



Building phonetic categories: an argument for the role of sleep

F. Sayako Earle^{1*} and Emily B. Myers^{1,2,3}

¹ Department of Speech, Language, and Hearing Sciences, University of Connecticut, Storrs, CT, USA

² Department of Psychology, University of Connecticut, Storrs, CT, USA

³ Haskins Laboratories, New Haven, CT, USA

Edited by:

Marcela Pena, Catholic University of Chile, Chile

Reviewed by:

LouAnn Gerken, University of Arizona, USA

Ina Bornkessel-Schlesewsky, University of South Australia, Australia

*Correspondence:

F. Sayako Earle, Department of Speech, Language, and Hearing Sciences, University of Connecticut, 850 Bolton Road, Unit 1085, Storrs, CT 06269, USA
e-mail: frances.earle@uconn.edu

The current review provides specific predictions for the role of sleep-mediated memory consolidation in the formation of new speech sound representations. Specifically, this discussion will highlight selected literature on the different ideas concerning category representation in speech, followed by a broad overview of memory consolidation and how it relates to human behavior, as relevant to speech/perceptual learning. In combining behavioral and physiological accounts from animal models with insights from the human consolidation literature on auditory skill/word learning, we are in the early stages of understanding how the transfer of experiential information between brain structures during sleep manifests in changes to online perception. Arriving at the conclusion that this process is crucial in perceptual learning and the formation of novel categories, further speculation yields the adjacent claim that the habitual disruption in this process leads to impoverished quality in the representation of speech sounds.

Keywords: speech perception, learning, memory, consolidation, phonology

INTRODUCTION

Categorical perception refers to the phenomenon by which listeners demonstrate non-linear perception of items across an acoustic-phonetic continuum. In speech, we observe this behaviorally in the more accurate discrimination of two different phonetic tokens lying across a category boundary, compared to two phonetic tokens of a comparable acoustic distance occurring within the boundaries of the same category (Liberman et al., 1957; see Repp and Liberman, 1987 for review). Phonological contrasts correspond to the sounds of speech that are capable of signaling change in the lexical meaning in the signal (Chomsky and Halle, 1968), and each phoneme encompasses many acoustically distinct phonetic tokens. During spoken language processing, listeners map incoming phonetic tokens to existing phonological representations. The formation of phonetic categories¹ is mediated by different stages of memory processing, which encode one's experience with novel speech sounds and lead to generalization away from the context of that initial encounter. Thus, the focus of this review will be on the different memory processes that aggregate auditory experience into functional knowledge, and how these processes support category acquisition. Given that instability in speech sound representation has been linked to developmental language-based disorders such as dyslexia (Nation and Snowling, 1998) and Specific Language Impairment (Joanisse and Seidenberg, 2003), understanding of the memory encoding processes that impact phonetic acquisition may provide specific predictions

for how disruptions manifest in behavioral symptoms, leading to the potential loci of breakdown or deficit.

In the present discussion, we review the theoretical models of category representation, arriving at the conclusion that two distinct cognitive processes support the perception of phonological categories: selective attention to certain features within the acoustic signal, and the recollection of a context within which those features are relevant. We argue that two types of learning are involved in the acquisition of categories: listeners must develop perceptual automaticity for selectively attending to familiar acoustic-phonetic features, and they must encode the context that dictates the relevance of these features. We then review selected literature on memory consolidation in auditory/perceptual skill and lexical form acquisition. Through this discussion, we illustrate that both types of learning are supported by memory encoding processes that take place during sleep.

MEMORY IN MODELS OF SPEECH SOUND CATEGORY REPRESENTATION

Category acquisition is subject to the lack of invariance problem, in that "categories" represent our relatively constant perception of a signal that is subject to acoustic variability according to phonetic context, environmental conditions, and speaker idiosyncrasy (Nygaard et al., 1994; Nygaard and Pisoni, 1998; Allen and Miller, 2004; Eisner and McQueen, 2005; Kraljic and Samuel, 2005). The construction of a new category representation therefore requires the abstraction of information away from the individual acoustic events, such that knowledge of phonetic structure can be applied generally to novel tokens, new talkers, and different phonological

¹We use "phonetic category" to describe the distribution of acoustic tokens which together are perceived as mapping to a phoneme in the listener's inventory.

contexts. Models of speech processing posit different theoretical constructs for how category knowledge is represented; however, they share in common the assumption that perception is influenced by experience (Kuhl, 1991; Iverson and Kuhl, 1995; Goldinger, 1998). This section will review how selected models of spoken language processing conceptualize phonetic categories, and how acoustic tokens might enter long-term store in order to exert an influence on online processing.

Attempts to resolve the lack of invariance problem must account for our propensity to attend specifically to the details of the acoustic signal that reflect features signaling a change in category. One framework, which posits the storage of idealized phonetic “prototypes,” is a popular construct by which to discuss the representation of speech sound categories (Iverson et al., 2003; Guenther et al., 2004; Athos et al., 2007). The notion of prototypes is realized by appropriating the generalized context model (GCM; Nosofsky, 1986), which allows the consideration of acoustic-phonetic features as dimensional structures within a perceptual space. Within this framework, language experience is thought to decrease the distance between acoustic tokens in the vicinity of prototypes, such that they behave as attractors that perceptually “pull” in acoustic items that fall within its boundaries [Native Language Magnet (NLM) model, Kuhl, 1991; Iverson and Kuhl, 1995]. Prototypes offer a potential solution to the lack of invariance problem by mapping many tokens onto one category, such that only one token, the prototype, is stored in memory for each phonetic category. While this is advantageous from the standpoint of efficiency in category retrieval, certain behavioral phenomena have prompted current iterations of magnet theory to consider the emergence of attractors as a consequence of statistical regularity in the input (Kuhl et al., 1992; Maye et al., 2002, 2008). The notion that one distribution in perceptual space reliably maps to one prototype makes it more difficult to account for context-dependent adjustments to the phonetic category boundary. For example, listeners adjust the perceptual boundary in the /s/ - /f/ continuum to the specific speaker after a period of exposure to the speaker’s particular pattern of articulation (Norris et al., 2003). Furthermore, listeners’ accommodations for speaker idiosyncrasy appear to be different for spectral and voice onset time (VOT) contrasts (Kraljic and Samuel, 2005). In addition, despite adults’ initial difficulty in perceiving non-native categories (Best, 1994; Iverson et al., 2003), perceptual training results in improved discriminability of non-native contrasts (Golestani and Zatorre, 2009; Swan and Myers, 2013). Under the prototype framework, this suggests that relatively little experience with non-native sounds can establish prototypes for novel categories, calling into question the amount and type of exposure that is necessary in order for categories to form.

An alternative is to consider a many-to-many mapping, such that each speech event is stored in memory as an individuated representation (Exemplar theory: Medin and Schaffer, 1978; Goldinger, 1998; see Pisoni and Levi, 2007 for review). Under this view, category identity is assigned through mapping the token to the category of similar sounds within previously encountered events. The traditional notion of a single prototypic exemplar acting as a perceptual magnet is replaced with the topography of perceptual space that is warped by the coherent convergence of

past exposure to native exemplars. Thus, a prediction of how a non-native token will be perceived would not be determined by the relationship between the token and the prototypical exemplar, but by the probability by which the closest previously encountered exemplar belongs to a particular category. Similarly, statistical/distributional accounts (Pierrehumbert, 2003; Maye et al., 2008; Feldman et al., 2009) allow for episodes, over time, to create valleys in perceptual space without the provision of context. In this sense, these accounts are not wholly incompatible with NLM, in that a hypothetical token possessing the “average” of the featural values of all previously encountered exemplars in a given category would be in essence the category prototype. The difference between these perspectives is the nature of what is represented in memory: whether one prototypical exemplar of a category is stored as a perceptual attractor, or if an attractor emerges from the amalgamation of every experience individually stored. Furthermore, by exemplar-based and distributional accounts, the literal existence of “categories” is rendered unnecessary, as categorical perception may be epiphenomenal to the lexical and indexical cues that are stored in the representations of individual speech events.

The many-to-many mapping viewpoint thus offers a solution to the lack of invariance problem that accounts for stable percepts of “category” that also flexibly adapt to context. Experience, however, increases the number of previously stored encounters, increasing the number of tokens against which new encounters must be considered; in other words, the literal storage of individuated exemplars may tax perceptual efficiency rather than facilitate it. Versions of the exemplar model address this issue through considering the reorganization of perceptual space as the modification of weights on the connections between tiers of information. For example, in Kruschke’s (1992) Attention Learning Covering Map (ALCOVE) model, the three-layer connectionist model architecture specifies stimulus dimension nodes that receive the input, a hidden layer in which every exemplar is represented by a single node, and category nodes at the output. The weights between the exemplar and category nodes are adjusted per trial by a learned association algorithm, while the connections between stimulus features and exemplars are adjusted by a selective attention algorithm such that “attending” to a particular feature during training will expand the perceptual distance along that continuum. Thus, the model’s mechanism of perceptual learning is to alter its degree of sensitivity to distinctions in features. This key innovation of the model in incorporating selective attention allows for active cognitive processing during the perceptual event to mediate the gradation of impact that experience has on the underlying perceptual topography. In this sense, while exemplars are the currency upon which the model is built, the model itself results in a warping of perceptual space, such that processing new tokens is achieved by active competition and cooperation between the increased automaticity of bottom-up processing and a set of static category boundaries.

There remains, however, an outstanding issue that is common to both many-to-one mapping and many-to-many mapping approaches that needs to be addressed. This is the problematic assumption that the perception of any exemplar is pre-determined, prior to the perceptual task and regardless of context,

by the relationship between the stimulus and the listener's topography of underlying perceptual space (Nusbaum and Magnuson, 1997). This assumption breaks down when we consider the contribution of listener expectation. For example, the listeners' expectations about the speaker's gender alter their perception of an ambiguous fricative along the /s/ - /ʃ/ continuum (Strand, 1999). Furthermore, prior knowledge of the message content can render an otherwise unintelligible message delivered by sine-wave speech intelligible (Remez et al., 1981). Finally, in a speeded target-monitoring task, Magnuson and Nusbaum (2007) found that participants instructed that they will hear two different talkers responded more slowly on a speech identification task containing alternating voices relative to participants who were given no such instruction, despite both groups receiving acoustically identical stimuli. Thus, the listener's expectations about the signal prior to exposure appears to mediate signal processing, which is incongruous with the notion that perception is determined only by the features of the incoming signal interacting with the pre-task topography of perceptual space. Memory of previous experience (one vs. multiple talkers in this case) appears to set up listener expectations that constrain the set of features to which he/she selectively attends.

In concert with this idea, the Attention-to-dimension approach (A2D; Francis et al., 2000; Francis and Nusbaum, 2002) assumes no default (pre-task) warping of perceptual space, and points to active selective attention giving rise to the differential weighting of acoustic features as relevant for the completion of a perceptual task. To illustrate, Francis et al. (2000) found that listeners given the same set of acoustic cues, but different category labels by which to assign them, learned to attend to different features in the signal. The authors suggest that the top-down input of category, which in this case was provided, inform the listener's allocation of selective attention to different features for completion of the task. To reiterate, in contrast to exemplar accounts, A2D posits that category identity is not determined by the information contained within the acoustic signal, but rather the listener's active cognitive processes acting upon the input. This distinction allows for listeners to receive identical acoustic inputs and yet differ behaviorally in a perceptual task according to their expectations about the signal (e.g., Magnuson and Nusbaum, 2007). This is a qualitatively different role of selective attention with respect to ALCOVE, in that attention's primary role is not to make lasting changes to perceptual space, but to give task-specific guidance for perception. This account better explains how adults with no prior foreign language experience can learn to perceive non-native contrasts quickly with training. Specifically, this framework suggests that while increased practice may reduce attentional effort, there is no default pre-task bias that determines perception. Rather, we rely on memory to provide a representation of context (such as the lexical context, speaker, or category label) during online processing that constrains the acoustic dimensions along which we allocate attentional resources.

This idea is further consistent with Goldstone's (1994) solution for category formation in visual perception: while the physiological constraints of signal perception may somewhat guide attention along certain dimensions, "category" is a malleable, abstract construct that provides top-down context for guiding attention

along certain features. In support of this idea, Goldstone (1995) demonstrated that participants' judgments on the saturation of a hue were mediated by the object on which the color was perceived (i.e., the redness on a typically red object is judged more red than the identical redness on a typically yellow object), indicating that one's abstract representation of an object's typical color category affects perception of the featural (color intensity) value. In speech terms, this allows language novices with little experience to have functional boundaries for novel categories, and each encounter with a new token will refine their knowledge about the context-dependent location of these boundaries.

In order to connect perceptual learning to the memory literature, we discuss what is learnt in terms of the memory system active during learning. We emphasize, however, that the conceptual underpinnings resonate with existing theoretical models of category representation. Kruschke's model posits that two types of information are considered relevant to perception; similarly, we propose that the acquisition of novel categories involves both the encoding of a context (talker/lexical/category label, etc.) that specifies the acoustic-phonetic features of interest (exemplar theory), and the reduction of effort/increased automaticity in attending to values along those continua (A2D). Furthermore, behavioral assessments of these two types of learning are likely to engage different memory systems. Specifically, automaticity of attention to features is proposed to be enhanced by the procedural memory system, while the top-down source of influence (context-specific location of category boundaries), are predicted to be encoded by the structures that support declarative memory formation. Before highlighting the rationale for this distinction, the discussion will first turn to a focused review of memory consolidation as it relates to speech-related learning, and the roles of sleep on procedural and declarative memory. In combining these insights, we will arrive at a means by which we can test potential loci of deficits in individuals who struggle to form stable representations.

OVERVIEW OF MEMORY CONSOLIDATION

As indicated above, memory processes are implicit in all models of category acquisition. In order to understand the nature of the representation that is formed, it is important to understand how phonetic details enter long-term memory through memory consolidation processes. The term "consolidation" was coined by Müller and Pilzecker (1900), who observed that the recall of a list of non-sense syllables were more accurate in participants tested following an interval of time, in comparison to those who were tested shortly after learning. The past half-century of research in the time course of selective changes to the memory trace has highlighted at least two different types of consolidation processes that take place (Dudai, 2004). These events are the local changes to neural structure immediately (minutes to hours) following the learning event (synaptic, early, or short-term consolidation), and the slow (days to months) reorganization of the neural network reflecting long-term storage (systems, late, or long-term consolidation). Therefore, "consolidation" is a collective reference to the stages of information processing that result in either local or systemic neural changes, with qualitative effects on the recollected memory trace observable in behavior. In relation to the acquisition of novel phonetic categories, we propose that synaptic

consolidation precipitates perceptual automaticity, while systems consolidation results in the context representation that specifies the acoustic-phonetic features of interest.

SYNAPTIC CONSOLIDATION

Synaptic consolidation is proposed to be the mechanism underlying enhanced procedural learning (Vertes, 2004), and therefore tied to acquiring perceptual automaticity. Following an inducing (perceptual) event, early long-term potentiation (LTP; Bliss and Lomo, 1973), in conjunction with late-phase LTP, or synaptic consolidation, stabilizes the initial expression of the memory trace. LTP is a persistent enhancement of synaptic potential immediately following an inducing event that allows weaker inputs to also strengthen synapses when activated. During late-phase LTP, or synaptic consolidation, protein synthesis yields structural changes to active dendritic spines (Bramham and Messaoudi, 2005), crucial to memory maintenance beyond the first few minutes to hours. A standard model of synaptic consolidation (Bliss and Collingridge, 1993; Dudai, 2004) summarizes that processes associated with early LTP, lasting immediately to a few hours following induction, yield local changes that decay in time. Long-term changes to synaptic strength further require induction by a strong stimulus that triggers protein synthesis, resulting in stabilization of the active synapse by enlarging the dendritic spine and chemically binding the active pre-/post-synaptic terminals. The resultant synapse is more efficient in signal transmission and resistant to modification. Below, we discuss the evidence that sleep promotes latent synaptic consolidation; behavioral correlates of sleep-mediated synaptic consolidation includes the enhanced ability for explicit recall of new information (e.g., Rasch et al., 2007), or enhanced proficiency in an acquired skill (e.g., Walker et al., 2003) including auditory perceptual skill (Brawn et al., 2010). We argue that similar processes promote perceptual automaticity in speech.

SYSTEMS CONSOLIDATION

We propose that distributed (and therefore generalizable) context representations (such as the lexical context, speaker, environment, etc.) that provide top-down guidance for online perception results from the abstraction of episodic information during systems consolidation (Davis et al., 2009). Systems consolidation is a process by which new memories are integrated with pre-existing information, and is primarily associated with hippocampal-encoded memory. The account of systems consolidation is based on a theoretical argument for the necessity of separate structures for temporary vs. permanent storage of memory (McClelland et al., 1995). During an experience, the hippocampus is suggested to automatically capture cortical activity and store an index to that pattern as a minimal representation that can be later reactivated for the recollection of an event (e.g., McNaughton and Morris, 1987; O'Reilly and Rudy, 2001). However, the hippocampal architecture is unlikely to provide a solution for permanent memory storage (Marr, 1971). McClelland et al. (1995) illustrated the necessity of separate structures for the temporary vs. permanent storage of memory, by revisiting Rumelhart's (1990) simulations of semantic learning. The network implements learning as the unsupervised discovery of shared structures following exposure

to various exemplars (as denoted by a set of features) of target concepts. Rapid, focused learning by such networks lead to catastrophic interference, such that new information overrides conflicting pre-existing information. This is resolved by modifying the training procedure to gradual, interleaved learning between trials of different concepts with shared features. During the capture of an experience, rapid encoding of sensory information does not allow for slow integration with pre-existing knowledge: thus, the authors proposed that the hippocampus gives temporary storage to rapid encoding of episodes, which are slowly integrated with the cortex off-line for long-term storage. Semantic knowledge may then arise from the shared structures that emerge in the cortex over a lifetime of experience. In speech learning, systems consolidation may give rise to context-independent, abstract category knowledge that allows listeners to recognize sounds in unfamiliar lexical items or spoken by novel talkers.

By comparison to synaptic consolidation, systems consolidation is less well understood. This is in part because systems consolidation is a slow process, whereas synaptic consolidation for a new trace is generally accepted to complete within 24 h. Furthermore, systems consolidation is difficult to establish in behavior in that performance must somehow assess the state of the memory trace as either in an integrated or non-integrated state. Candidate behavioral (e.g., Dumay and Gaskell, 2007) and physiological observations (e.g., Ji and Wilson, 2007) that potentially reflect this off-line transfer of information between short and long-term storage is discussed in the following section.

THE RELEVANCE OF SLEEP TO SPEECH-RELATED LEARNING

There is evidence to suggest that sleep supports various aspects of language acquisition (see Gomez et al., 2011 for review). For example, naps have been found to support the abstraction of syntactic dependencies of an artificial language (Gómez et al., 2006), and in the retention of these abstractions 24-h post-exposure (Hupbach et al., 2009). Exposure to speech sound stimuli during sleep has been further demonstrated to lead to significant change in preattentive discrimination in newborns (Cheour et al., 2002). In the wider consolidation literature, there is debate as to the necessity of sleep to memory consolidation, in part due to the confusion in terms; for example, consolidation has been linked to stabilization against interference (Brashers-Krug et al., 1996), skill enhancement (Karni et al., 1994), integration (Tamminen et al., 2010), or selective forgetting (Saletin et al., 2011). Interpretation of consolidation effects as dependent on a period of time spent in sleep state, or as merely the result of passage of time in either wake or sleep, largely depends on the behavioral metric used. Below, we address how these differences in consolidation effects reflect learning that engages different memory systems according to the type of information acquired or the manner in which it is acquired (Tulving, 1985; Nissen and Bullemer, 1987; Squire, 1992). This is important to keep in mind, as the task used to measure consolidation in phonetic learning will affect the presence and nature of effects that are specific to mediation by sleep. To specify, "procedural" learning is considered as the implicit acquisition of skill (Doyon, 1997), and "declarative" learning indicates acquisition of information that can be explicitly recalled (O'Keefe and Nadel, 1978; O'Reilly and Rudy, 2001). It

has been suggested that rapid-eye movement (REM) sleep preferentially consolidates procedural memory, while non-REM/slow-wave sleep (SWS) promotes consolidation of declarative memory (Marshall and Born, 2007). This review will discuss the evidence for these claims, in relation to the kinds of learning that are relevant to our discussion of phonetic learning.

Consolidation during REM sleep and auditory skill learning

Learning-induced perceptual automaticity is considered here as analogous to the acquisition of an auditory/perceptual skill. There is a wide literature documenting the sleep-mediated improvement of procedural memory (see Smith, 2001; Stickgold, 2005 for review). A period of sleep, as compared to a comparable period of wakefulness, has been observed to lead to improved visual perceptual discrimination (Stickgold et al., 2000a,b), motor skill learning (Walker et al., 2002, 2003; Fischer et al., 2005), and auditory discrimination (Brawn et al., 2010). The memory literature often discusses auditory/perceptual skill acquisition as an example of procedural learning, in that this type of learning is measured by the implicit improvement in skill/task performance rather than in the declarative recall of experience (e.g., Gaab et al., 2004; Walker and Stickgold, 2004; Margoliash and Fenn, 2008). Within this literature, convergent evidence points to a time-dependent stabilization of a learned auditory skill, followed by a sleep-mediated increase in perceptual automaticity. For example, Gaab et al. (2004) investigated the retention of pitch learning in the 24 h following initial training at three time points post-training (immediate, 12-h post, and 24-h post). The authors found that the participants, regardless of when they were trained, showed latent improvement in discrimination only after the post-sleep interval. Similarly, Brawn et al. (2010) found that adult starlings trained on discrimination of segments of natural birdsong improved in performance following an interval containing sleep, whereas no improvement was observed following comparable intervals of wake state. On the other hand, Gottselig et al. (2004) measured improvement after training in participants grouped in one of four post-training conditions: no break (NB; immediate post-test administered after training), restful waking (RW; training-to-test interval spent awake in a dark, quiet room), busy waking (BW; training-to-test interval spent watching a film), and sleep (training-to-testing interval spent napping). While the authors interpret their findings as a lack of sleep-dependent effects, an alternative interpretation for their reported *p*-values is that the differences between the sleep group and the NB/RW groups respectively appear to be approaching significance, but not between NB and RW. In other words, performance is stable across time in the absence of interfering information, but task *improvement* without additional training is observed only after sleep.

Enhancement of procedural/perceptual skill has been tied to REM sleep. The REM stage in sleep has been found to be longer in duration and more frequent in sleep immediately following procedural training (e.g., Smith, 1985; Smith and Lapp, 1991). REM sleep deprivation has been linked to selective impairments in implicit learning, while leaving performance on declarative tasks intact (Smith and Kelly, 1988; Conway and Smith, 1994; Smith, 1995). Karni et al. (1994) demonstrated that selective disruptions

to REM sleep, but not to non-REM sleep, prevent improvement on a learned visual perceptual task. Taken together, these studies have been taken as evidence that REM plays a consolidation role specific to procedural learning (see Stickgold et al., 2000a,b for review). Furthermore, Schwartz et al. (2002) established using fMRI that the neural correlates of sleep-mediated improvement in monocular visual texture discrimination are localized changes in functional connectivity in the primary sensory cortex. The authors concluded that the mechanism of implicit perceptual learning involves strengthening of local connections in the sensory cortex engaged during the wake-state activity. In other words, the mechanism underlying the perceptual skill improvement appears to be supported by processes similar to latent synaptic consolidation (Vertes, 2004).

Revisiting the concept that perceptual training for speech involves learning to selectively attend to relevant acoustic features (Francis et al., 2000), the initial acquisition phase of the novel phonetic information would imply an effortful attentional shift to the features that are distinctive for the novel category. Synaptic consolidation, in increasing the efficiency of neural transmission through structural change, would thereby reduce the amount of stimulus signal necessary to elicit a response (Bramham and Messaoudi, 2005), and the corresponding effort required to selectively attend to relevant features (Atienza et al., 2004). The previous discussion on the consolidation of procedural/implicit learning during REM sleep suggests that improvement in perceptual skills reflect latent synaptic consolidation. This interpretation is consistent with Atienza et al. (2004), who reported on the post-training changes to the event-related potential (ERP) evoked during auditory discrimination. They found that while sleep deprivation did not prevent the enhancement in magnitude of mismatch negativity (MMN) response over time, an automatic shift of attention to stimuli, as measured by the P3a component, failed to develop in the sleep-deprived condition. The authors suggest that the role of sleep for auditory skill learning is to reduce the amount of effort required for attentional shift to learned stimuli, thereby increasing perceptual automaticity. Thus, in speech, the perceptual discrimination of novel speech tokens is likely to be similarly strengthened during REM sleep following perceptual training.

Consolidation during non-REM sleep and novel word form learning

In contrast to the REM sleep-mediated enhancement of perceptual automaticity, we propose that the representation of a context (such as lexical context, information about the speaker, category label, etc.) that guides the acoustic-phonetic features of interest is mediated by the structures associated with the declarative memory system. Qualitative changes to declarative memory are associated with neural activity observed during non-REM sleep. Evidence that sleep plays a role in systems consolidation of hippocampal memories comes from the observation of neural “replay” activity (Wilson and McNaughton, 1994; Skaggs and McNaughton, 1996; Ji and Wilson, 2007) in hippocampal place cells (O’Keefe and Dostrovsky, 1971) of the sleeping rat. By taking unit-cell recordings in rats trained to sleep immediately before and after spatial exploration, Wilson and McNaughton (1994) observed that hippocampal cells that fired together during

wake-state experience had a significantly higher likelihood of firing together during post-task, but not pre-task, SWS. In establishing “frames” of place cell firing sequences, Wilson and McNaughton (1994) further observed that hippocampal activity during sleep was “replays” of exact neural sequences active during the pre-sleep spatial exploration. Building on this, Ji and Wilson (2007) established that hippocampal replay events are temporally coordinated with cortical replays in the visual cortex during SWS, indicating that the hippocampal–cortical dialog during sleep reflects co-activity during wake-state experience. In addition, the same replay events were absent in the same animals sleeping ~24 h post-run, indicating that memory reactivation during sleep reflects wake-state activity immediately prior to sleep. These results are consistent with models of systems consolidation that posit an off-line transfer of episodic information from the hippocampus to long-term storage (McClelland et al., 1995), suggesting that the hippocampal–cortical temporal coordination reflects this transfer.

Following this rationale, several lines of research have worked to establish a link between non-REM sleep and systems integration of declarative memory. Considerable behavioral support for the role of sleep in the transfer and integration of information is provided by the novel word learning literature (Gaskell and Dumay, 2003; Dumay et al., 2004; Dumay and Gaskell, 2007; Tamminen and Gaskell, 2008; Davis et al., 2009; Tamminen et al., 2010). These studies share the premise that word recognition undergoes a process of activation of all possible candidate items in the mental lexicon, followed by the determination of the most likely lexical candidate for the input based on the information that unfolds over time (e.g., Cohort model, Marslen-Wilson, 1987; TRACE, McClelland and Elman, 1986). Words that are simultaneously activated are said to be in competition; task performance is therefore mediated by word frequency and phonological neighborhood density (Neighborhood activation model, Goldinger et al., 1989; Luce and Pisoni, 1998): high frequency words, and/or words occurring in a sparse phonological neighborhood, are recognized more quickly than low-frequency words in a dense neighborhood. For phonological sequences to be considered as having “word” status, they must exert similar effects of competition on existing words in the lexicon. Gaskell and Dumay (2003) trained participants on the phonological structure of novel items such as *cathedruke*, thereby introducing a competitor to a pre-existing lexical item such as *cathedral* in a sparse phonological neighborhood. Following training, the new item facilitated faster response times in lexical decision to their competitor English words (e.g., “cathedral”). However, 1 week following the training, response times in lexical decision on the pre-existing words were significantly slowed in comparison to performance on the day of training. The authors claimed that sometime during that interval, the novel words were inserted into the mental lexicon such that they became active competitors to their phonological neighbors. In a similar paradigm assessed over 24-h, Dumay and Gaskell (2007) further demonstrated that lexical competition effects emerge following the first overnight interval, but not during a comparable daytime interval. Furthermore, these lexical competition effects that emerge over the initial 24 h are still present 8 months following the initial acquisition, indicating

that the insertion of a phonological string into the mental lexicon reflects a lasting change in representational status (Tamminen and Gaskell, 2008).

Since then, overnight emergence of lexical competition has been confirmed through variations on this paradigm, such as through employing semantic training on the novel words (Dumay et al., 2004), investigating effects of lexicalization on speech segmentation (Dumay and Gaskell, 2012), or effects measured through pause detection paradigms (Dumay et al., 2004). The degree of increase in reaction time (interpreted as degree of lexical integration) has been found to correspond to duration of SWS (Tamminen et al., 2010), and furthermore the relationship of the learned novel semantic features to semantic feature neighborhood density has been shown to affect subsequent non-REM sleep architecture (Tamminen et al., 2013). Davis et al. (2009) further obtained neural correlates of the “lexicalization” effect through fMRI imaging. The authors trained novel and existing words on 2 days, and recorded participants’ neural responses to novel familiar unconsolidated, novel unfamiliar, novel consolidated, and existing words. The authors found that the initial acquisition of novel words appeared to recruit the hippocampus, and while the pattern of cortical activity for newly learned words was most similar to unfamiliar words on day 1, it came to resemble activation for existing words on day 2. The interpretation of the collective works has been that novel word forms integrate with the mental lexicon overnight through off-line interleaving of novel information (Davis and Gaskell, 2009), consistent with the *Complementary Systems Account* of learning (McClelland et al., 1995). It has been suggested that the repeated reactivation of hippocampal memories during neocortical slow oscillations result in the redistribution of episodic information to cortical networks, leading to a reorganization of abstract knowledge (Rasch and Born, 2013). Taken together with the evidence on word learning, we propose that the analogous role of systems consolidation in speech learning is to abstract and integrate novel phonetic tokens into the pre-existing phonology, such that abstract knowledge may exert top-down influence on perception.

Dissociation in sleep stages reflects synaptic vs. systems consolidation

The above discussion on sleep highlights the dissociation between non-REM sleep as preferentially responsible for declarative memory consolidation, and REM sleep for procedural memory consolidation. Before moving to our discussion on phonetic learning, it is important to first clarify that the declarative/procedural dissociation in consolidation effects is a historical convention of the wider memory literature (see Smith, 2001 for review) that current theories now consider an artifact of the task used to measure sleep-mediated effects (Diekelmann and Born, 2010). To illustrate the distinction, Cairney et al. (2011) reported that the abstraction of statistical regularities in tone sequences emerged after sleep as a function of SWS quality, even though the task tapped an implicitly learned skill. Durrant et al. (2013) examined the neural correlates of implicitly learned statistical regularities using fMRI, and found that the duration of SWS corresponded to stronger striatal, and weaker parahippocampal, responses following sleep. Taken together with the behavioral evidence, the authors claim

that abstractions for statistical regularities involve information transfer from the hippocampus to the striatum during sleep. Thus, it appears that the effects of systems consolidation (abstraction) are not limited to learning by the declarative system.

Therefore, the alternative is to consider the differential effects observed during different sleep stages as attributable to synaptic vs. systems consolidation. Diekelmann and Born (2010) propose a two-step model of active, sequential consolidation during the REM–SWS cycles of sleep to illustrate this concept. Specifically, the authors suggest that spindle activity during phase 2 non-REM sleep initiates a reorganization of the cortex, followed by the reactivation of selective wake-state experiences through coordinated hippocampal–neocortical replay in SWS that result in distributed representations. During the subsequent phase of REM, regions of the brain are proposed to undergo localized synaptic consolidation, enhancing the automaticity of signal processing within these regions. Thus, the association of REM/non-REM sleep with improvement in procedural vs. declarative tasks is thought to be reflective of the consolidation type (synaptic vs. systems).

Our predictions for how selective disruptions to specific sleep stages affects perceptual learning is built upon our assumptions for how Diekelmann and Born's (2010) two-step model applies to the two types of learning that we have highlighted as relevant in acquiring categories. It is therefore important to keep in mind that consolidation effects to be expected for phonetic learning depend on the design of the training and the task used to measure learning: specifically, which type of consolidation (synaptic, systems, or both) is likely to improve performance on a particular task. For the purposes of discussion, however, we will continue to refer to the procedural/declarative distinction to align the memory system with tasks that primarily engage each system. Specifically, perceptual learning measured by tasks that rely on acquired auditory skill, we will refer to as “procedural,” while tasks that measure the explicit recall of information, we will refer to as “declarative.” This is not to imply that we believe these memory systems to act independently in encoding phonetic information, but rather that the tasks used to assess perceptual learning differentially manipulate the relative reliance on auditory skill or explicit recall.

PREDICTIONS FOR THE ROLE OF SLEEP IN PERCEPTUAL LEARNING OF SPEECH

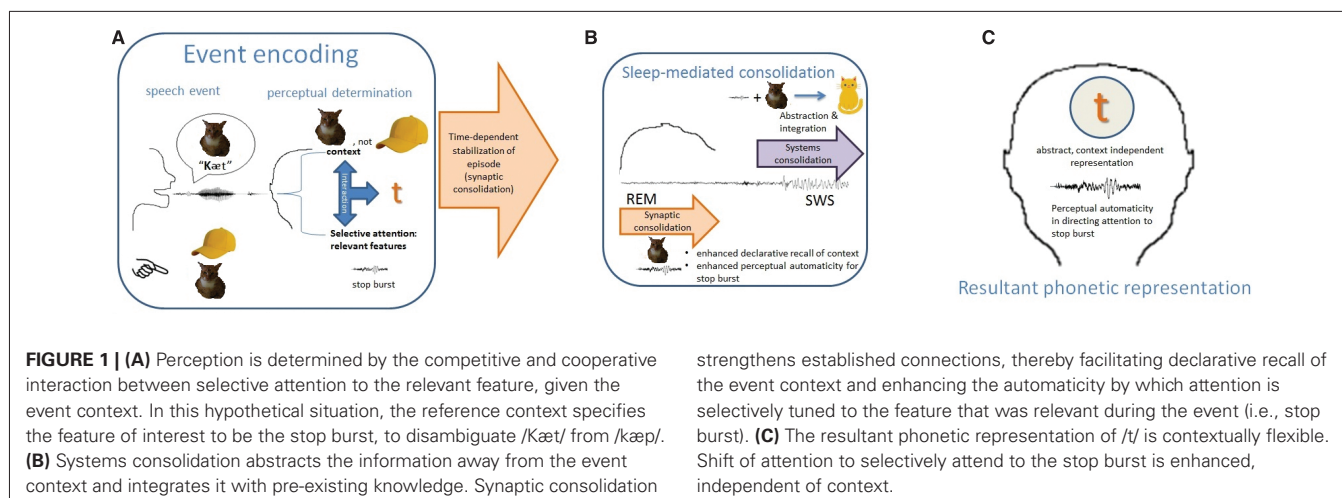
Few studies have directly investigated the role of consolidation in perceptual learning relating to speech, with some studies suggesting a role of sleep, and others not. For example, lexically guided perceptual adjustments to the categorical boundary along a /f/-/s/ continuum have been found to emerge immediately, and to remain stable over the 24-h experiment period (Eisner and McQueen, 2006), suggesting no performance change associated with sleep or the passage of time. In contrast, in a syllable-identification-in-noise task, Roth et al. (2005) showed either no gains or degraded performance up to 5 h post-training, but significant improvement 6–12 h after training regardless of whether or not the interval between training and retest contained sleep. Thus, some training-induced changes in speech perception appear not to derive benefit from sleep. In contrast, Fenn et al. (2003) observed a different pattern, in which changes to

performance over time depended on when the training took place in relation to sleep. Their task was to identify words presented in synthetic speech; no word was used more than once, such that training required that listeners generalize the mapping of synthetic speech to pre-existing phonology. For those trained at 9 p.m., the overnight between-session interval appeared to have a stabilizing effect against degradation in performance. In those trained at 9 a.m., performance appeared to degrade during the day and then was restored overnight. These studies therefore suggest different roles for consolidation over time and sleep depending on the perceptual measure used.

This inconsistency may be potentially resolved by considering the phonetic information contributing to the pre-existing phonology by the training task. In both the studies by Eisner and McQueen (2006) and Roth et al. (2005), the perceptual adjustments are made within pre-established category boundaries. In other words, neither study requires any reorganization of the phonological system, or any new information to be added to inform pre-existing category structure. In contrast, the mapping task in Fenn et al. (2003) required a systemic re-definition of representations to include synthetic speech as acceptable exemplars. Thus, the sleep-mediated changes to behavioral performance may be more salient in tasks that benefit from the addition of phonetic information to, or systemic reorganization of, the established phonology. This interpretation is supported by a recent study by Fenn et al. (2013), comparing the sleep-mediated effects of rote vs. generalized training of mapping synthetic speech to native phonology. They found that the rote-trained group, trained on a closed set of the same 20 words, improved after sleep in the trained words only, but the generalization-trained group, trained on unique stimuli for every trial, demonstrated significant improvement on novel words after sleep.

Applying the above insights, we can expect several sleep-related changes to performance on tasks that measure the encoding of novel phonetic information (see **Figure 1**). In the procedural memory system, synaptic consolidation may promote automaticity in the implicitly acquired shift of attention to selective features, leading to enhanced performance on perceptual discrimination of novel speech sounds. In addition, synaptic consolidation in the declarative memory system may lead to improvement in the explicit recall of the category label, resulting in faster identification. Systems consolidation on the other hand, would promote the abstraction and integration/reorganization of novel phonetic information within the pre-existing mental phonology, such that the identification of novel phonetic tokens are no longer bound to episodic information such as the specific talker encountered in the learning event. We would expect to observe this in the generalization of novel phonetic information to the identification of the newly learned phonetic items spoken by a different talker, or occurring in a different phonetic environment, with respect to tokens used during learning.

To reiterate, this account predicts that the two behavioral tasks often used to measure phonetic learning, discrimination and identification, rely in different degrees on information encoded by two different memory systems. Improvement in perceptual discrimination may rely on learned selective attention to relevant



features, which, as an implicitly acquired perceptual skill, is primarily mediated by the procedural memory system. The task of identification, as an explicit recall of a category label, is mediated by the declarative memory system. This claim is supported by differences recently observed in our lab on the sleep-mediated effects on task performance between discrimination and identification tasks (Earle and Myers, 2014). Specifically, we found that passive exposure to conflicting phonetic information following the perceptual training of non-native consonants prevented the overnight improvement of perceptual discrimination, consistent with retroactive interference effects observed in the sleep-mediated consolidation of procedural learning in other domains (e.g., Walker et al., 2003). The interference condition, however, did not prevent improved identification of the trained non-native phonetic tokens over time, suggesting that the information recalled for identification is not susceptible to the same interference effects that are detrimental to perceptual discrimination. This robustness of performance on the identification task relative to the discrimination task may be due to the contribution of episodic information through the engagement of the declarative memory system during the identification task. This interpretation further predicts that the generalization of the trained speech sounds across different talkers and different acoustic contexts will be more likely to be observed in a task involving explicit recall, such as identification, over an application of implicit skill, such as discrimination.

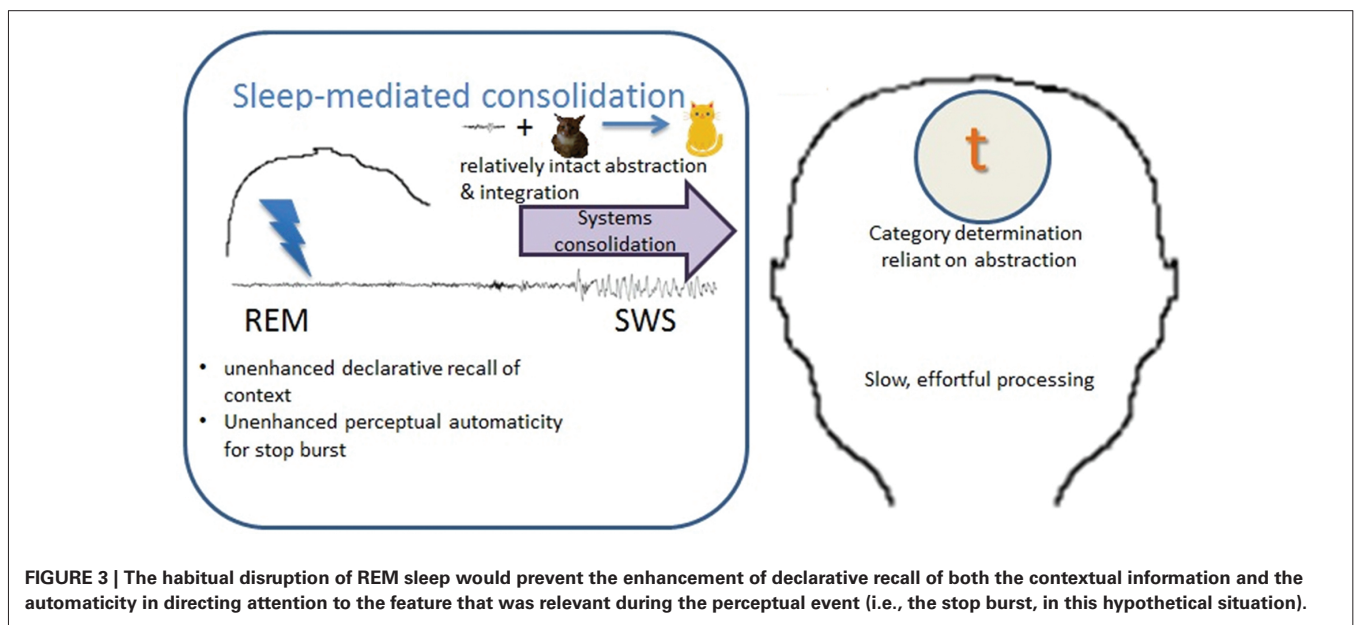
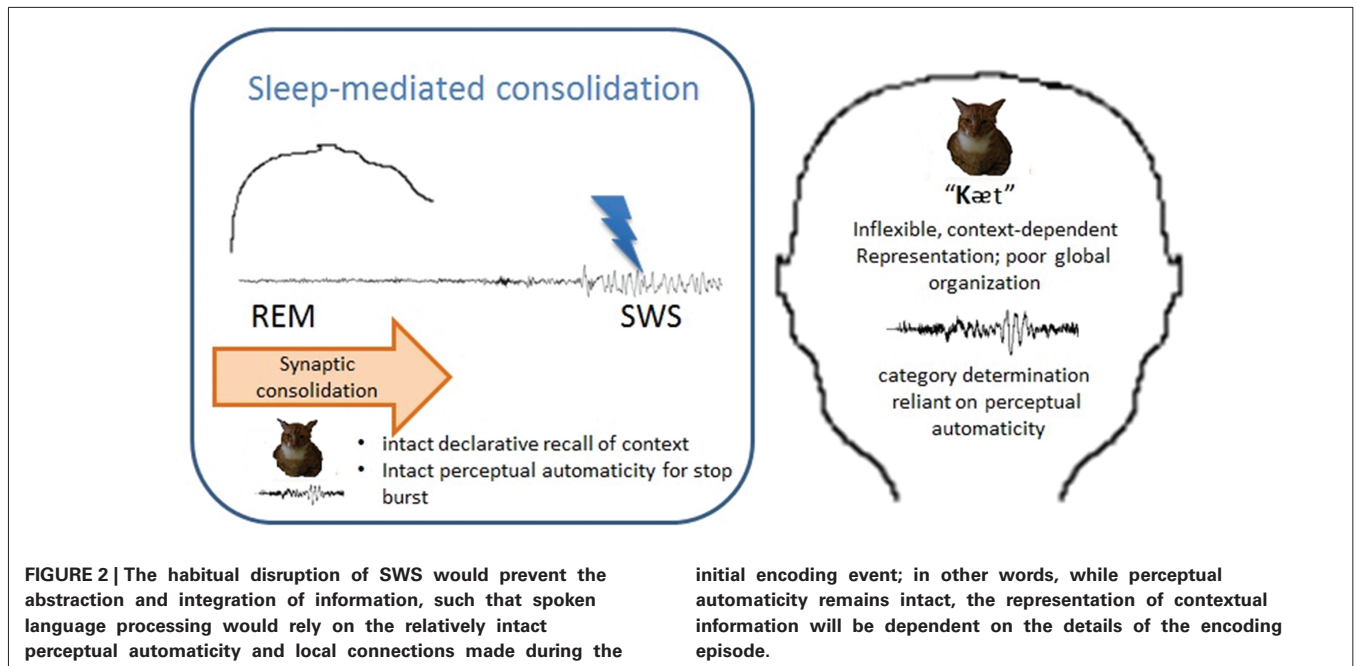
The task dissociation furthermore reflects the two sources of information that are highlighted above as necessary for efficient online spoken language processing: the implicitly acquired perceptual automaticity in directing attention selective to relevant features in the signal, and the context in which features are evaluated as similar/dissimilar. Learned selective attention, as might be preferentially improved through synaptic consolidation, underlies changes to behavioral performance in discrimination. The encoding of the context that specifies category boundaries, we propose is bound to the declarative memory system; the category boundaries are initially tied to the episodic capture of the experience, followed by an off-line abstraction away from the episode-specific details.

PREDICTED CONSEQUENCES TO HABITUAL DISRUPTION IN SLEEP

The current discussion further yields specific predictions for how habitual sleep disruptions may affect the encoding of phonetic information. This topic is motivated by reports in the developmentally language-impaired population of paroxysmal activity during sleep (Duvelleroy-Hommet et al., 1995; Picard et al., 1998), along with abnormal development in brain structures associated with the procedural memory system (Procedural Deficit Hypothesis; Ullman and Pierpont, 2005). Language impairment in autism has furthermore been linked to epileptiform activity observed during sleep (Patzold et al., 1998; Ballaban-Gil and Tuchman, 2000); furthermore, the onset of language regression in Landau-Kleffner is associated with epileptic activity observed during SWS (Massa et al., 2000; McVicar and Shinnar, 2004). As epileptic activity implies potential structural damage, it is not the intent of the current review to argue that the language impairments in these populations are directly caused by disruptions to sleep. Rather, the following discussion will outline the potential consequences to phonological stability that we might expect from habitual disruptions to the discrete structures or processes involved in building phonological representations.

First, structural abnormality in the procedural memory system leads to a compromised ability to build speed/automaticity in selectively attending to features that are relevant in the acoustic signal. This may lead to slowed and/or effortful signal processing, increasing the cognitive demands placed on online comprehension of spoken language (see Figure 2). In contrast, declarative systems mediate explicit recall of episodic and semantic memory; thus, structural/connectivity differences in the declarative system structures, may result in deficits in the encoding/recall of episodic details of the signal such as talker features and in the abstraction of these details that underlie category structure.

In considering a functional abnormality, we might expect that selective habitual disruption to REM sleep may prevent both the enhancement of perceptual automaticity and of episodic recall. A selective habitual disruption to SWS, on the other hand, may



prevent the formation of abstract categories and the transfer of information between systems, leading to poor global organization of categories with relatively intact local connections. Behaviorally, this may manifest in sensitivity to non-distinctive cues in the signal, without contextual or top-down knowledge guiding efficient organization of the input (see **Figure 3**). Examples of speech-specific difficulties for such a system may be in the comprehension of a degraded signal, such as over the phone, or in adverse conditions, such as in background noise. Furthermore, such systems may be less adaptable to accented or novel talkers. Finally, a structural or functional abnormality in

either system can lead to impoverished representational quality overall, in compromising the cohesion between the two sources of input.

In conclusion, there is much work to be done on the role of sleep in phonetic learning. There are several outstanding issues highlighted here regarding the contribution of discrete memory encoding processes on speech perception. In defining the time course of typical encoding of novel phonetic tokens, we may begin to disentangle if/how these encoding processes are compromised in those with language symptoms concurrent to developmental disability.

ACKNOWLEDGMENTS

This work was supported by NIH NIDCD grants R03 DC009495 and R01 DC013064 to Emily B. Myers, and NIH NICHD grant P01 HD001994 (Rueckl, PI). The content is the responsibility of the authors and does not necessarily represent official views of the NIH, NIDCD, or NICHD. We further thank Etan Markus, Rachel Theodore, Tammie Spaulding, and Erika Skoe for their helpful advice and feedback on a previous document that served as the foundation for the current paper.

REFERENCES

- Allen, J. S., and Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *J. Acoust. Soc. Am.* 115, 3171. doi: 10.1121/1.1701898
- Athos, E. A., Levinson, B., Kistler, A., Zemansky, J., Bostrom, A., Freimer, N., et al. (2007). Dichotomy and perceptual distortions in absolute pitch ability. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14795–14800. doi: 10.1073/pnas.0703868104
- Atienza, M., Cantero, J. L., and Stickgold, R. (2004). Posttraining sleep enhances automaticity in perceptual discrimination. *J. Cogn. Neurosci.* 16, 53–64. doi: 10.1162/08989290432275557
- Ballaban-Gil, K., and Tuchman, R. (2000). Epilepsy and epileptiform EEG: association with autism and language disorders. *Ment. Retard. Dev. Disabil. Res. Rev.* 6, 300–308. doi: 10.1002/1098-2779(2000)6:4<300::AID-MRDD9>3.0.CO;2-R
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: a perceptual assimilation model. *Dev. Speech Percept.* 167, 224.
- Bliss, T. V., and Collingridge, G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* 361, 31–39. doi: 10.1038/361031a0
- Bliss, T. V., and Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol.* 232, 331–356.
- Bramham, C. R., and Messaoudi, E. (2005). BDNF function in adult synaptic plasticity: the synaptic consolidation hypothesis. *Prog. Neurobiol.* 76, 99–125. doi: 10.1016/j.pneurobio.2005.06.003
- Brashers-Krug, T., Shadmehr, R., and Bizzi, E. (1996). Consolidation in human motor memory. *Nature* 382, 252–255. doi: 10.1038/382252a0
- Brawn, T. P., Nusbaum, H. C., and Margoliash, D. (2010). Sleep-dependent consolidation of auditory discrimination learning in adult starlings. *J. Neurosci.* 30, 609–613. doi: 10.1523/JNEUROSCI.4237-09.2010
- Cairney, S. A., Durrant, S. J., Musgrove, H., and Lewis, P. A. (2011). Sleep and environmental context: interactive effects for memory. *Exp. Brain Res.* 214, 83–92. doi: 10.1007/s00221-011-2808-7
- Cheour, M., Martynova, O., Näätänen, R., Erkkola, R., Sillanpää, M., Kero, P., et al. (2002). Speech sounds learned by sleeping newborns. *Nature* 415, 599–600. doi: 10.1038/415599b
- Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.
- Conway, J., and Smith, C. (1994). REM sleep and learning in humans: a sensitivity to specific types of learning tasks. *J. Sleep Res.* 3(Suppl. 1), 48.
- Davis, M. H., Di Betta, A. M., Macdonald, M. J., and Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *J. Cogn. Neurosci.* 21, 803–820. doi: 10.1162/jocn.2009.21059
- Davis, M. H., and Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3773–3800. doi: 10.1098/rstb.2009.0111
- Diekelmann, S., and Born, J. (2010). The memory function of sleep. *Nat. Rev. Neurosci.* 11, 114–126. doi: 10.1038/nrn2762
- Doyon, J. (1997). Skill learning. *Int. Rev. Neurobiol.* 41, 273–294. doi: 10.1016/S0074-7742(08)60356-6
- Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol.* 55, 51–86. doi: 10.1146/annurev.psych.55.090902.142050
- Dumay, N., and Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychol. Sci.* 18, 35–39. doi: 10.1111/j.1467-9280.2007.01845.x
- Dumay, N., and Gaskell, G. M. (2012). Overnight lexical consolidation revealed by speech segmentation. *Cognition* 123, 119–132. doi: 10.1016/j.cognition.2011.12.009
- Dumay, N., Gaskell, M. G., and Feng, X. (2004). “A day in the life of a spoken word,” in *Proceedings of the Twenty-sixth Annual Conference of the Cognitive Science Society* (Mahwah, NJ: Erlbaum), 339–344.
- Durrant, S. J., Cairney, S. A., and Lewis, P. A. (2013). Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cereb. Cortex* 23, 2467–2478. doi: 10.1093/cercor/bhs244
- Duvelleroy-Hommet, C., Gillet, P., Billard, C., Loisel, M. L., Barthez, M. A., Santini, J. J., et al. (1995). Study of unilateral hemisphere performance in children with developmental dysphasia. *Neuropsychologia* 33, 823–834. doi: 10.1016/0028-3932(95)00020-4
- Earle, S., and Myers, E. B. (2014). Native language interference on the overnight consolidation of a learned nonnative contrast. *J. Acoust. Soc. Am.* 135, 2352–2352. doi: 10.1121/1.4877733
- Eisner, F., and McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Percept. Psychophys.* 67, 224–238. doi: 10.3758/BF03206487
- Eisner, F., and McQueen, J. M. (2006). Perceptual learning in speech: stability over time. *J. Acoust. Soc. Am.* 119, 1950. doi: 10.1121/1.2178721
- Feldman, N. H., Griffiths, T. L., and Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychol. Rev.* 116, 752. doi: 10.1037/a0017196
- Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (2013). Sleep restores loss of generalized but not rote learning of synthetic speech. *Cognition* 128, 280–286. doi: 10.1016/j.cognition.2013.04.007
- Fenn, K. M., Nusbaum, H. C., and Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature* 425, 614–616. doi: 10.1038/nature01951
- Fischer, S., Nitschke, M. F., Melchert, U. H., Erdmann, C., and Born, J. (2005). Motor memory consolidation in sleep shapes more effective neuronal representations. *J. Neurosci.* 25, 11248–11255. doi: 10.1523/JNEUROSCI.1743-05.2005
- Francis, A. L., Baldwin, K., and Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Percept. Psychophys.* 62, 1668–1680. doi: 10.3758/BF03212164
- Francis, A. L., and Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 349. doi: 10.1037/0096-1523.28.2.349
- Gaab, N., Paetzold, M., Becker, M., Walker, M. P., and Schlaug, G. (2004). The influence of sleep on auditory learning: a behavioral study. *Neuroreport* 15, 731–734. doi: 10.1097/00001756-200403220-00032
- Gaskell, M. G., and Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition* 89, 105–132. doi: 10.1016/S0010-0277(03)00070-2
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105, 251. doi: 10.1037/0033-295X.105.2.251
- Goldinger, S. D., Luce, P. A., and Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: effects of competition and inhibition. *J. Mem. Lang.* 28, 501–518. doi: 10.1016/0749-596X(89)90009-0
- Goldstone, R. L. (1994). The role of similarity in categorization: providing a groundwork. *Cognition* 52, 125–157. doi: 10.1016/0010-0277(94)90065-5
- Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychol. Sci.* 6, 298–304. doi: 10.1111/j.1467-9280.1995.tb00514.x
- Golestani, N., and Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain Lang.* 109, 55–67. doi: 10.1016/j.bandl.2008.01.005
- Gómez, R. L., Bootzin, R. R., and Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychol. Sci.* 17, 670–674. doi: 10.1111/j.1467-9280.2006.01764.x
- Gomez, R. L., Newman-Smith, K. C., Breslin, J. H., and Bootzin, R. R. (2011). Learning, memory, and sleep in children. *Sleep Med. Clin.* 6, 45–57. doi: 10.1016/j.jsmc.2010.12.002
- Gottselig, J. M., Hofer-Tinguely, G., Borbely, A. A., Regel, S. J., Landolt, H. P., Retey, J. V., et al. (2004). Sleep and rest facilitate auditory learning. *Neuroscience* 127, 557–561. doi: 10.1016/j.neuroscience.2004.05.053
- Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., and Tourville, J. A. (2004). Representation of sound categories in auditory cortical maps. *J. Speech Lang. Hear Res.* 47, 46. doi: 10.1044/1092-4388(2004)005
- Hupbach, A., Gomez, R. L., Bootzin, R. R., and Nadel, L. (2009). Nap-dependent learning in infants. *Dev. Sci.* 12, 1007–1012. doi: 10.1111/j.1467-7687.2009.00837.x

- Iverson, P., and Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *J. Acoust. Soc. Am.* 97, 553–562. doi: 10.1121/1.412280
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47–B57. doi: 10.1016/S0010-0277(02)00198-1
- Ji, D., and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* 10, 100–107. doi: 10.1038/nn1825
- Joanisse, M. F., and Seidenberg, M. S. (2003). Phonology and syntax in specific language impairment: evidence from a connectionist model. *Brain Lang.* 86, 40–56. doi: 10.1016/S0093-934X(02)00533-3
- Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J., and Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science* 265, 679–682. doi: 10.1126/science.8036518
- Kraljic, T., and Samuel, A. G. (2005). Perceptual learning for speech: is there a return to normal? *Cogn. Psychol.* 51, 141–178. doi: 10.1016/j.cogpsych.2005.05.001
- Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychol. Rev.* 99, 22. doi: 10.1037/0033-295X.99.1.22
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Percept. Psychophys.* 50, 93–107. doi: 10.3758/BF03212211
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606–608. doi: 10.1126/science.1736364
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358. doi: 10.1037/h0044417
- Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear Hear.* 19, 1.
- Magnuson, J. S., and Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 391. doi: 10.1037/0096-1523.33.2.391
- Margoliash, D., and Fenn, K. M. (2008). “Sleep and memory consolidation in audition,” in *The Senses: A Comprehensive Reference*, Vol. 3, *Audition*, eds A. I. Basbaum, A. Kaneko, G. M. Shepherd, and G. Westheimer (San Diego: Academic Press), 895–912.
- Marr, D. (1971). Simple memory: a theory for archicortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 23–81. doi: 10.1098/rstb.1971.0078
- Marshall, L., and Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends Cogn. Sci.* 11, 442–450. doi: 10.1016/j.tics.2007.09.001
- Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition* 25, 71–102. doi: 10.1016/0010-0277(87)90005-9
- Massa, R., de Saint-Martin, A., Hirsch, E., Marescaux, C., Motte, J., Seegmuller, C., et al. (2000). Landau-Kleffner syndrome: sleep EEG characteristics at onset. *Clin. Neurophysiol.* 111, S87–S93. doi: 10.1016/S1388-2457(00)00407-7
- Maye, J., Weiss, D. J., and Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Dev. Sci.* 11, 122–134. doi: 10.1111/j.1467-7687.2007.00653.x
- Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111. doi: 10.1016/S0010-0277(01)00157-3
- McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0
- McClelland, J. L., McNaughton, B. L., and 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. *Psychol. Rev.* 102, 419. doi: 10.1037/0033-295X.102.3.419
- McNaughton, B. L., and Morris, R. G. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.* 10, 408–415. doi: 10.1016/0166-2236(87)90011-7
- McVicar, K. A., and Shinnar, S. (2004). Landau-Kleffner syndrome, electrical status epilepticus in slow wave sleep, and language regression in children. *Ment. Retard. Dev. Disabil. Res. Rev.* 10, 144–149. doi: 10.1002/mrdd.20028
- Medin, D. L., and Schaffer, M. M. (1978). Context theory of classification learning. *Psychol. Rev.* 85, 207. doi: 10.1037/0033-295X.85.3.207
- Müller, G. E., and Pilzecker, A. (1900). *Experimentelle Beiträge zur Lehre vom Gedächtnis*. Leipzig: Verlag von Johann Ambrosius Barth.
- Nation, K., and Snowling, M. J. (1998). Individual differences in contextual facilitation: evidence from dyslexia and poor reading comprehension. *Child Dev.* 69, 996–1011. doi: 10.1111/j.1467-8624.1998.tb06157.x
- Nissen, M. J., and Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cogn. Psychol.* 19, 1–32. doi: 10.1016/0010-0285(87)90002-8
- Norris, D., McQueen, J. M., and Cutler, A. (2003). Perceptual learning in speech. *Cogn. Psychol.* 47, 204–238. doi: 10.1016/S0010-0285(03)00006-9
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *J. Exp. Psychol. Gen.* 115, 39. doi: 10.1037/0096-3445.115.1.39
- Nusbaum, H., and Magnuson, J. (1997). “Talker normalization: phonetic constancy as a cognitive process,” *Talker Variability in Speech Processing*, eds K. Johnson and J. W. Mullennix (San Diego, CA: Academic), 109–132.
- Nygaard, L. C., and Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Percept. Psychophys.* 60, 355–376. doi: 10.3758/BF03206860
- Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychol. Sci.* 5, 42–46. doi: 10.1111/j.1467-9280.1994.tb00612.x
- O’Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* 34, 171–175. doi: 10.1016/0006-8993(71)90358-1
- O’Keefe, J., and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*, Vol. 3. Oxford: Clarendon Press.
- O’Reilly, R. C., and Rudy, J. W. (2001). Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychol. Rev.* 108, 311. doi: 10.1037/0033-295X.108.2.311
- Patzold, L. M., Richdale, A. L., and Tonge, B. J. (1998). An investigation into sleep characteristics of children with autism and Asperger’s disorder. *J. Paediatr. Child Health* 34, 528–533. doi: 10.1046/j.1440-1754.1998.00291.x
- Picard, A., Heraut, F. C., Bouskraoui, M., Lemoine, M., Lacert, P., and Delattre, J. (1998). Sleep EEG and developmental dysphasia. *Dev. Med. Child Neurol.* 40, 595–599. doi: 10.1111/j.1469-8749.1998.tb15424.x
- Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Lang. Speech* 46, 115–154. doi: 10.1177/00238309030460020501
- Pisoni, D. B., and Levi, S. V. (2007). “Some observations on representations and representational specificity in speech perception and spoken word recognition,” in *The Oxford Handbook of Psycholinguistics*, ed. M. G. Gaskell (Oxford: Oxford University Press), 3–18.
- Rasch, B., and Born, J. (2013). About sleep’s role in memory. *Physiol. Rev.* 93, 681–766. doi: 10.1152/physrev.00032.2012
- Rasch, B., Büchel, C., Gais, S., and Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science* 315, 1426–1429. doi: 10.1126/science.1138581
- Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science* 212, 947–949. doi: 10.1126/science.7233191
- Repp, B. H., and Liberman, A. M. (1987). “Phonetic category boundaries are flexible,” in *Categorical Perception: The Groundwork of Cognition*, ed. S. Harnad (New York: Cambridge University Press), 89–112.
- Roth, D. A. E., Kishon-Rabin, L., Hildesheimer, M., and Karni, A. (2005). A latent consolidation phase in auditory identification learning: time in the awake state is sufficient. *Learn. Mem.* 12, 159–164. doi: 10.1101/87505
- Rumelhart, D. E. (1990). “Brain style computation: learning and generalization,” in *An Introduction to Neural and Electronic Networks*, eds S. F. Zornetzer, J. L. Davis, and C. Lau (San Diego, CA: Academic), 405–420.
- Saletin, J. M., Goldstein, A. N., and Walker, M. P. (2011). The role of sleep in directed forgetting and remembering of human memories. *Cereb. Cortex* 21, 2534–2541. doi: 10.1093/cercor/bhr034
- Schwartz, S., Maquet, P., and Frith, C. (2002). Neural correlates of perceptual learning: a functional MRI study of visual texture discrimination. *Proc. Natl. Acad. Sci. U.S.A.* 99, 17137–17142. doi: 10.1073/pnas.242414599
- Skaggs, W. E., and McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* 271, 1870–1873. doi: 10.1126/science.271.5257.1870

- Smith, C. (1985). Sleep states and learning: a review of the animal literature. *Neurosci. Biobehav. Rev.* 9, 157–168. doi: 10.1016/0149-7634(85)90042-9
- Smith, C. (1995). Sleep states and memory processes. *Behav. Brain Res.* 69, 137–145.
- Smith, C. (2001). Sleep states and memory processes in humans: procedural versus declarative memory systems. *Sleep Med. Rev.* 5, 491–506. doi: 10.1053/smr.2001.0164
- Smith, C., and Kelly, G. (1988). Paradoxical sleep deprivation applied two days after end of training retards learning. *Physiol. Behav.* 43, 213–216. doi: 10.1016/0031-9384(88)90240-5
- Smith, C., and Lapp, L. (1991). Increases in number of REMS and REM density in humans following an intensive learning period. *Sleep* 14, 325–330.
- Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* 99, 195. doi: 10.1037/0033-295X.99.2.195
- Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature* 437, 1272–1278. doi: 10.1038/nature04286
- Stickgold, R., James, L., and Hobson, J. A. (2000a). Visual discrimination learning requires sleep after training. *Nat. Neurosci.* 3, 1237–1238. doi: 10.1038/81756
- Stickgold, R., Whidbee, D., Schirmer, B., Patel, V., and Hobson, J. A. (2000b). Visual discrimination task improvement: a multi-step process occurring during sleep. *J. Cogn. Neurosci.* 12, 246–254. doi: 10.1162/089892900562075
- Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *J. Lang. Soc. Psychol.* 18, 86–100. doi: 10.1177/0261927X99018001006
- Swan, K., and Myers, E. (2013). Category labels induce boundary-dependent perceptual warping in learned speech categories. *Second Lang. Res.* 29, 391–411. doi: 10.1177/0267658313491763
- Tamminen, J., and Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical competition effects. *Q. J. Exp. Psychol.* 61, 361–371. doi: 10.1080/17470210701634545
- Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., and Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *J. Neurosci.* 30, 14356–14360. doi: 10.1523/JNEUROSCI.3028-10.2010
- Tamminen, J., Ralph, M. A. L., and Lewis, P. A. (2013). The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *J. Neurosci.* 33, 15376–15381. doi: 10.1523/JNEUROSCI.5093-12.2013
- Tulving, E. (1985). How many memory systems are there? *Am. Psychol.* 40, 385. doi: 10.1037/0003-066X.40.4.385
- Ullman, M. T., and Pierpont, E. I. (2005). Specific language impairment is not specific to language: the procedural deficit hypothesis. *Cortex* 41, 399–433. doi: 10.1016/S0010-9452(08)70276-4
- Vertes, R. P. (2004). Memory consolidation in sleep: dream or reality. *Neuron* 44, 135–148. doi: 10.1016/j.neuron.2004.08.034
- Walker, M. P., Brakefield, T., Hobson, J. A., and Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature* 425, 616–620. doi: 10.1038/nature01930
- Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., and Stickgold, R. (2002). Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron* 35, 205–211. doi: 10.1016/S0896-6273(02)00746-8
- Walker, M. P., and Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron* 44, 121–133. doi: 10.1016/j.neuron.2004.08.031
- Wilson, M. A., and McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science* 265, 676–679. doi: 10.1126/science.8036517

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 June 2014; accepted: 02 October 2014; published online: 28 October 2014.

Citation: Earle FS and Myers EB (2014) Building phonetic categories: an argument for the role of sleep. *Front. Psychol.* 5:1192. doi: 10.3389/fpsyg.2014.01192

This article was submitted to Language Sciences, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Earle and Myers. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.