLoS One*, *2*(4), e376.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology*, *55*, 51–86.

Dumay, N., & Gaskell, M.G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35–39.

Earle, F.S., Landi, N., & Myers, E.B. (2017). Sleep duration predicts behavioural and neural differences in adult speech sound learning. *Neuroscience Letters*, *636*, 77–82.

Earle, F.S., & Myers, E.B. (2014). Building phonetic categories: an argument for the role of sleep. *Frontiers in Psychology*, *5*, 1192.

Earle, F.S., & Myers, E.B. (2015a). Sleep and native language interference affect non-native speech sound learning. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1680–1695.

Earle, F.S., & Myers, E.B. (2015b). 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.

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Leipzig: Verlag von Duncker und Humblot.

Eimas, P.D., Siqueland, E. R., Juscyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306.

Eisner, F., & McQueen, J.M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, *119*(4), 1950–1953.

Ellenbogen, J.M., Hulbert, J.C., Jiang, Y., & Stickgold, R. (2009). The sleeping brain's influence on verbal memory: boosting resistance to interference. *PLoS One*, *4*(1), e4117.

Ellenbogen, J.M., Hulbert, J.C., Stickgold, R., Dinges, D.F., & Thompson-Schill, S.L. (2006). Interfering with theories of sleep and memory: sleep, declarative memory, and associative interference. *Current Biology*, *16*(13), 1290–1294.

Fenn, K.M., Margoliash, D., & Nusbaum, H.C. (2013). Sleep restores loss of generalized but not rote learning of synthetic speech. *Cognition*, *128*, 280–286.

Fenn, K.M., Nusbaum, H., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language of perceptual learning. *Nature*, *425*(6958), 614–616.

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*. Timonium, MD: York Press, pp. 229–273.

Flege, J.E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities*. Berlin: Mouton de Gruyter, pp. 319–355.

Flege, J.E., Frieda, E.M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*(2), 169–186.

Flege, J.E., Munro, M.J., & MacKay, I.R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125–3134.

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*(2), 349–366.

Freed, B.F., Segalowitz, N., & Dewey, D.P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*(2), 275–301.

Fuhrmeister, P. Myers, E.B. (2017). Non-native phonetic learning is destabilized by exposure to phonological variability before and after training. *The Journal of the Acoustical Society of America*. *142*(5), EL448–EL454.

Galantucci, B., Fowler, C.A., & Turvey, M.T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377.

Goedert, K.M., & Willingham, D.B. (2002). Patterns of interference in sequence learning and prism adaptation inconsistent with the consolidation hypothesis. *Learning & Memory*, *9*(5), 279–292.

Golestani, N., & Zatorre, R.J. (2004). Learning new sounds of speech: reallocation of neural substrates. *Neuroimage*, *21*(2), 494–506.

Golestani, N., & Zatorre, R.J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language*, *109*(2–3), 55–67.

Granena, G., & Long, M.H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311–343.

Hauptmann, B., Reinhart, E., Brandt, S. A., & Karni, A. (2005). The predictive value of the leveling off of within-session performance for procedural memory consolidation. *Cognitive Brain Research*, *24*(2), 181–189.

Heim, S., Klann, J., Schattka, K.I., Bauhoff, S., Borcherding, G., Nosbüsch, N., Struth, L., Binkofski, F.C., &Werner, C.J. (2017). A nap but not rest or activity consolidates language learning. *Frontiers in Psychology*, *8*, 665.

Kuhl, P.K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, *4*(6), 812–822.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D.,Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1493), 979–1000.

Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21.

Kuriyama, K.,Stickgold, R., & Walker, M.P. (2004). Sleep-dependent learning and motor-skill complexity. *Learning & Memory*, *11*(6), 705–713.

Lahl, O., Wispel, C., Willigens, B., & Pietrowsky, R. (2008). An ultra short episode of sleep is sufficient to promote declarative memory performance. *Journal of Sleep Research*, *17*(1), 3–10.

Lim, S.J., & Holt, L.L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science*, *35*(7), 1390–1405.

Lively, S.E., Logan, J.S., & Pisoni, D.B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic

environment and talker variability in learning new perceptual categories. *The Journal of Acoustical Society of America*, *94*(3), 1242–1255.

Lively, S.E., Pisoni, D.B., Yamada, R.A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of Acoustical Society of America*, *96*(4), 2076–2087.

Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of Acoustical Society of America*, *89*(2), 874–886.

MacKay, I.R., Meador, D., & Flege, J.E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica*,*58*(1–2), 103–125.

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences*, *11*(10), 442–450.

Mazza, S., Gerbier, E., Gustin, M.P., Kasikci, Z., Koenig, O., Toppino, T.C., & Magnin, M. (2016). Relearn faster and retain longer: Along with practice, sleep makes perfect. *Psychological Science*, *27*(10), 1321–1330.

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, & Behavioural Neuroscience*, *2*(2), 89–108.

McClelland, J.L. (1998). Complementary learning systems in the brain: A connectionist approach to explicit and implicit cognition and memory. *Annals of the New York Academy of Sciences*, *843*(1), 153–169.

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*(3), 419–457.

McCloskey, M., & Cohen, N.J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.) *Psychology of learning and motivation*, Vol. 24, Academic Press. pp. 109–165.

Muellbacher, W., Ziemann, U., Wissel, J., Dang, N., Kofler, M., Facchini, S., Boroojerdi, B.,Poewe, W., & Hallett, M. (2002). Early consolidation in human primary motor cortex. *Nature*, 415(6872), 640–644.

Müller, G.E., & Pilzecker, A. (1900). *Experimentelle Beiträge zur Lehre vom Gedächtnis*. (Vol. 1). JA Barth.

Myers, E.B., & Swan, K. (2012). Effects of category learning on neural sensitivity to non-native phonetic categories. *Journal of Cognitive Neuroscience*, *24*(8), 1695–1708.

Neufeld, G.G. (1979). Towards a theory of language learning ability. *Language Learning*, *29*(2), 227–241.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioural plasticity in speech perception. *Cognition*, *64*(3), B9–B17.

Piske, T., MacKay, I.R., & Flege, J.E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215.

Poldrack, R.A., Clark, J., Pare-Blagoev, E.J., & Shohamy, D. (2001). Interactive memory systems in the human brain. *Nature*, *414*(6863), 546–550.

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*(2), 159–164.

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 of the United States of America*, *102*(41), 14895–14900.

Shibata, K., Sasaki, Y., Bang, J.W., Walsh, E.G., Machizawa, M.G., Tamaki, M., Chang, L.-H. & Watanabe, T. (2017). Overlearning hyper-stabilizes a skill by rapidly making neurochemical processing inhibitory-dominant. *Nature Neuroscience*, *20*(3), 470–475.

Silbert, N.H., Smith, B.K., Jackson, S.R., Campbell, S.G., Hughes, M.M., & Tare, M. (2015). Non-native phonemic discrimination, phonological short term memory, and word learning. *Journal of Phonetics*, *50*, 99–119.

Snow, C.E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, *49*(4), 1114–1128.

Squire, L.R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171–177.

Stickgold, R., LaTanya, J., & Hobson, J.A. (2000). Visual discrimination learning requires sleep after training. *Nature Neuroscience*, *3*(12), 1237–1238.

Swan, K., & Myers, E. (2013). Category labels induce boundary-dependent perceptual warping in learned speech categories. *Second Language Research*, *29*(4), 391–411.

Tamminen, J., & Gaskell, M.G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *The Quarterly Journal of Experimental Psychology*, *66*(5), 1001–1025.

Tsao, F.M., Liu, H.M., & Kuhl, P.K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, *75*(4), 1067–1084.

Tucker, M.A., & Fishbein, W. (2008). Enhancement of declarative memory performance following a daytime nap is contingent on strength of initial task acquisition. *Sleep*, *31*(2), 197–203.

Vlahou, E.L., Protopapas, A., & Seitz, A.R. (2012). Implicit training of nonnative speech stimuli. *Journal of Experimental Psychology: General*, *141*(2), 363–381.

Wade, T., & Holt, L.L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *The Journal of the Acoustical Society of America*, *118*(4), 2618–2633.

Walker, M.P. (2005). A refined model of sleep and the time course of memory formation. *Behavioural and Brain Sciences*, *28*(1), 51–64.

Walker, M.P., Brakefield, T., Hobson, J.A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature*, *425*(6958), 616–620.

Werker, J.F., & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63.

Xie, X., Earle, F.S. & Myers, E.B. (2017). Sleep facilitates generalisation of accent adaptation to a new talker. *Language, Cognition and Neuroscience*, *33*(2), 196–210.

Yi, H.G., Maddox, W.T., Mumford, J.A., & Chandrasekaran, B. (2014). The role of corticostriatal systems in speech category learning. *Cerebral Cortex*, *26*(4), 1409–1420.