JSLHR

Research Article

Sleep Promotes Phonological Learning in Children Across Language and Autism Spectra

Victoria C. P. Knowland,[a] Fay Fletcher,[a] Lisa-Marie Henderson,[a] Sarah Walker,[a] Courtenay F. Norbury,[b] and M. Gareth Gaskell[a]

Purpose: Establishing stable and flexible phonological representations is a key component of language development and one which is thought to vary across children with neurodevelopmental disorders affecting language acquisition. Sleep is understood to support the learning and generalization of new phonological mappings in adults, but this remains to be examined in children. This study therefore explored the time course of phonological learning in childhood and how it varies by structural language and autism symptomatology.
Method: Seventy-seven 7- to 13-year-old children, 30 with high autism symptomatology, were included in the study; structural language ability varied across the sample. Children learned new phonological mappings based on synthesized speech tokens in the morning; performance was then charted via repetition (without feedback) over 24 hr and followed up 4 weeks later. On the night following learning, children's sleep was monitored with polysomnography.
Results: A period of sleep but not wake was associated with improvement on the phonological learning task in childhood. Sleep was associated with improved performance for both trained items and novel items. Structural language ability predicted overall task performance, though language ability did not predict degree of change from one session to the next. By contrast, autism symptomatology did not explain task performance. With respect to sleep architecture, rapid eye movement features were associated with greater phonological generalization.
Conclusions: Children's sleep was associated with improvement in performance on both trained and novel items. Phonological generalization was associated with brain activity during rapid eye movement sleep. This study furthers our understanding of individual differences in the acquisition of new phonological mappings and the role of sleep in this process over childhood.
Supplemental Material: https://doi.org/10.23641/asha.11126732

Phonological representations are units of knowledge in long-term memory describing the sounds that make up words (Stackhouse & Wells, 1997). The representation of speech sounds is perhaps the most fundamental element of language perception and production, forming the basis of phonological and orthographic word form and being critical to the processing of morphosyntactic structure (Joanisse & Seidenberg, 2003). Speech sounds are understood to be represented categorically, with greater perceptual distance between points of equivalent acoustic distance when those points cross a category boundary compared to when they fall within a category (Chang et al., 2010; Liberman, Harris, Hoffman, & Griffith, 1957). The establishment and use of phonological representations demand a fine balance. Phonological categories must be sufficiently well defined to support speech perception by allowing the recognition of distinct phonemes, yet the system must also be flexible enough to allow an adaptive response to the changing linguistic environment, allowing new phoneme categories to be formed and established categories to be adjusted when listeners face inter- and intraspeaker variation (see Earle & Myers, 2014). When listening to someone with an unfamiliar accent, for example, a listener "tunes in" over time, expanding and adapting the category requirements for different phonological categories and generalizing learned adaptations to new contexts as they are heard (e.g., Whitteman, Bardhan, Weber, & McQueen, 2015).

[a]Department of Psychology, University of York, United Kingdom
[b]Division of Psychology and Language Sciences, UCL, London, United Kingdom

Correspondence to Victoria C. P. Knowland: Victoria.knowland@york.ac.uk

Editor-in-Chief: Bharath Chandrasekaran
Editor: Chao-Yang Lee

Received March 4, 2019
Revision received June 21, 2019
Accepted August 7, 2019
https://doi.org/10.1044/2019_JSLHR-S-19-0098

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Children are thought to vary in the specificity with which speech sounds are represented. Although not ubiquitously (Coady, Kluender, & Evans, 2005), many studies have found that children with developmental disorders of oral language perception and production have noisy or poorly specified phonological representations. For example, children with developmental language disorder (DLD) show weak discrimination of speech sounds across boundary points (Robertson, Joanisse, Desroches, & Ng, 2009; Stark & Heinz, 1996; Vanderwalle, Boets, Ghesquière, & Zink, 2012), struggle to perceive speech in noise (Knowland, Evans, Snell, & Rosen, 2016; Vanderwalle et al., 2012; Ziegler, Pech-Georgel, George, Alario, & Lorenzi, 2005; Ziegler, Pech-Georgel, George, & Lorenzi, 2011), and exhibit poor lexical decision making when nonwords are phonologically similar to existing words (Maillart, Schelstraete, & Hupet, 2004). Poorly specified, noisy phonological representations have been proposed to underlie morphological difficulties, which are a hallmark of DLD (Joanisse, 2004; Joanisse & Seidenberg, 2003). The precision of phonological representations also varies between typically developing children within an age group (Anthony et al., 2010) and across developmental time (Hazan & Barett, 2000). We might expect that the specificity with which speech sounds are represented is therefore associated with structural language ability.

Individuals with autism spectrum disorders (ASDs) do not tend to show atypicalities in the perceptual categorization of speech sounds (Constantino et al., 2007; Stewart, Petrou, & Ota, 2018; although see Wang, Wang, Fan, Huang, & Zhang, 2017, for a study on the categorization of lexical tone). However, there is some evidence for reduced phonetic generalization in this population (Järvinen-Pasley, Wallace, Ramus, Happé, & Heaton, 2008), that is, a reduced ability to apply learning about speech sounds from one situation to another. Compared to controls, individuals with ASD demonstrate less generalization across speech sounds when trained to imitate words with a modified phonetic feature (increased voice onset time; Mielke, Nielsen, & Magloughlin, 2013). The ability to generalize is critical to the flexibility of the phonological system. Generalization allows the listener to adapt to interspeaker variation, such as a previously unencountered accent, or intraspeaker variation, such as phonetic variability as speakers encounter different acoustic backgrounds (see Kleinschmidt & Jaeger, 2015). Generalization also allows listeners to transfer that learning to new words from the accented speaker or new speakers in the same noisy conditions.

In this study, we assessed the time course of phonological learning in a task that evaluated performance on both trained items and novel items (assumed to rely more heavily on generalization). Our aim was to probe the nature of the phonological system in children who were thought to vary in the stability and flexibility of their phonological representations, that is, in children who vary in structural language ability and autism symptomatology. Notably, many children with an ASD also show broad weaknesses in structural language ability (see Williams, Botting, & Boucher, 2008, for a review), and many children with poor structural language show symptoms of ASD. By sampling across spectra of ability, here, we were able to assess the implications of poor structural language and autism symptomatology in the same children.

Phonological Learning Is Supported by Sleep

Perceptual learning related to speech sounds can be rapid, but the establishment of stable representations takes time and, as some evidence suggests, sleep. In typically developing young adults, Fenn and colleagues (Fenn, Margoliash, & Nusbaum, 2013; Fenn, Nusbaum, & Margoliash, 2003) have demonstrated that sleep promotes the generalization of perceptual learning when listening to synthesized "text-to-speech" tokens. Fenn et al. (2013) tested adults on their ability to understand single synthesized words at four time points: before and after a perceptual training session at 9 a.m. and then 12 and 24 hr later. Two groups of participants were included. The generalization-trained group was trained by listening to 300 different synthetic words while being presented simultaneously with the written words. After 30 min of training, participants showed around a 20% improvement in their ability to type out previously unheard synthesized words; this training effect had decayed 12 waking hours later but was restored to posttraining levels after sleep. A separate rote-trained group was trained on a limited set of 20 items, each repeated 15 times over the course of training. This group showed improvement when tested on the same items they were exposed to during training, but not when tested on novel items, even after sleep. Fenn et al. suggest that sleep therefore supports the generalization of phonological remapping in adults. The importance of sleep for generalization is supported by previous work from this same group, showing that phonological generalization after training on synthesized speech is restored to posttraining levels after sleep in adults trained in the evening, but not after an equivalent period of wake in adults trained in the morning (Fenn et al., 2003).

Other work on the time course of phonological generalization has considered adult listeners' ability to generalize between foreign-accented speakers after, for example, adapting to initially ambiguous speech sounds during lexically guided learning. After perceptual adaptation to one accented speaker, immediate generalization to others has been demonstrated when acoustic properties relevant to perceptual contrasts of interest are similar between speakers (Xie, Earle, & Myers, 2016). When speakers differ phonetically, however, listeners' ability to generalize perceptual learning across speakers was observed only after a period of sleep and not an equivalent wake period (Earle & Myers, 2015; Xie, Earle, & Myers, 2018). This suggests that sleep supports the abstraction of higher level category-relevant information from speech. Not all studies of phonological learning have shown a privileged role for sleep, however; for example, perceptual adjustments such as adapting to altered voice onset times (Collet et al., 2012) or place of articulation (Eisner & McQueen, 2006) for a single phonological contrast are not usually shown to be sleep dependent in adults. This suggests that relatively straightforward modifications to the phonological system or to individual phonemic features are stable enough to not require overnight consolidation (see Earle & Meyers, 2014). The exact conditions under which sleep may support phonetic learning are not clear though, as even the identification and discrimination of a single nonnative contrast has been shown to be sleep dependent (Earle, Landi, & Myers, 2017).

Recently, Earle, Landi, and Myers (2018) trained and immediately tested adults with and without DLD on nonnative speech sounds in the evening and then retested performance the following morning. Adults with DLD showed comparable gains to age-matched controls during training but did not show overnight consolidation on the identification and discrimination of the newly learned sounds. The literature therefore suggests that sleep promotes complex phonological learning in typically developing adults but may not provide the same degree of support for individuals who show relatively poor structural language skills.

The mechanisms by which sleep supports human memory in general are beginning to come into focus. Sleep can be broadly divided into rapid eye movement (REM) sleep and non-REM (NREM) sleep (consisting of Stage 1, Stage 2, and slow wave sleep [SWS]). Over the course of a night, humans cycle through these stages, alternating between NREM and REM sleep around four or five times. One influential two-phase model of the role of sleep in memory consolidation (Diekelmann & Born, 2010) proposes that each stage of sleep actively contributes to different components of consolidation in a complementary manner, with the cyclical nature of sleep stages illustrating the interplay between those components. Diekelmann and Born suggest that NREM sleep supports *system consolidation*, whereby memory traces, encoded during the preceding period of wakefulness, are re-activated through coordinated activity between the hippocampus and the neocortex. The re-activation of memory traces is thought to drive the relocation of trace-specific activity from short-term hippocampal to long-term neocortical storage. The mechanisms of mnemonic redistribution during system consolidation have been further supported and refined through evidence of hierarchical, nested activity during SWS (Staresina et al., 2015). Specifically, neocortically generated slow oscillations during SWS, thalamocortically generated spindles (bursts of activity at around 10–16 Hz), and hippocampally generated ripples (bursts of activity at around 80–100 Hz) are functionally coupled in the human hippocampus. This process is believed to be orchestrated by slow oscillations, with behavioral change in hippocampally dependent memory being associated with the coupling of oscillatory activity (Cox, Hofman, & Talamini, 2012; Latchourmane, Ngo, Born, & Shin, 2017). Wei, Krishnan, Komarov, and Bazhenov (2018) further propose that spindles in Stage 2 sleep promote the reactivation and consolidation of multiple competing memory traces, while SWS preferentially consolidates stronger traces.

The second component of Diekelmann and Born's two-phase model proposes that the period of REM sleep following NREM sleep promotes the so-called *synaptic consolidation* in localized neocortical regions, allowing the stabilization of memory traces in long-term storage and enhancing the automaticity of signal processing. Sequential stages of sleep are therefore thought to act in concert to promote the reorganization and consolidation of memory, with the action of stages depending on the type of memory, the task used, and the strength of the memory trace at encoding.

Earle and Myers (2014) have suggested that the characteristics of sleep architecture (patterns of overnight brain activity) that phonological learning is associated with should vary depending on the task at hand. Tasks that rely more on auditory skills might be expected to be more strongly associated with REM-related synaptic consolidation, promoting a shift of attention to selective auditory features resulting in enhanced perceptual performance. By contrast, tasks that involve a greater degree of explicit recall or the integration of new memory traces with existing knowledge are likely to be associated, to a greater extent, with NREM-related system consolidation.

The Current Study

In this study, we trained children, who varied in their structural language ability and autism symptomatology, to listen to synthesized speech tokens. We tracked changes in performance immediately after training (in the morning) and then around 12 and 24 hr after training. At each test point, children were asked to identify words on which they had been trained (trained condition) and to generalize their learning to untrained words (novel condition). Our behavioral task was based on the paradigm developed by Fenn et al. (2013), with the crucial difference that, while these authors used a between-participants design with separate generalized and rote learning regimes, here we used a single learning regime in a within-participant design. This change was made to ease the recruitment of difficult-to-reach populations. While this approach allowed us to compare performance on trained and novel items in the same participants, it prevented an analysis of the links between sleep and rote learning per se, as participants could bring to bear their broader experience during training when listening to both trained and novel items. To assess the role of sleep in the time course of phonological learning, children were asked to wear polysomnography (PSG) sensors for the night after learning. Analysis of these data charts the time course of phonological learning and generalization, as well as relationships between sleep characteristics and phonological learning.

The current task required children to develop new phonological mappings by including synthetic tokens as acceptable exemplars of existing phonological categories and to apply that learning to trained and novel contexts. As this involves complex remapping of phonological representations, we expected to see an improvement in performance overnight (based on previous adult data from Earle & Myers, 2014; Fenn et al., 2013, 2003), but not over the wake interval. We further made two hypotheses regarding the relationship between behavioral change and aspects of sleep architecture. First, we hypothesized that phonological learning would be associated with synaptic consolidation during REM sleep, as indexed by theta power. This hypothesis is based on the prediction by Earle and Myers (2014) that tasks requiring a shift of attention toward relevant acoustic features of phonological stimuli to enhance perceptual performance will be most strongly associated with REM sleep. This notion is supported both by research showing that phonological learning is associated with attentional shifts to relevant acoustic cues and that perceptual learning is linked to REM sleep. The task at hand here, for both trained and novel conditions, requires participants to attend to linguistically informative acoustic cues in the synthesized speech. Exactly this kind of selective attentional shift to acoustic cues has been demonstrated in adults to support the categorization of ambiguous synthesized diphones, as well as generalization to novel contexts (Francis, Baldwin, & Nusbaum, 2000). A specific role for REM sleep is suggested by the finding that REM stabilizes performance on a visual perceptual learning task (Tamaki, Watanabe, & Sasaki, 2017). We hypothesized that generalization would be additionally related to system consolidation, as indexed by spindle activity and power during NREM sleep, as it relies on the integration of new information with existing phonological knowledge. Indeed, the generalization of grammatical rules has been experimentally boosted by re-exposing participants to an artificial grammar during SWS (Batterink & Paller, 2017), and over a nap, the extraction of linguistic rules has been linked to coordinated activity across SWS and REM sleep (Batterink, Oudiette, Reber, & Paller, 2014). In addition, Earle et al. (2017) found an association between performance on the identification of a newly learned nonnative phoneme and NREM sleep (Stages 1 and 2), possibly because this task required the integration of a new contrast into an existing phonological category. Here, we could expect to see a relationship between NREM sleep parameters and overnight change in performance on either condition, given that participants will be able to use their broader experience during training to support perception in both novel and trained conditions. We might expect stronger associations between REM sleep and performance on trained items and between NREM sleep and performance on novel items. We therefore consider brain activity during both REM and NREM (Stage 2 and SWS) sleep in relation to both conditions here.

Although there is less work considering the role of sleep in phonological learning across childhood, existing evidence supports our hypothesis that sleep should play a role in the establishment of new phonological knowledge. Recognition and recall of newly learned word forms, as well as the integration of word forms with existing lexical representations, improves 24 hr after learning in 7- to 10-year-old children (Henderson, Devine, Weighall, & Gaskell, 2015). Similar improvements are only seen over a 12-hr interval for 7- to 12-year-olds when that interval includes sleep (Henderson, Weighall, Brown, & Gaskell, 2012). It is possible that the benefits of sleep for novel word learning seen over the primary school years are linked to the particular dominance of SWS in childhood, with SWS both accounting for a greater proportion of total sleep time in childhood (Ohayon, Carskadon, Guilleminault, & Vitiello, 2004) and being of greater amplitude compared to adults (American Academy of Sleep Medicine [AASM], 2016).

We hypothesized that children with high autism symptomatology (above diagnostic threshold on a parent report screener) would show less phonological generalization than their peers with low autism symptomatology (Järvinen-Pasley et al., 2008; Mielke et al., 2013) and that sleep variables would be associated with these difficulties in phonological generalization. This association was expected given that children with a profile of ASD have frequently been reported to show atypical behavioral patterns of sleep and different sleep architectures compared to typically developing peers. In a meta-analysis of objective sleep measures (including PSG and actigraphy) in children with ASD, Elrod and Hood (2015) report "small but measurable" differences compared to typically developing peers, including longer sleep onset time (the time it takes to get to sleep once the lights are off) and shorter overall sleep time, with problems shown to be exacerbated in lower functioning children. With respect to the architecture of sleep, few studies exist, but those that do (see Díaz-Román, Zhang, Delorme, Beggiato, & Cortese, 2018, for a meta-analysis) point most consistently to a decrease in the density of spindles in Stage 2 sleep compared with typically developing peers (Godbout, Bergeron, Limoges, Stip, & Mottron, 2000; Limoges, Mottron, Bolduc, Berthiaume, & Godbout, 2005; Tessier et al., 2015). One study did not find a difference in spindle density (Maski et al., 2015), though the team did report reduced time in REM sleep, a finding that has been mirrored elsewhere (Limoges et al., 2005). The literature on sleep in ASD points to the hypothesis that reduced sleep and a reduced number of spindles in sleep may contribute to atypical phonological generalization the day after training. Children in our sample also varied with respect to structural language ability, which we hypothesized would predict the extent and rate of phonological learning for both trained and novel conditions. No hypotheses were made about how language variability might relate to sleep, as while some preliminary evidence exists of sleep difficulties in children with language disorders based on parent report (Botting & Baraka, 2018; Dominick, Ornstein Davis, Lainhart, Tager-Flusberg, & Folstein, 2007), the relationship remains largely unexplored.

Method

Participants

Participants were recruited to this study as part of the SleepSmart project at the University of York. For this phase of the project, 79 participants were recruited, two of whom did not provide behavioral data relevant to this study and were therefore not included in the analysis. Seventy-seven children (47 boys, 30 girls), with an average age of 10;1 (years; months), completed the study (SD = 17 months, range:

7;1–12;9). Children were excluded from participation if they had been regularly exposed to a language other than English from birth, had a history of epileptic seizures, had a genetic syndrome, or did not have sufficient oral language to give informed oral assent to take part. Four children were reported to be taking melatonin to support sleep at the time of the study, three of whom had a diagnosis of ASD and one was undergoing assessment for ASD; children were not asked to refrain from taking their usual medication to participate.

Measuring Autism and Language Ability

Children were recruited to provide substantial variability in language ability in those with low and high autism symptomatology. Participants were divided into two groups based on autism symptomatology as measured by the Gilliam Autism Rating Scale–Third Edition (GARS-3; Gilliam, 2013). The GARS-3 is a parent report screening measure that was norm-referenced with an autistic cohort. It provides an Autism Index (AI) score indicating whether a diagnosis of autism is "unlikely," "probable," or "very likely" for a given child or young adult. The AI score is a composite that reflects subscales covering restrictive and repetitive behaviors, social interaction, and cognitive and emotional processing. In the current study, the low-AI group consisted of 47 children (age: M = 9;10, SD = 17.0 months; 24 boys and 23 girls). Six children in the low-AI group scored below the 10th percentile on two or more standardized language tasks (see Table 1), consistent with a profile of DLD; one further child scored below the 10th percentile on the Elision subscale of the Comprehensive Test of Phonological Processing–Second Edition (CTOPP-2). Five children in the low-AI group had a sibling with a diagnosis of ASD, and three of these children scored in the "probable" range on the GARS-3 (AI: M = 56.8, SD = 8.1). A further three children, who were otherwise believed to be typically developing, scored within the "probable" range. The remaining 32 children in the low-AI group performed within the normal range on all cognitive measures and within the "unlikely" range on the GARS-3.

The high-AI group consisted of 30 children (age: M = 10;6, SD = 16.6 months), 23 of whom were male and seven were female. All of the children in this group scored within the "very likely" range on the GARS-3 (AI: M = 99, SD = 13.5). Within the high-AI group, 14 children had a diagnosis on the autism spectrum from a pediatrician, clinical psychologist, or multidisciplinary team; a further 11 were undergoing diagnosis at the time of testing. Although not all children in our sample had a diagnosis at the time of testing, on average, it takes 3.5 years to get a diagnosis in the United Kingdom (Crane, Chester, Goddard, Henry, & Hill, 2015), and four of the children awaiting diagnosis at the time of testing are known to have received their diagnosis by the time this article was submitted for publication. Of these 25 children, five had a language profile consistent with DLD even though they did not have additional diagnoses. In addition to the 25 children who had confirmed or pending diagnoses of ASD, a further five were included in the "high-AI" group as they scored in the "very likely" range on the GARS-3; four of these children had a diagnosis of a developmental disorder of language.

Although some children met clinical criteria for DLD, in this study, we treat language as a continuous variable. We did not adopt this approach with the GARS-3 as this measure resulted in two distinct groups (i.e., at either end of the spectrum of GARS-3 scores), which were not described by a normal distribution; in addition, the GARS-3 is a parental report measure, which may be biased by preexisting parental views, potentially reflecting the stage families are at in the diagnostic process. The distribution of GARS-3 scores reflected our recruitment efforts. By contrast, our language measures were objective, standardized tests. Children's Communication Checklist–Second Edition (CCC-2; Bishop, 2003) profiles suggest that the sample as a whole spanned a wide range of ability along both structural and social language continua (see Figure 1 for the distribution of scores on the GARS-3 and CCC-2, along with cognitive profiles by group in Table 1).

Materials

Auditory stimuli were generated using a Votrax SC-01-A text-to-speech synthesizer[1] (Gagnon, 1978). We used synthesized speech first to align our paradigm with Fenn et al. (2013, 2003) and, second, because the degradation seen in synthesized speech is nonuniform across cues, such that listeners cannot make simple perceptual adjustments in order to map degraded speech to established phonology. The Votrax synthesizer is a formant-based synthesis-by-rule program, whereby formant synthesized speech is built on the fly, based on rules regarding the acoustic properties of formant transitions. Synthesized words are built using a directory of 64 phonemes, concatenated using rules to model coarticulation and inflection. The overall impression is of a male robot with a slight American accent. Tokens were individually synthesized and downloaded in WAV format with a sampling rate of 44.1 kHz.

The stimulus set was composed of nine lists of 25 concrete nouns: 225 words in total (see Appendix). Over the course of the experiment, all participants were exposed to all lists: one at pretest, four plus the pretest list in training, and four in posttests (one plus the pretest list in each posttest). The word lists were balanced with respect to syllable and phoneme number, frequency of occurrence in spoken U.S. English (Brysbaert & New, 2009), age of acquisition (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012), phonotactic probability (segment average and biphone average; Vitevitch & Luce, 2004), concreteness (Brysbaert et al., 2014), and the percentage of the words that were made up of the 10 most common phonemes in spoken American English (Mines, Hanson, & Shoup, 1978; see Table 2 for details of these variables across lists). A pilot study with adults was run in order to make sure that naive listeners would not perform at floor on the task. Words between one and four syllables in length were selected for inclusion in

[1]Can be found at http://real-votrax.no-ip.org/.

Table 1. Descriptive statistics for cognitive measures, split by Autism Index (AI) group, with group differences shown.

Group	BPVS	Recall. Sen.	Word Defs.	Elision	NWR	RAN Digits	Lang Comp	Matrices	Back digit
Low AI	106.6	108.0	106.8	104.8	110.4	102.4	106.4	100.8	105.8
High AI	95.4	93.4	96.5	95.4	96.9	91.9	94.6	95.3	98.5
t	2.710**	3.055**	2.483*	2.486*	3.545***	3.524***	4.133***	1.213	1.492
Total	102.3	102.4	102.9	101.5	105.6	98.7	101.8	98.7	103.1

Note. BPVS = British Picture Vocabulary Scale; Recall. Sen. = Recalling Sentences subtest from the Clinical Evaluation of Language Fundamentals–Fourth Edition; Word Defs. = Word Definitions subtest from the British Ability Scales–Third Edition (BAS-3); Elision, NWR, & RAN Digits = Elision, Nonword Repetition, and Rapid Automatic Naming Digits subscales from the Comprehensive Test of Phonological Processing–Second Edition; Lang Comp = language composite measure; Matrices = matrices from BAS-3; Back digit = backward digit span from BAS-3.

**p* < .05. ***p* < .01. ****p* < .001.

the lists if between 20% and 90% of adult listeners correctly identified them on first hearing, if they were highly concrete, and if they had an age-of-acquisition of less than 7 years. A few items had an age-of-acquisition over 7 years, including “dustbin,” “badger,” and “hedgehog,” but these items were included as the Kuperman ratings are from an American sample and British children were judged to learn these words at a younger age than their American counterparts. Each token was normalized, and any DC offset was removed using Audacity (Audacity Team, 1999–2016). Stimuli were always presented on a laptop with over-ear Superlux headphones at a comfortable listening volume.

Design and Procedure

Participants were seen at five time points: once for an initial assessment of cognitive ability, three times over the course of 24 hr (Day1:AM, in the morning of the first day; Day1:PM, in the evening around 12 hr later; and Day2:AM, on the morning of the second day), and then once again at follow-up around 4 weeks later (Follow-Up). The dependent variable in this study was proportion of words correctly repeated (Accuracy). Independent variables were Group (high AI, low AI), Language Composite Score and Sleep Architecture Between Participants, and Condition (trained, novel) and Session (Day1:AM-pre, Day1:AM-post, Day1:PM, Day2:AM, Follow-Up) within participants.

Each participant was designated one trained list, which they heard at pretest, during training, and at each subsequent posttest after training. Four additional lists were used during training; then, at each posttest, participants heard one novel list along with their trained list. Pseudorandom counterbalancing was achieved by establishing nine different permutations of list exposure and assigning participants to one of these permutations sequentially as testing progressed. Testing was completed in a one-to-one setting with one of a small team of researchers and was conducted either at the child’s home, at school, or in the Department of Psychology at the University of York. All children who completed behavioral testing for this study were also participants in a study of semantic learning, training and testing for which were completed during the Day1:PM and Day2:AM sessions (see Fletcher et al., 2019).

Figure 1. (a) Distribution of scores on the Children’s Communication Checklist–Second Edition; the box indicates scores considered to fall in the range of autism spectrum disorder (ASD), and cases are marked by group and whether the child has a diagnosis of ASD within the high–Autism Index (AI) group. (b) Distribution of Gilliam Autism Rating Scale–Third Edition (GARS-3) scores showing the split between high- and low-AI groups (low AI, *n* = 47, *M* = 51.2; high AI, *n* = 30, *M* = 96.1).

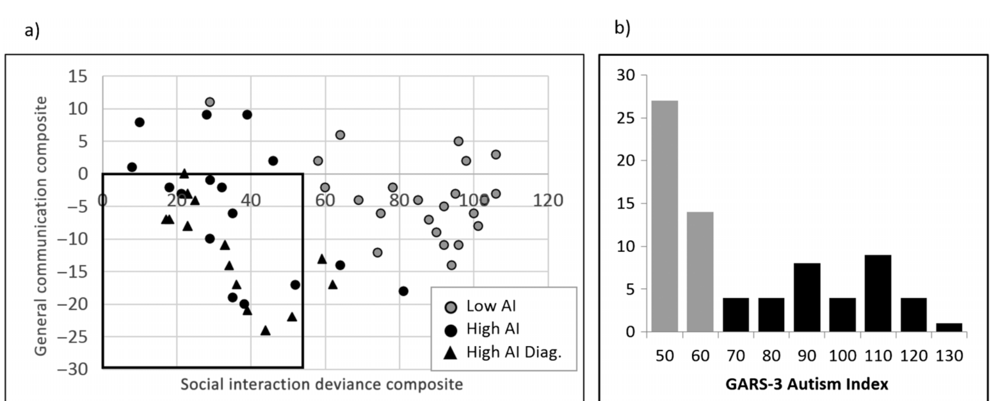


Table 2. Characteristics of trial items across lists.

List	%correct@pilot	Freq.	AoA	Phon count	Phon p	Biphone p	Concrete
1	54.4	28.31	5.08	5.2	1.23	1.01	4.8
2	55.6	29.69	5.19	4.9	1.22	1.02	4.8
3	55.6	22.56	5.45	5.2	1.25	1.02	4.8
4	57.8	33.39	5.44	5.1	1.26	1.02	4.8
5	58.6	19.24	5.42	5.2	1.24	1.02	4.8
6	54.8	16.09	4.94	5.4	1.25	1.02	4.8
7	54.4	21.31	5.21	5.2	1.24	1.01	4.8
8	54.8	34.40	4.74	5.4	1.24	1.01	4.8
9	56.0	18.74	5.10	5.4	1.27	1.02	4.7
Av.	55.8	24.9	5.2	5.2	1.24	1.02	4.8

Note. %correct@pilot = percentage of items correctly identified by a pilot sample of 10 adults; Freq. = frequency of occurrence per million using the SUBTLEX-US frequency norms of Brysbaert and New (2009); AoA = age of acquisition from Kuperman et al. (2012); Phon count = number of phonemes; Phon p and Biphone p give average phonotactic probabilities for phonemes and biphones within the words, using an online calculator by Vitevitch and Luce (2004); Concrete = concreteness ratings from Brysbaert et al. (2014); Av. = average.

Cognitive Assessments

The following standardized assessments were administered in accordance with published guidelines, always in the same order: British Picture Vocabulary Scale–Third Edition (Dunn, Dunn, & Styles, 2009); Matrices, Forward Digit Recall, Word Definitions, and Backward Digit Recall subscales from the British Ability Scales–Third Edition (Elliott & Smith, 2011); Recalling Sentences subscale from the Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel & Wiig, 2006); and Elision, Rapid Automatic Naming (RAN Digits and Letters), and Non-Word Repetition from the CTOPP-2 (Wagner, Torgesen, Rashotte, & Pearson, 2013)—Non-Word Repetition was rerecorded by a female speaker of British English with a southern accent, trained in phonetics. The parents of all children were also asked to complete a series of questionnaires about their child: the Children's Sleep Habits Questionnaire (Owens, Spirito, & McGuinn, 2000), the Child Behaviour Checklist (Achenbach & Edelbrock, 1983), the CCC-2 (Bishop, 2003), and the GARS-3 (Gilliam, 2013).

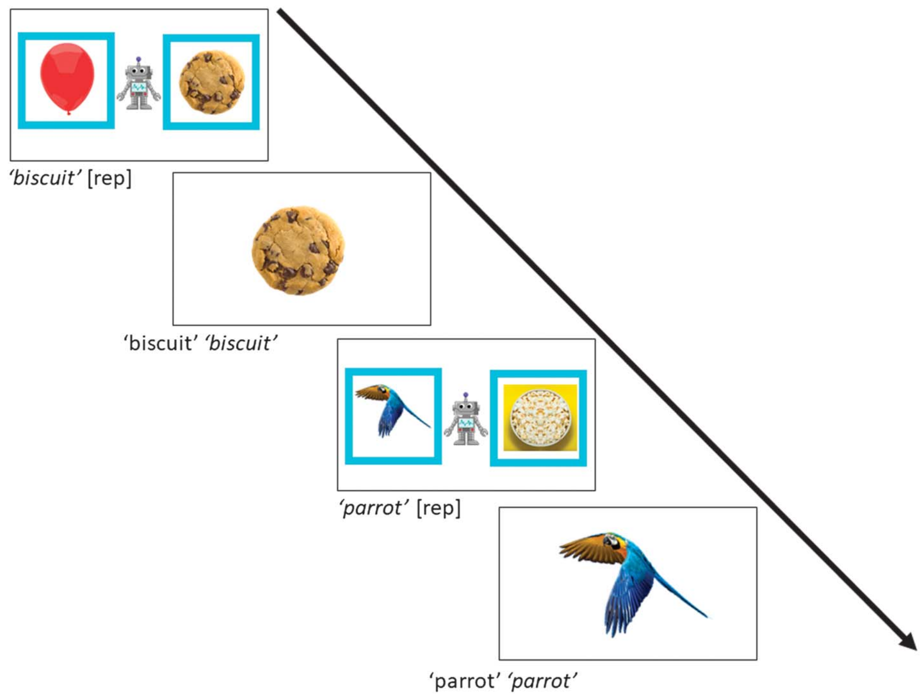
Experimental Session 1 (Day1:AM)

On the morning of Day 1, at around 9 a.m., participants were tested on 25 novel synthesized words (Day1:AM-pre), with this word list going on to be the trained list for that participant at all subsequent time points. After the task was introduced, each item was presented in isolation, with a picture of a robot on the computer screen. Participants were asked to repeat what the robot said, take a guess, or say that they did not know. The experimenter recorded "correct" or "incorrect" with a mouse click, which moved the paradigm on to the next test item without providing feedback for the child. Item order was randomized, and no feedback was given during this testing phase. Stimuli were presented, and verbal responses were recorded by E-Prime2 (Psychology Software Tools, 2012).

After the Day1:AM-pre test, participants completed a training phase. During training, participants heard one novel stimulus item per trial and were asked to perform a two alternative forced choice task with a verbal repetition response. During the presentation of each auditory token, two pictures were displayed on the computer screen: one being the target and the other being a distractor with the same number of syllables and the same initial phoneme, for example, "parrot" and "popcorn" (see Figure 2). The picture onset was aligned to the start of the auditory stimuli, and each remained on the screen until the participant made their response, at which point the experimenter marked it as correct or incorrect with a mouse click. The experimenter mouse click triggered participant feedback in the form of a clear auditory token spoken by a British man, followed by the synthesized token again after a 500-ms interval. During feedback, the target picture remained in the center of the screen. Each participant completed training with five lists of 25 words, with one of these lists being the participant's "trained" list; list order and items within lists were randomized during training. Training lasted around 20 min, including breaks between lists as needed. Feedback was included here as studies have shown that reinforcement facilitates the overnight consolidation of transitive inference (Werchen & Gómez, 2013), and hearing clear than distorted feedback during the perceptual learning of distorted speech markedly facilitates learning (Hervais-Adleman, Davis, Johnstrude, & Carlyon, 2008). Task demands were kept minimal, with low working memory requirements.

After training, participants were immediately tested on two lists (Day1:AM-post): the list that they were tested on before training (their trained list) plus one novel list. Presentation of these 50 items (25 novel and 25 trained) was randomized across the test. At each experimental session, participants also completed a psychomotor vigilance test (PVT) based on one developed by Basner, Mollicone, and Dinges (2011), which took around 4 min to complete. During the test, which was presented on a laptop using E-Prime2 software, participants were required to give a button press response as quickly as possible when a star appeared on the screen. The PVT was included to assess sustained attention at each test session, allowing a consideration of fatigue

Figure 2. Example trials for “biscuit” and “parrot”; synthesized speech is shown in italics.



during the evening session that could lead to spurious overnight improvements in performance.

Subsequent Experimental Sessions

On the evening of Day 1 at around 6 p.m., the Day1:PM session was conducted, during which participants were again tested on the synthesized speech stimuli; the test consisted of repeating what the robot said when presented with the participant’s trained list plus one novel list: a total of 50 items randomly presented. The Day2:AM experimental session, conducted at around 9 a.m. on the second day, and the Follow-Up session, followed the same format as the Day1:PM session but with different novel lists.

Sleep Measurement

During the night between Day 1 and Day 2, participants had their brain activity recorded with PSG. One of four electronically identical portable recording devices was used, either a Titanium by Embla or a Morpheus by Micromed. Recordings were taken at a sampling frequency of 256 Hz from six locations on the scalp (C3/C4/F3/F4/O1/O2), plus lower left and upper right EOG and two electromyography channels on the chin. The ground electrode was placed on the forehead, and Cz acted as an online reference, with offline re-referencing to the contralateral mastoid. Recordings were made in RemLogic 3.4.

Ethical approval for this study was granted by the research ethics committee in the Department of Psychology at the University of York. All parents provided informed written consent for their children to take part, and all children provided informed oral assent on the first day of testing and before PSG.

Results

Behavioral Data Analysis

Data can be found on the Open Science Framework at https://osf.io/82eqm/files/. Data were analyzed using R (R Core Team, 2017), with the “lme4” package (Bates, Mächler, & Bolker, 2012) with plots made using the “ggplot2” package (Wickham, 2016). Three binomial logistic mixed-effects models are presented with Accuracy as the dependent variable (see Table 3 for performance descriptives). Models were fitted in two stages. A backward model selection procedure was adopted to establish a parsimonious fixed effects structure, starting with a maximal fixed effects structure (i.e., simple fixed effects and all interaction terms), along with random intercepts. Fixed effects were then individually knocked down starting with highest order interactions. The removal of each fixed effect was assessed via likelihood ratio tests, and the removal of a fixed effect was justified if there was no evident reduction in model fit ($p > .2$). At each stage, the fixed effects that contributed least to model fit were removed first (the largest p value via likelihood ratio test). Having established which simple fixed effects and interactions contributed to model fit, random intercepts were justified. Finally, random slopes were added one-by-one to see if each alone or in combination

Table 3. Accuracy means (and standard deviations) for all participants by session, condition, and group for the phonological learning task and reciprocally transformed reaction times and lapses for the psychomotor vigilance task (PVT).

Task	Group	Day1:AM-pre	Day1:AM-post	Day1:PM	Day2:AM
Phonological learning: Proportion correct					
Novel	Low AI	0.331 (0.172)	0.618 (0.119)	0.596 (0.131)	0.642 (0.165)
	High AI	0.251 (0.151)	0.574 (0.173)	0.569 (0.205)	0.609 (0.193)
	Total	0.291 (0.166)	0.596 (0.149)	0.583 (0.170)	0.626 (0.179)
Trained	Low AI	—	0.824 (0.104)	0.832 (0.105)	0.863 (0.094)
	High AI	—	0.805 (0.133)	0.766 (0.166)	0.826 (0.124)
	Total	—	0.815 (0.119)	0.797 (0.142)	0.845 (0.111)
PVT					
1/RT	Low AI	—	2.90 (0.40)	2.75 (0.41)	2.80 (0.51)
	High AI	—	2.74 (0.55)	2.61 (0.47)	2.55 (0.55)
	Total	—	2.85 (0.45)	2.70 (0.43)	2.72 (0.53)
Lapses	Low AI	—	11.09 (11.97)	16.94 (12.95)	16.44 (16.56)
	High AI	—	18.03 (20.00)	23.5 (16.71)	27.90 (20.02)
	Total	—	13.79 (15.84)	19.46 (14.76)	20.75 (18.67)

Note. Em dashes indicate that this measure was not taken pretraining. AM1 = morning of the first day; PM = evening around 12 hr later; AM2 = morning of the second day; AI = Autism Index; RT = response time.

made a difference to the fit of the model under a liberal criterion of $p < .2$ (Bates, Mächler, Bolker, & Walker, 2015) compared to the fixed effects and intercepts-only model. Slopes were built up in stages, with the slope that contributed most to model fit being kept at each stage. The best fitting models are described below for (a) data spanning Day1:AM-pre and Day1:AM-post (the "training model"), (b) data describing the time course of performance at 0 and then approximately 12 and 24 hr after training (Day1:AM-post, Day1:PM, Day2:AM; the "time-course model"), and (c) data from the Follow-Up session (the "follow-up model").

For each model, session and AI group (high AI vs. low AI) were entered as factorial predictors. AI group was coded using simple coding (low AI: −0.5, high AI: 0.5), and language composite (language) as a continuous predictor, with this measure being scaled and centered. The language composite was formed from an average of each child's standard scores over six measures of language ability: British Picture Vocabulary Scale–Third Edition (Dunn et al., 2009); Word Definitions British Ability Scales–Third Edition (Elliott & Smith, 2011); Recalling Sentences subscale from the Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel & Wiig, 2006); and Elision, RAN Digits, and Non-Word Repetition from the CTOPP-2 (Wagner et al., 2013). RAN Letters was not used as nine children were not familiar enough with the names of the letters to complete the task without resorting to letter sounds, such that it could not be considered a test of phonological access for those children. For the time-course and follow-up models, condition (novel vs. trained) was entered as a factorial predictor using simple coding (novel: −0.5, trained: 0.5). Intercepts for participants and items were included as random effects, along with by-item random slopes for the effects of session, language, and AI and by-participant slopes for session.

In each model, influential cases were detected using the package influence.ME (Nieuwenhuis, te Grotenhuis, & Pelzer, 2012) to calculate DFBETAS (Belsley, Kuh, & Welsch, 1980) for each of the simple fixed effects and their interactions in the final models. Z scores were calculated from DFBETAS, and participants were removed from the model if they had z scores more extreme than ±3.29 for any significant fixed effect or interaction (two participants from the high-AI group were removed from the training model, three from the low-AI group were removed from the time-course model, and two from the high-AI group and one from the low-AI group were removed from the follow-up model).

The Training Model

Details of the best fit training model are given in Table 4. The fixed effects of session (Day1:AM-pre vs. Day1:AM-post coded using simple coding: Day1:AM-pre, −0.5; Day1:AM-post, 0.5) and language significantly contributed to the model fit, with accuracy on novel items post-training (see Figure 3a) being better than accuracy during the pretraining test and better performance on the language composite predicting better overall performance in the Day1:AM session (see Figure 3b). No interactions between fixed effects significantly contributed to the model fit. Training therefore successfully improved performance for the sample as a whole, and while language ability predicted performance on the task, it did not predict the benefit of training.

Time-Course Model

We next analyzed the time course of performance on trained and novel items after training with degraded speech. Session (Day1:AM-post, Day1:PM, Day2:AM) was entered as a factorial predictor with three levels; therefore, planned contrasts were coded using forward difference coding (UCLA Statistical Consulting Group, n.d.) to include a comparison between the Day1:AM and Day1:PM sessions (daytime change: Day1:AM-post = 2/3, Day1:PM = −1/3, Day2:AM = −1/3) and the Day1:PM versus Day2:AM sessions (overnight change: Day1:AM-post = 1/3,

Table 4. Fixed and random effects for model of performance accuracy pre- and posttraining in Session 1: training model.

	Fixed effects					Random effects	
			95% CI			Item	Participant
Term	*b*	*SE*	Lower	Upper	*z*	*SD*	*SD*
(Intercept)	**−0.253**	**0.119**	−0.486	−0.020	**−2.123***	1.647	0.240
Session (pre–post)	**1.888**	**0.151**	1.592	2.184	**12.495*****	0.355	0.620
Group (low–high AI)	−0.064	0.161	−0.380	0.252	−0.398	—	—
Language (composite)	**0.277**	**0.0839**	0.113	0.441	**3.299****	—	—
Session × Group	0.254	0.281	−0.297	0.805	0.905	—	—
Session × Language	−0.086	0.147	−0.374	0.202	−0.584	—	—
AI Group × Language	0.122	0.169	−0.209	0.453	0.723	—	—
Session × Group × Language	0.234	0.298	−0.350	0.818	0.784	—	—

Note. Model formed from 3,726 observations: 75 participants across 225 items and two sessions. AI = Autism Index; CI = confidence interval.

*$p < .05$. **$p < .01$. ***$p < .001$.

Day1:PM = 1/3, Day2:AM = −2/3). It was hypothesized, in line with previous research with adults, that performance on novel items (indicating generalization) would decline between Day1:AM-post (0 hr) and Day1:PM (12 hr) and then improve at the Day2:AM session (24 hr) after a night of sleep. In comparison, trained items were expected to show maintained performance at all points after training. Overall performance was hypothesized to be predicted by language ability, while those in the high-AI group were predicted to show poorer performance on the novel condition and to demonstrate less improvement in performance overnight (Day1:PM to Day2:AM).

Over the wake retention period (Day1:AM-post to Day1:PM), overall performance did not significantly change (Day1:AM-post: $M = 0.706$, $SD = 0.171$; Day1:PM: $M = 0.694$, $SD = 0.189$). By contrast, performance was observed to significantly improve over the course of the sleep period (Day2:AM: $M = 0.737$, $SD = 0.183$). As shown in Table 5, condition (novel vs. trained) also contributed to model fit, with performance on the trained condition ($M = 0.820$,

Figure 3. Training model. (a) Performance pre- and posttraining at Session1; posttraining, only novel items are included. Error bars show standard error. (b) Relationship between the language composite measure and average performance over Session 1. One participant was removed from this analysis after being identified as an influential case with DFBETAS—they are shown in red but excluded from the figure summary statistics.

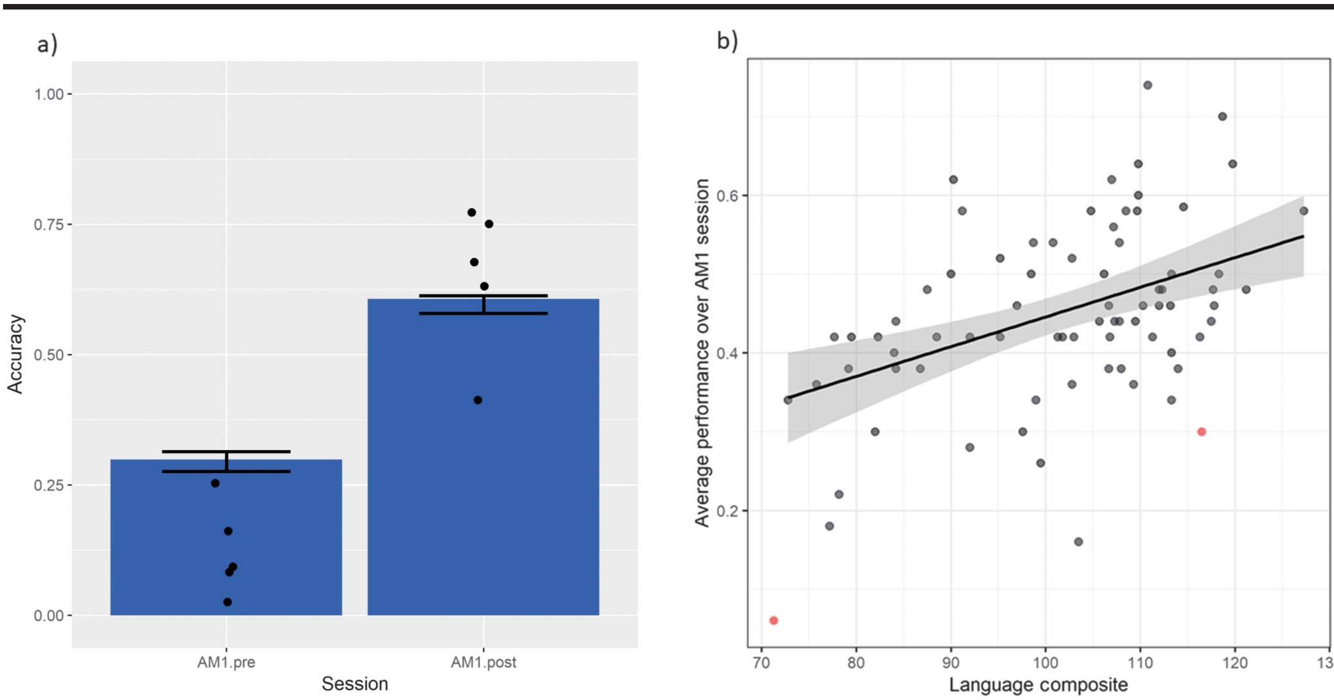


Table 5. Fixed and random effects for model of performance accuracy at the AM1-post, PM, and AM2 session: time-course model.

	Fixed effects					Random effects	
			95% CI			Item	Participant
Term	*b*	*SE*	Lower	Upper	*z*	*SD*	*SD*
(Intercept)	**1.597**	**0.138**	**1.327**	**1.868**	**11.573*** **	1.471	0.657
Session 1 (AM1-post–PM)	0.067	0.063	−0.057	0.191	1.056	—	—
Session 2 (PM–AM2)	**−0.300**	**0.064**	**−0.424**	**−0.175**	**−4.706*** **	—	—
Condition (novel–trained)	**1.737**	**0.118**	**1.505**	**1.968**	**14.713*** **	0.9587	0.405
Group (low–high AI)	0.166	0.187	−0.200	0.532	0.891	—	—
Language (composite)	**0.468**	**0.101**	**0.269**	**0.666**	**4.625*** **	0.211	—
Condition × Group	0.275	0.192	−0.101	0.651	1.433	1.018	—
Condition × Language	**0.281**	**0.092**	**0.101**	**0.460**	**3.066** **	—	—
Group × Language	0.165	0.203	−0.233	0.563	0.811	0.451	—
Condition × Group × Language	**−0.409**	**0.185**	**−0.771**	**−0.047**	**−2.216* **	—	—

Note. Model formed from 11,015 observations: 74 participants across 225 items and three sessions. Data in bold are significant. AM1 = morning of the first day; PM = evening around 12 hr later; AM2 = morning of the second day; CI = confidence interval; AI = Autism Index.
*Significant at $p < .05$. **Significant at $p < .01$. ***Significant at $p < .001$.

SD = 0.125) being substantially better than performance on the novel condition (*M* = 0.604, *SD* = 0.165), though no interaction between condition and session was observed. The significant contribution of the Day1:PM versus Day2:AM comparison is consistent with the hypothesis that sleep is associated with phonological learning on this task (see Figure 4), and the lack of interaction with condition indicates that both novel and trained items benefited. As we saw with the training model, children who showed a better language ability also performed better on the task overall, with language contributing significantly to model fit. A two-way interaction between condition and language also emerged. This interaction was driven by a tendency toward ceiling effects in the trained condition for children with a better language ability (evident in Figure 5). This interpretation is supported by significant heteroscedasticity in the trained condition but not in the novel condition, as demonstrated by nonconstant variance score tests on linear regression models predicting

Figure 4. Time-course model. Performance 0 (AM1), 12 (PM1), and 24 (AM2) hr after; error bars show standard error. Two participants were removed from this analysis after being identified as influential cases with DFBETAS—they are shown in red but excluded from the figure summary statistics.

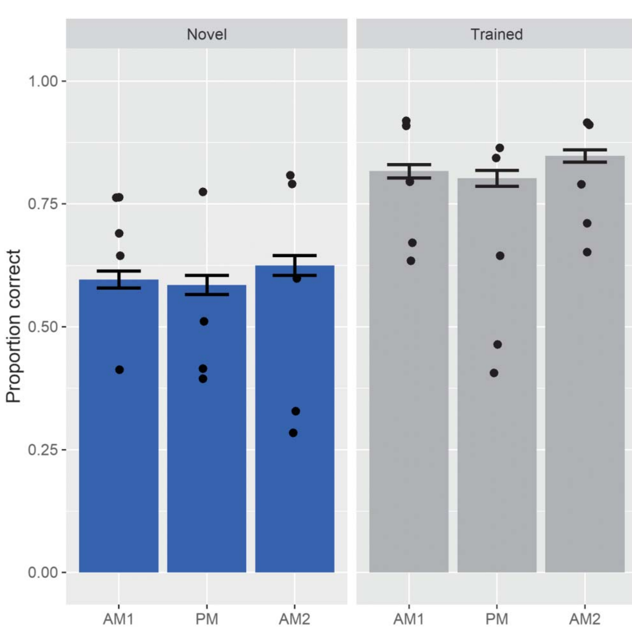
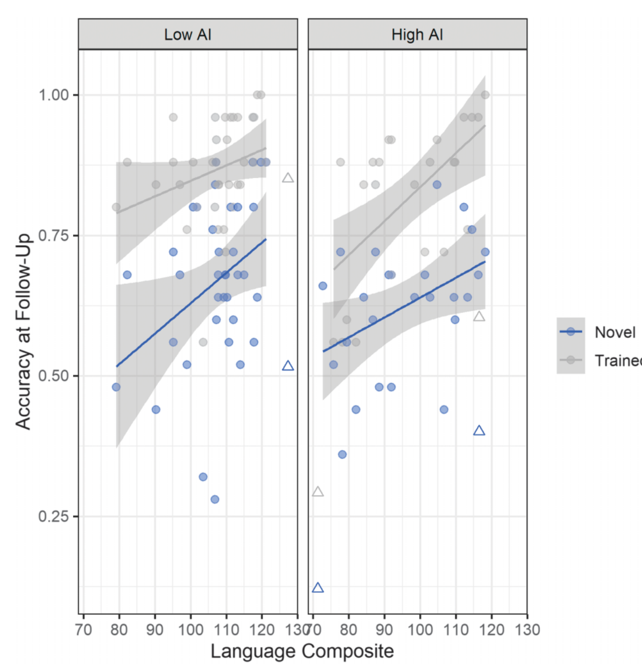


Figure 5. Interaction shown in the time-course model between language composite, Autism Index (AI) group, and condition. Data points shown as triangles were excluded from the model on the grounds of being overly influential according to DFBETAS; these participants are not included in the figure summary statistics.



overall performance in each model by language: trained, $\chi^2(1) = 157.3$, $p < .001$; novel, $\chi^2(1) = 2.2$, $p = .142$. An interaction between condition, language, and group also emerged, as shown in Figure 5, which suggests that language ability predicts performance better in the trained condition than in the novel condition for those with low autism symptomology. This interpretation was supported using the package emmeans (Lenth, Singmann, & Buerkner, 2019): We found that accuracy as a function of language composite differed between conditions for the low-AI group, while this was not true for the high-AI group (low AI: novel–trained estimate = −0.485, SE = 0.150, z ratio = −3.230, p = .0012; high AI: novel–trained estimate = −0.076, SE = 0.106, z ratio = −0.717, p = .4734; with Tukey correction for multiple comparisons). This seems to be driven by the particularly shallow slope seen for novel items in the low-AI group—here, having a lower language ability does not confer a disadvantage on the novel condition. AI did not interact with condition, suggesting that the high-AI group did not show a specific deficit in phonological generalization on this task. We consider change in performance overnight in relation to the sleep variables of interest in more detail below.

Attention

The PVT data were analyzed by dividing each response time (RT) in milliseconds by 1,000 and then reciprocally transforming (1/RT, "response speed," in line with Basner & Dinges, 2011) before calculating averages for each participant over the Day1:AM, Day1:PM, and Day2:AM sessions (see Table 3). RTs were excluded from this process if they were shorter than 100 ms. A mixed analysis of variance with session (Day1:AM, Day1:PM, Day2:AM) and group (by AI) revealed a main effect of session, $F(2, 146) = 12.69$, $p < .001$, which was driven by a difference between Day1:AM (retransformed M = 399.77 ms, SD = 250.57) and Day1:PM (retransformed M = 436.67 ms, SD = 384.50; $p < .001$) and between Day1:AM and Day2:AM (Day2:AM: retransformed M = 460.95 ms, SD = 406.92; p = .001). No main effect of group or interaction between group and session was observed. We also analyzed the number of lapses observed at each session, defined as the number of reaction times greater than 500 ms. Lapses showed a main effect (correcting for violation of sphericity where appropriate) of session, $F(1.7, 128.3) = 24.875$, $p < .001$, driven by a difference between Day1:AM (M = 13.16, SD = 14.93) and Day1:PM (M = 19.53, SD = 14.74) at $p < .001$ and between Day1:AM and Day2:AM (M = 20.97, SD = 18.69) at $p < .001$. A main effect of group was observed in the lapse data, $F(1, 74) = 4.755$, $p = .032$, as well as a marginally significant interaction between session and group, $F(1.7, 128.3) = 3.217$, $p = .050$, driven by a difference between groups only in Session 3, $t(75) = 2.718$, $p = .008$ (low AI: M = 16.70, SD = 16.36; high AI: M = 27.90, SD = 20.02). While the faster reaction times and fewer lapses at Session 1 likely indicate that children were tired of the repetitive task (as supported by feedback from children), the lack of difference between Day1:PM and Day2:AM indicates that overnight improvement in performance cannot be attributed to enhanced attention in the morning.

Follow-Up Model

Finally, we considered performance at the follow-up test approximately 4 weeks after initial training (follow-up). The follow-up was completed by 64 children, an average of 32.7 days (min = 25, max = 47, SD = 5.9) after training. Follow-up performance was not compared to prior sessions as not all children completed the follow-up session, although broadly speaking, performance at this test point suggests that learning was well maintained in the sample. As shown in Table 6, condition and language both contributed to model fit, supporting findings from the time-course model that the trained condition was easier for children (novel: M = 0.650, SD = 0.477; trained: M = 0.847, SD = 0.360) and that those with stronger language skills were able to perform at a higher level. An interaction between condition, language, and group also emerged, as in the time-course model. Here, emmeans revealed a somewhat different pattern to that seen in the time-course model: with the novel

Table 6. Fixed and random effects for model of performance accuracy at Session 3 and the follow-up session approximately 4 weeks later: follow-up model.

	Fixed effects					Random effects	
			95% CI			Item	Participant
Term	*b*	*SE*	Lower	Upper	*z*	*SD*	*SD*
(Intercept)	**1.694**	**0.151**	**1.398**	**1.990**	**11.249*****	1.354	0.573
Condition (novel–trained)	**1.671**	**0.179**	**1.320**	**2.022**	**9.351*****	0.918	—
Group (low–high AI)	0.064	0.210	−0.348	0.476	0.303	—	—
Language (composite)	**0.476**	**0.111**	**0.258**	**0.694**	**4.284*****	—	—
Group × Condition	−0.036	0.256	−0.538	0.466	−0.140	—	—
Language × Condition	0.199	0.133	−0.062	0.460	1.492	—	—
Group × Language	0.047	0.224	−0.392	0.486	0.211	—	—
Condition × Group × Language	**0.555**	**0.276**	**0.014**	**1.096**	**2.012***	—	—

Note. Model formed from 3,050 observations: 61 participants across 225 items. CI = confidence interval; AI = Autism Index.
*p < .05. ***p < .001.

and trained slopes becoming more similar over time for the low-AI group and more distinct for the high-AI group (low AI: novel–trained estimate = 0.079, SE = 0.202, z ratio = 0.390, p = .6965; high AI: novel–trained estimate = −0.477, SE = 0.182, z ratio = −2.623, p = .0087; with Tukey correction for multiple comparisons). Four weeks after training, low language scores were less of a disadvantage on the novel condition for the high-AI group (see Supplemental Material S1 for illustration).

In each of the mixed-effects models presented here, both language and AI group have been included; as the AI groups differ with respect to language ability (see Table 1), this is a potential confound. In order to make sure that any effects of group were not hidden by those children with primary language difficulties who were included in the high-AI group on the account of having a GARS-3 score within the "very likely" range, we plotted the performance of the five children who did not have a current or pending diagnosis of ASD relative to the rest of the sample. The graphs shown in Supplemental Materials S2 and S3 suggest that these children were spread across the range of performance on this task. This check, in combination with the fact that we see consistently strong relationships between performance accuracy and language ability, but never AI group, suggests that autism symptomatology is not contributing in any meaningful way to performance or performance change.

Sleep Data

Staging of Electroencephalogram (EEG) Data

Each data set was hand-scored independently by two out of a pool of three researchers in accordance with the AASM rules for children (AASM, 2016). Independent scoring resulted in 82.9% concordance. Where scorers disagreed on the staging of 10 or more consecutive 30-s epochs, the data were reconsidered, and if an agreement could not be reached, the staging given by the designated first scorer was used. Of the 77 participants who contributed behavioral data in this study, 54 contributed enough stageable PSG data to identify a sufficient amount of both REM and NREM sleep for the extraction of sleep parameters (35 boys and 29 girls, with a mean age of 121.17 months, SD = 16.40). Of these, 33 participants were from the low-AI group, and 21 were from the high-AI group. The loss of PSG data is primarily due to the loss of key electrodes over the course of the night (particularly toward the morning), which impacted on the quality of REM data toward the end of recordings.

To ensure that overnight EEG measurements were representative of a normal night of sleep, participants were given an Ambulatory Monitoring actigraph watch to wear for approximately four nights, including the night of learning between Day1:PM and Day2:AM. Some participants chose to wear the watch longer. For the 54 participants who contributed staged EEG data, on average, the watch was worn for 4.56 nights (min = 3, max = 7), with an average of 3.35 nights (min = 0, max = 5) when the child was in their normal bedtime routine. Children were taken to be in their normal routine on a week night during term time; for the two children who wore the actigraph watch and were home schooled, any night of the week was counted as routine. We compared total sleep time on the night of learning with an equivalent night (in or out of routine): Overall, the participants did not differ between the two nights (p = .506), and split by AI group, there were no differences on this measure either (low-AI group: learning night mean = 481.89 min, comparison night mean = 463.96 min, p = .344; high-AI group: learning night mean = 445.31 min, comparison night mean = 446.94 min, p = .957). We can therefore assume that the EEG measurements taken on the night of learning were reasonably representative of sleep for our sample.

Spindle Density and Power Calculations

Spindles were identified and counted using an algorithm written by Tsanas and Gifford (2015), which uses a continuous wavelet transform to identify characteristic patterns of activity between 10 and 15 Hz. Absolute power was calculated in MATLAB (MathWorks, 2017) at each scalp electrode for fast (10.00–12.49 Hz) and slow (12.50–14.99 Hz) spindle ranges, delta (0.30–3.99 Hz), alpha (9.00–12.99 Hz), and beta (13.00–35.00 Hz); natural log transformations were applied to all power variables before analysis. In order to maximize usable data sets, we opted to analyze the first 3 hr of NREM data (Stage 2 and SWS) and the first 1.25 hr of REM data. For the 27 children who required the least artifact rejection over the course of the night, the correlation between absolute power in NREM stages in the first 3 hr and all night was r = .91 for the slow spindle range (10.00–12.49 Hz) frontally and r = .91 for the fast spindle range (12.50–14.99 Hz) centrally; for central REM theta power and between the first 1.25 hr and all night, it was r = .95. Children were included in the analysis of spindles and power if they contributed all parameters to both REM and NREM data sets; on these grounds, 48 children were included, 31 from the low-AI group and 17 from the high-AI group. For spindle and power analyses, data were extracted from C4 and F4 where possible, but in 11 cases, C3 was deemed to be the cleaner channel, and in eight cases, F3 was taken.

The hypothesis-driven predictors included were central spindle density, frontal slow spindle power (10.00–12.49 Hz), and central fast spindle power (12.50–14.99 Hz) for Stage 2 and Stage 3 separately; SWS frontal delta power (0.30–3.99 Hz); and central REM theta power; AI, language, and age in months were also included. Relationships between overnight change in accuracy and characteristics of sleep architecture were considered using linear regression. The leaps package (Lumley, 2017) was used to exhaustively search for the subset of variables that provided the highest adjusted R^2 (AdjR^2) with these variables then being used to predict change in performance for each condition.

None of the log-transformed sleep variables considered here differed between AI groups (see Table 7). That being said, in a partially overlapping sample of 17 children with typical language ability and a diagnosis or pending diagnosis of ASD from the same cohort, presented in a separate article (Fletcher et al., 2019), it did show significantly reduced overall spindle power across Stages 2 and 3 when

Table 7. Means (and standard deviations) for the sleep parameters used to predict change in overnight performance, presented by group.

Group	Central spindle density		Frontal slow spindle power		Central fast spindle power		Frontal delta power	Central theta power
	Stage 2	Stage 3	Stage 2	Stage 3	Stage 2	Stage 3	Stage 3	REM
Low AI	7.59 (2.59)	1.46 (0.92)	2.51 (0.51)	2.203 (0.50)	1.57 (0.49)	1.21 (0.36)	7.74 (0.31)	2.98 (0.42)
High AI	7.66 (3.61)	1.25 (1.40)	2.36 (0.61)	2.219 (0.64)	1.33 (0.58)	1.05 (0.44)	7.64 (0.46)	2.85 (0.37)
t	−0.072	0.557	0.900	−0.091	1.458	1.319	0.761	1.185
All pts.	7.61 (2.95)	1.38 (1.11)	2.46 (0.55)	2.210 (0.55)	1.48 (0.53)	1.15 (0.39)	7.70 (0.37)	2.94 (0.41)

Note. REM = rapid eye movement; AI = Autism Index; All pts. = All participants.

compared to 28 typically developing controls. We present sleep data from all the children who contributed data to the SleepSmart sample and provide group differences for all sleep parameters considered across each paper via the Open Science Framework at https://osf.io/82eqm/files/.

In the case of the novel condition, the selected subset of predictors consisted of central Stage 2 spindle density, Stage 2 frontal slow spindle power, Stage 3 frontal delta, central REM theta, and age. A model with these predictors was able to significantly predict overnight change in performance on the novel condition, AdjR^2 = .225, $F(7, 39) = 2.909$, $p = .0151$, with central spindle density in Stage 2 ($B = 0.009$) and Stage 3 ($B = 0.047$), frontal slow spindle power in Stage 2 ($B = -0.071$) and Stage 3 ($B = -0.092$), frontal delta power in Stage 3 ($B = 0.093$), central REM theta ($B = 0.173$), and age ($B = -0.002$) emerging as significant predictors (see Table 8). Variance inflation factors were checked and did not exceed 3.8; variance inflation factor for REM theta power was 1.4. One participant was removed as overly influential based on DFBETAS. These data suggest that REM sleep contributes to the generalization of phonological learning in this task, as supported by REM theta power being a significant predictor of overnight change in performance on the novel condition when entered alone, $B = 0.122$, $t = 2.278$, $p = .0276$ (see Figure 6).

For overnight changes in trained performance (see Table 8), the selected subset of variables consisted only of

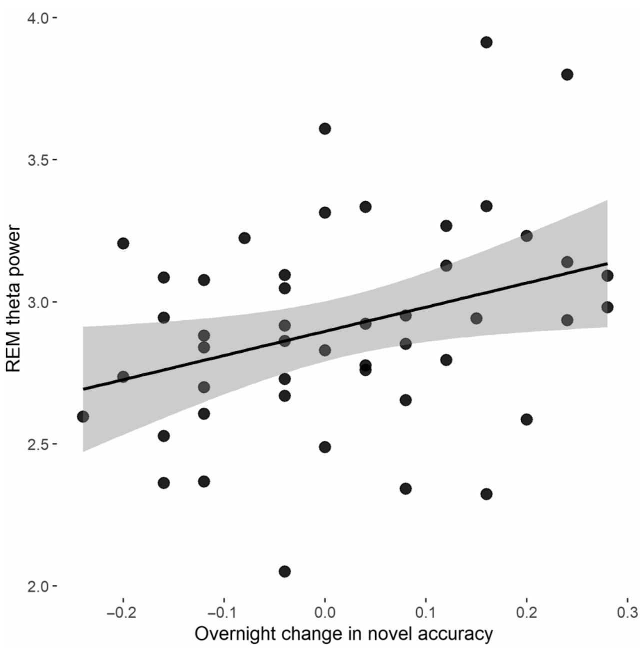
Table 8. Regression model describing relationship between change in performance on the novel and trained conditions overnight and sleep parameters.

Condition		*B*	*SE*	Lower 95% CI	Upper 95% CI	*t*
Novel	Intercept	−0.682	0.504	−1.670	0.307	−1.352
	Age (in months)	−0.002	0.001	−0.005	0.000	−1.880
	Language (composite)					
	Group (low–high AI)					
	S2 central spindle density	0.009	0.011	−0.013	0.031	0.776
	S2 frontal slow spindle power	−0.071	0.065	−0.197	0.056	−1.094
	S2 central fast spindle power					
	S3 central spindle density	0.047	0.029	−0.010	0.104	1.629
	S3 frontal slow spindle power	−0.092	0.061	−0.211	0.027	−1.508
	S3 central fast spindle power					
	S3 frontal delta power	0.093	0.067	−0.038	0.224	1.389
	REM central theta power	0.173	0.057	0.061	0.285	3.040**
Trained	Intercept	0.212	0.0791	0.057	0.367	2.680*
	Age (in months)					
	Language (composite)	−0.001	0.001	−0.003	0.001	−2.069*
	Group (low–high AI)					
	S2 central spindle density					
	S2 frontal slow spindle power					
	S2 central fast spindle power	−0.027	0.022	−0.070	0.016	−1.226
	S3 central spindle density					
	S3 frontal slow spindle power					
	S3 central fast spindle power					
	S3 frontal delta power					
	REM central theta power					

Note. Values are provided for any variables included in the regression model having run leaps subset selection with all eight predictors; cells have been grayed out where a variable was excluded by the leaps package on the account of not contributing explanatory power to the best fitting regression model. CI = confidence interval; AI = Autism Index; REM = rapid eye movement.

$^*p < .05$. $^{**}p < .01$.

Figure 6. Relationship between rapid eye movement (REM) theta power and overnight change in performance on the novel condition.



Language and Stage 2 Central Fast Spindle Power. These variables did not predict change, AdjR^2 = .062, $F(2, 45)$ = 2.55, p = .089, although Language Composite Score did emerge as a unique predictor of overnight change (B = −0.001). Looking at the extent to which performance at the Day1:PM session predicts performance at the Day2:AM session, we might expect this stark contrast between the predictive ability of sleep parameters across the two conditions. For the trained condition, performance at the Day2:AM session is well predicted by performance the evening before during the Day1:PM session (R^2 = .789), while predictive power for the novel condition is markedly lower (R^2 = .385), leaving considerably more variability to explain for the novel condition.

Discussion

In this study, we charted the time course of phonological learning and generalization over the course of 24 hr in a sample of children who varied with respect to structural language ability and autism symptomatology. Training was carried out with text-to-speech synthesized speech tokens, which required children to remap phonological representations to include new exemplars for existing phonological categories. On the night following training, children wore PSG sensors to measure nocturnal sleep parameters that may be associated with the consolidation of new phonological information. Successful performance on this task required children to shift decision boundaries in the identification of phonemes, somewhat like learning to listen to an unfamiliar speaker with a strong accent. The listener's task is further complicated in the case of synthetic speech as remapping rules are not systematic: Making adjustments for one phoneme does not help listeners tune in to another. When tested on novel items, the task required children to transfer their phonological learning to new tokens, necessitating phoneme-by-phoneme recognition rather than word-level auditory pattern recognition.

Training was effective in our paradigm, with children showing a substantial average improvement of around 30% on novel tokens (concrete nouns) pre- to posttraining. This training effect supports the effectiveness of clear and degraded feedback on performance (Hervais-Adleman et al., 2008). Over the course of the day following training, performance on the task as a whole did not change significantly. This was contrary to expectations based on Fenn et al.'s (2013) adult data, which showed a dip in performance on a generalization task over the course of a day, though the current task also differed from that of Fenn et al. in important ways (as outlined in the introduction). The lack of a significant reduction in performance may, however, represent a genuine developmental effect, as it is consistent with literature showing intersession decline in performance for adults but not children on linguistic tasks, including artificial grammar learning (Ferman & Karni, 2010) and nonword repetition (Bishop, Barry, & Hardiman, 2012). However, in order to determine whether differences in methodology or sample characteristics explain our results, children, adolescents, and adults would need to be tested using the same procedure.

Overnight, we saw an overall improvement in performance. This pattern could not be attributed to a sleep debt in the evening affecting attentional control, as performance on the PVT did not differ between sessions before and after sleep (Day1:PM vs. Day2:AM). Testing in the morning was conducted after children arrived at school, long enough after waking to avoid effects of sleep inertia (see Trotti, 2017), such that the decline seen in PVT reaction time seen after the first experimental session is likely explainable by task fatigue. No interaction emerged between overnight change and condition, suggesting that sleep promotes phonological learning in children for both trained items and generalization to novel items. This finding is a deviation from the specific effect of sleep found for the generalization of phonological information in adults (Fenn et al., 2013). This difference could be attributable to a developmental shift in the influence of sleep on the consolidation of different types of memory, with enhanced benefits of sleep for explicit aspects of task performance (Wilhelm et al., 2013) in children. Alternatively, we could attribute the contrast of our results with Fenn et al. to a difference in methodology. Fenn et al. utilized a between-subjects design such that participants in the rote condition only ever experienced rote items in training, while participants in the novel condition experienced a variety of items in training. Here, children were trained on multiple items but tested on both trained and novel items at each time point. This suggests that generalization is likely to contribute to performance in the trained condition and the novel condition. According to this explanation, sleep may act preferentially on generalization, rather than other types of phonological learning in childhood, but our within-participants design could not isolate generalization.

We hypothesized that performance on both trained and novel items would be associated with REM theta power and that generalization would be additionally related to spindle activity and power during NREM sleep stages. These hypotheses were based on the work of Earle and Myers (2014), suggesting that the exact relationships between sleep characteristics and phonological learning would depend heavily on the nature of the task itself given that phonological learning is not easily classified as either declarative or procedural in nature. As hypothesized, we saw that theta power in REM sleep predicted overnight change in performance on the novel condition. This finding is consistent with an active role for sleep in phonological generalization and, specifically, a role for synaptic consolidation in the type of phonological generalization shown here. Notably though, we did not find any evidence for a relationship between behavioral change and NREM sleep parameters.

This study is one of a small number to consider the role of sleep in phonological generalization (Earle & Myers, 2015; Fenn et al., 2013, 2003; Xie et al., 2016, 2018). Together, these articles pose the questions: What aspects of phonological generalization are sleep associated with, and if sleep actively supports such generalization, then what are the mechanisms of that support? In previous work, a consolidation period containing sleep has been shown to improve listeners' ability to generalize perceptual learning across speakers who show phonemic variability (Earle & Myers, 2015; Xie et al., 2016, 2018). In these studies, generalizing learning to a new speaker means being able to adjust specific category boundaries to allow for interspeaker variability, suggesting that sleep is relevant to the abstraction of higher level category-relevant phonemic information. In the task adopted by Fenn et al. (2013), adults were required to use the new phonological mappings to which they had been exposed to understand previously unheard synthesized words. Sleep benefited the recombination of those new mappings but did not support those in the rote-trained condition to extract any higher level information about the synthetic voice to generalize to new words containing previously unheard phonemes. In the current study, the training that children experienced included comprehensive exposure across all phonemes necessary to succeed at test.[2] Children did not have to adapt the new mappings when tested but rather recombine them to understand new words (with the exception of allowing for phonological assimilation). Collectively, these results beg the question: What is sleep doing exactly?

The hypothesis that REM sleep might relate to behavioral change on this paradigm was based on the understanding that the task as a whole required participants to shift attention toward relevant acoustic features of speech stimuli in order to learn new phonological mappings. The exact association we found, however, was between theta power in REM sleep and the ability to recombine those newly mapped phonological categories. Unfortunately, we were unable to properly assess how specific this relationship was as overnight change in performance on trained items was not associated with any sleep variables. The most likely reason for this is that performance after sleep was very well predicted by presleep performance, and for children with superior language ability, performance neared ceiling levels after sleep such that interparticipant variability in overnight change may have been curtailed, reducing the likelihood of seeing relationships with sleep parameters. Change in performance on the trained condition was only predicted (negatively) by language ability, with children who had poorer language ability also having more room left to improve on the task. This tendency toward ceiling effects in the trained condition on Day 2 may also have prohibited an interaction between condition and overnight change, which, had it emerged, would have demonstrated a sleep advantage for word-level auditory pattern recognition in addition to generalization.

That sleep plays a role in phonological generalization in children adds to our knowledge and understanding of the mechanisms of sleep and its role in the consolidation of new knowledge at this age. Here, sleep can be seen to support perceptual learning in the phonological domain, to help the system remain stable and flexible as it encounters new items, thereby supporting adaptation to the environment. Interestingly, over the course of 4 weeks, learning remained relatively stable, with performance at the follow-up session showing the same pattern as after encoding; performance level also appeared stable (though this was not tested statistically), suggesting that perceptual learning in the phonological domain shows good retention over weeks despite no intervening practice or relevance. Similar stability of perceptual learning after training with synthesized speech has been demonstrated in adults (Schwab, Nusbaum, & Pisoni, 1985).

This study aimed to consider phonological learning in children who differed with respect to the stability and flexibility of phonological representations. It was hypothesized that children with poorer structural language would perform less well than their peers overall and show a slower rate of learning on the account of having less stable phonological representations to map new exemplars on to. In reality, many different aspects of the language system could lead to poorer performance on this task. For example, vocabulary knowledge may impact on phonological processing given that semantic and phonological representations are interdependent (see McClelland, Mirman, & Holt, 2006). We tried to ensure that all items would be familiar to all children, but the extent of familiarity, the depth of semantic knowledge, and the speed of lexical retrieval are likely to have varied considerably in this sample. Unfortunately, we were unable to verify this given the

[2]During training, all children were exposed to all 44 phonemes in the English language, with the exceptions that one group did not hear /ʒ/ and two groups did not hear /ʊ/.

large number of items presented to participants over the course of the study. Language ability emerged as a strong predictor of overall task performance in all models but did not predict change in performance, with the exception of overnight change in the trained condition, as discussed above.

In individuals with high autism traits, we hypothesized we would see poorer performance on novel items compared to trained items, reflecting a hypothesized tendency in those with ASD to allocate greater attentional resources to features that are unique to individual items and less to features common across items compared to typically developing peers (Plaisted, 2001). We did not find an overall group difference here; indeed, phonological learning more broadly is often a relative strength in the language profiles of children with ASD, with some studies showing enhanced performance on tasks that demand the learning of new phonological forms (e.g., Henderson, Powell, Gaskell, & Norbury, 2014; Norbury, Griffiths, & Nation, 2010). However, we did observe a deficit in generalization (performance on the novel condition) for those individuals with high autism symptomology and low language ability. Language ability constrained performance on the novel items for those with high autism traits. If difficulties with generalization do exist in this population, enhanced phonological skills may act as a protective factor in this domain for some individuals. Finally, no group differences were observed with respect to the specific sleep parameters of interest between children with high and low autism symptomatology in this sample (though see https://osf.io/82eqm/files/ for sleep results from the full sample). Given that these sleep parameters explained variance in overnight change in generalization, it is perhaps not surprising that group differences in behavior did not emerge here.

Limitations and Conclusions

The findings of this study should be considered in the light of its main limitation: the nature of the atypical samples recruited to this study. Children recruited on the account of falling along the autism spectrum did not have severe symptomatology. This recruitment bias was expected given that children with severe ASD symptoms often show hypersensitivity to tactile stimuli, in particular touching of the head or body (see Marco, Hinkley, Hill, & Nagarajan, 2011, for a review), such that the overnight electrophysiological measurement would be difficult to tolerate for many. We were also keen to only involve children who we felt confident could give informed oral assent to the procedure. The future inclusion of children with more severe symptoms in a behavioral study of phonological generalization could address more clearly the flexibility of phonological representations in this population. We were also only able to include a relatively small number of children here with severe language difficulties due to issues with recruitment. This is likely to have limited the extent to which stability of phonological representations varied over the sample. Children with language disorders are believed to show some degree of behavioral sleep problems (Botting & Baraka, 2018; Dominick et al., 2007), but no studies have examined the architecture of sleep in this population in over 20 years (Duvelleroy-Hommet et al., 1995; Picard et al., 1998). More thorough consideration of how phonological learning relates to sleep in this population may therefore be crucial in trying to understand the long-term nature of phonological development in these children.

This study considered the role of sleep in phonological learning in children who varied as a function of structural language skills and autism symptomatology. We showed that sleep is associated with phonological learning in childhood, with phonological generalization being supported by theta power in REM sleep. Language skill was found to predict overall task performance, although the role of REM sleep did not differ as a function of language ability. This work adds to a growing literature exploring the importance of sleep for the stability of new learning and the integration of new representations into existing networks in childhood.

Acknowledgments

This work was supported by an Economic and Social Research Council grant (reference: ES/N009924/1) awarded to Lisa Henderson, Gareth Gaskell, and Courtenay Norbury. We thank Alex Bettarini, Alex Bond, Lolly Hernandez, Natasha Thompson, Amanda Olsson, and Lois Perry for helping with data collection and the children, families, and schools who agreed to participate.

References

Achenbach, T. M., & Edelbrock, C. S. (1983). *Manual for the child behavior checklist and revised child behavior profile* (p. 230). Burlington: University of Vermont/Department of Psychiatry.

American Academy of Sleep Medicine. (2016). *The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications*. Darien, IL: Author.

Anthony, J. L., Williams, J. M., Aghara, R. G., Dunkelberger, M., Novak, B., & Mukherjee, A. D. (2010). Assessment of individual differences in phonological representation. *Reading and Writing, 23,* 969–994.

Audacity Team. (1999–2016). Audacity (Version 2.1.3) [Audio editor and recorder]. Retrieved from http://audacityteam.org/

Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the psychomotor vigilance task (PVT) to sleep loss. *Sleep, 34*(5), 581–591.

Basner, M., Mollicone, D., & Dinges, D. F. (2011). Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronautica, 69,* 949–959.

Bates, D., Mächler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes* (R Package Version 0.999999-0). Retrieved from http://cran.r-project.org/web/packages/lme4/index.html

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Batterink, L. J., Oudiette, D., Reber, P. J., & Paller, K. A. (2014). Sleep facilitates learning a new linguistic rule. *Neuropsychologia, 65,* 169–179.

Batterink, L. J., & Paller, K. A. (2017). Sleep-based memory processing facilitates grammatical generalization: Evidence from targeted memory reactivation. *Brain and Language, 167,* 83–93.

Belsley, D., Kuh, E., & Welsch, R. (1980). *Regression diagnostics. Identifying influential data and sources of collinearity* (p. 292). New York, NY: Wiley.
Bishop, D. V. M. (2003). *Children's Communication Checklist–Second Edition (CCC-2)*. London, United Kingdom: Pearson.
Bishop, D. V. M., Barry, J. G., & Hardiman, M. J. (2012). Delayed retention of new word-forms is better in children than adults regardless of language ability: A factorial two-way study. *PLOS ONE, 7*(5), e37326.
Botting, N., & Baraka, N. (2018). Sleep behaviour relates to language skills in children with and without communication disorders. *International Journal of Developmental Disabilities, 64,* 238–243. https://doi.org/10.1080/20473869.2017.1283766
Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behaviour Research Methods, 41*(4), 977–990.
Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand known English word lemmas. *Behavioural Research Methods, 46*(3), 904–911.
Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience, 13*(11), 1428–1432.
Coady, J. A., Kluender, K. R., & Evans, J. L. (2005). Categorical perception of speech by children with specific language impairments. *Journal of Speech, Language, and Hearing Research, 48*(4), 944–959.
Collet, G., Schmitz, R., Urbain, C., Leybaert, J., Colin, C., & Peigneux, P. (2012). Sleep may not benefit learning new phonological categories. *Frontiers in Neurology, 3,* 97. https://doi.org/10.3389/fneur.2012.00097
Constantino, J. N., Yang, D., Gray, T. L., Gross, M. M., Abbacchi, A. M., Smith, S. C., . . . Kuhl, P. K. (2007). Clarifying the associations between language and social development in autism: A study of non-native phoneme recognition. *Journal of Autism and Developmental Disorders, 37*(7), 1256–1263.
Cox, R., Hofman, W. F., & Talamini, L. M. (2012). Involvement of spindles in memory consolidation is slow wave sleep-specific. *Learning and Memory, 19,* 264–267.
Crane, L., Chester, J. W., Goddard, L., Henry, L. A., & Hill, E. (2015). Experiences of autism diagnosis: A survey of over 1000 parents in the United Kingdom. *Autism, 20*(2), 153–162.
Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*(2), 114–126.
Díaz-Román, A., Zhang, J., Delorme, R., Beggiato, A., & Cortese, S. (2018). Sleep in youth with autism spectrum disorders: Systematic review and meta-analysis of subjective and objective studies. *Evidence-Based Mental Health, 21*(4), 146–154.
Dominick, K. C., Ornstein Davis, N., Lainhart, J., Tager-Flusberg, H., & Folstein, S. (2007). Atypical behaviors in children with autism and children with a history of language impairment. *Research in Developmental Disabilities, 28,* 145–162.
Dunn, L. M., Dunn, D. M., & Styles, B. (2009). *British Picture Vocabulary Scale–Third Edition (BPVS-3)*. London, United Kingdom: GL Assessment.
Duvelleroy-Hommet, C., Billard, C., Lucas, B., Gillet, P., Barthez, M. A., Santini, J. J., . . . Autret, A. (1995). Sleep EEG and developmental dysphasia: Lack of a consistent relationship with paroxysmal EEG activity during sleep. *Neuropediatrics, 26*(1), 14–18.
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., Landi, N., & Myers, E. B. (2018). Adults with specific language impairment fail to consolidate speech sounds during sleep. *Neuroscience Letters, 666,* 58–63.
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. (2015). Overnight consolidation promotes generalization across talkers in the identification of non-native speech sounds. *The Journal of the Acoustical Society of America, 137*(1), EL91–EL97. https://doi.org/10.1121/1.4903918
Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America, 119*(4), 1950–1953.
Elliott, C. D., & Smith, P. (2011). *British Ability Scales–Third Edition (BAS-3)*. London, United Kingdom: GL Assessment.
Elrod, M. G., & Hood, B. S. (2015). Sleep differences among children with autism spectrum disorders and typically developing peers: A meta-analysis. *Journal of Developmental and Behavioral Pediatrics, 36*(3), 166–177.
Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2013). Sleep restores loss of generalised but not rote learning of synthetic speech. *Cognition, 128,* 280–286.
Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature, 425,* 614–616.
Ferman, S., & Karni, A. (2010). No childhood advantage in the acquisition of skill in using an artificial language rule. *PLOS ONE, 5*(10), e13648.
Fletcher, F., Knowland, V. C. P., Walker, S., Gaskell, M. G., Norbury, C., & Henderson, L.-M. (2019). Atypicalities in sleep and semantic consolidation in autism. *Developmental Science.* in press.
Francis, A. L., Baldwin, K., & Nusbaum, H. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics, 62*(8), 1668–1680.
Gagnon, R. T. (1978). Votrax real time hardware for phoneme synthesis of speech. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), 78,* 175–178.
Gilliam, J. E. (2013). *Gilliam Autism Rating Scale–Third Edition (GARS-3)*. London, United Kingdom: Pearson.
Godbout, R., Bergeron, C., Limoges, E., Stip, E., & Mottron, L. (2000). A laboratory study of sleep in Asperger's syndrome. *NeuroReport, 11*(1), 127–130.
Hazan, V. L., & Barett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics, 28*(4), 377–396.
Henderson, L., Devine, K., Weighall, A., & Gaskell, G. (2015). When the daffodat flew to the intergalactic zoo: Off-line consolidation is critical for word learning from stories. *Developmental Psychology, 51*(3), 406–417.
Henderson, L., Powell, A., Gaskell, M., & Norbury, C. (2014). Learning and consolidation of new spoken words in autism spectrum disorder. *Developmental Science, 17,* 858–871.
Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science, 15*(5), 674–687. https://doi.org/10.1111/j.1467-7687.2012.01172.x
Hervais-Adleman, A., Davis, M. H., Johnstrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 460–474.
Järvinen-Pasley, A., Wallace, G. L., Ramus, F., Happé, F., & Heaton, P. (2008). Enhanced perceptual processing of speech in autism. *Developmental Science, 11*(1), 109–121.

Joanisse, M. F. (2004). Specific language impairments in children: Phonology, semantics and the English past tense. *Current Directions in Psychological Science, 13*(4), 156–160.

Joanisse, M. F., & Seidenberg, M. S. (2003). Phonology and syntax in specific language impairment: Evidence from a connectionist model. *Brain and Language, 86,* 40–56.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122*(2), 148–203.

Knowland, V. C., Evans, S., Snell, C., & Rosen, S. (2016). Visual speech perception in children with language learning impairments. *Journal of Speech, Language, and Hearing Research, 59,* 1–14.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978–990.

Latchourmane, C. V., Ngo, H. V., Born, J., & Shin, H. S. (2017). Thalamic spindles promote memory formation during sleep through triple phase-locking of cortical, thalamic, and hippocampal rhythms. *Neuron, 95*(2), 424.e6–435.e6.

Lenth, R., Singmann, H., & Buerkner, P. (2019). *emmeans: Estimated marginal means, aka least-squares means* (R package Version 1.3.5). Retrieved from http://CRAN.R-project.org/package=emmeans

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54,* 358–368.

Limoges, E., Mottron, L., Bolduc, C., Berthiaume, C., & Godbout, R. (2005). Atypical sleep architecture and the autism phenotype. *Brain, 128,* 1049–1061.

Lumley, T. (2017). *Leaps: Regression on subset selection* (R package Version 3.0) [based on Fortran code by Alan Miller]. Retrieved from http://CRAN.R-project.org/package=leaps

Maillart, C., Schelstraete, M. A., & Hupet, M. (2004). Phonological representations of children with SLI: A study of French. *Journal of Speech, Language, and Hearing Research, 47,* 187–198.

Marco, E. J., Hinkley, L. B., Hill, S. S., & Nagarajan, S. S. (2011). Sensory processing in autism: A review of neurophysiologic findings. *Pediatric Research, 69*(5), 48R–54R.

Maski, K., Holbrook, H., Manoach, D., Hanson, E., Kapur, K., & Stickgold, R. (2015). Sleep dependent memory consolidation in children with autism spectrum disorder. *Sleep, 38*(12), 1955–1963.

MathWorks. (2017). MATLAB Release 2017a. Natick, MA: Author.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception. *Trends in Cognitive Science, 10*(8), 363–369.

Mielke, J., Nielsen, K., & Magloughlin, L. V. (2013). Phonetic imitation by individuals with autism spectrum disorders: Investigating the role of procedural and declarative memory. *Proceedings of Meetings on Acoustics, 19,* 060142.

Mines, M. A., Hanson, B. F., & Shoup, J. E. (1978). Frequency of occurrence of phonemes in conversational English. *Language and Speech, 21*(3), 221–241.

Nieuwenhuis, R., te Grotenhuis, M., & Pelzer, B. (2012). influence.ME: Tools for detecting influential data in mixed effects models. *The R Journal, 4,* 38–47. Retrieved from https://journal.r-project.org/archive/2012/RJ-2012-011/RJ-2012-011.pdf

Norbury, C. F., Griffiths, H., & Nation, K. (2010). Sound before meaning: Word learning in autistic disorders. *Neuropsychologia, 48*(14), 4012–4019.

Ohayon, M. M., Carskadon, M. A., Guilleminault, C., & Vitiello, M. V. (2004). Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: Developing normative sleep values across the human lifespan. *Sleep, 27*(7), 1255–1273.

Owens, J. A., Spirito, A., & McGuinn, M. (2000). The Children's Sleep Habits Questionnaire (CSHQ): Psychometric properties of a survey instrument for school-aged children. *Sleep, 23*(8), 1043–1051.

Picard, A., Cheliout-Heraut, F., Bouskraoui, M., Lemione, M., Lacert, P., & Dalattre, J. (1998). Sleep EEG and developmental dysphasia. *Developmental Medicine and Child Neurology, 40,* 595–599.

Plaisted, K. (2001). Reduced generalisation in autism: An alternative to weak central coherence. In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 149–169). Mahwah, NJ: Erlbaum.

Psychology Software Tools. (2012). E-Prime 2.0. Retrieved from http://www.pstnet.com

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Robertson, E. K., Joanisse, M. F., Desroches, A. S., & Ng, S. (2009). Categorical speech perception deficits distinguish language and reading impairments in children. *Developmental Science, 12*(5), 753–767.

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27,* 395–408.

Semel, E., & Wiig, E. (2006). *Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4)*. London, United Kingdom: Pearson.

Stackhouse, J., & Wells, B. (1997). *Children's speech and literacy difficulties: A psycholinguistic framework*. London, United Kingdom: Whurr.

Staresina, B. P., Bergmann, T. O., Bonnefond, M., van der Meij, R., Jensen, O., Deuker, L., . . . Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nature Neuroscience, 18*(11), 1679–1686.

Stark, R. E., & Heinz, J. M. (1996). Perception of stop consonants in children with expressive and receptive–expressive language impairments. *Journal of Speech and Hearing Research, 39*(4), 676–686.

Stewart, M. E., Petrou, A. M., & Ota, M. (2018). Categorical speech perception in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders, 48*(1), 72–82.

Tamaki, M., Watanabe, T., & Sasaki, Y. (2017). REM sleep stabilizes visual perceptual learning which was rendered fragile by NREM sleep. *Sleep, 40*(1), A85.

Tessier, S., Lambert, A., Chicoine, M., Scherzer, P., Soulières, I., & Godbout, R. (2015). Intelligence measures and Stage 2 sleep in typically developing and autistic children. *International Journal of Psychophysiology, 97*(1), 58–65.

Trotti, L. M. (2017). Waking up is the hardest thing I do all day: Sleep inertia and sleep drunkedness. *Sleep Medicine Reviews, 35,* 76–84.

Tsanas, A., & Gifford, G. D. (2015). Stage-independent, single lead EEG sleep spindle detection using the continuous wavelet transform and local weighted smoothing. *Frontiers in Human Neuroscience, 8*(9), 181.

UCLA Statistical Consulting Group. (n.d.). *Introduction to the features of SAS: SAS learning modules*. Retrieved from https://stats.idre.ucla.edu/sas/modules/introduction-to-the-features-of-sas/

Vanderwalle, E., Boets, B., Ghesquière, P., & Zink, I. (2012). Auditory processing and speech perception in children with specific language impairment: Relations with oral language and literacy skills. *Research in Developmental Disabilities, 33,* 635–644.

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behaviour Research Methods, Instruments, & Computers, 36*(3), 481–487.

Wagner, R., Torgesen, J., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing–Second Edition (CTOPP-2)*. Oxford, United Kingdom: Pearson.

Wang, X., Wang, S., Fan, Y., Huang, D., & Zhang, Y. (2017). Speech-specific categorical perception deficit in autism: An event-related potential study of lexical tone processing in Mandarin-speaking children. *Scientific Reports, 7,* 43254.

Wei, Y., Krishnan, G. P., Komarov, M., & Bazhenov, M. (2018). Differential roles of sleep spindles and slow oscillations in memory consolidation. *PLoS Computational Biology, 14*(7), e1006322. https://doi.org/10.1371/journal.pcbi.1006322

Werchen, D. M., & Gómez, R. L. (2013). Generalizing memories over time: Sleep and reinforcement facilitate transitive inference. *Neurobiology of Learning and Memory, 100,* 70–76.

Whitteman, M. J., Bardhan, N. P., Weber, A., & McQueen, J. M. (2015). Automaticity and stability of adaptation to a foreign accented speaker. *Language and Speech, 58*(2), 168–189.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* New York, NY: Springer-Verlag.

Wilhelm, I., Rose, M., Imhof, K. I., Rasch, B., Büchel, C., & Born, J. (2013). The sleeping child outplays the adult's capacity to convert implicit into explicit knowledge. *Nature Neuroscience, 16,* 391–393.

Williams, D., Botting, N., & Boucher, J. (2008). Language in autism and specific language impairment: Where are the links. *Psychological Bulletin, 134*(6), 944–963.

Xie, X., Earle, F. S., & Myers, E. B. (2016). Sleep talker generalization of accent adaptation. *The Journal of the Acoustical Society of America, 140*(4), 3342.

Xie, X., Earle, F. S., & Myers, E. B. (2018). Sleep facilitates generalisation of accent adaptation to a new talker. *Language, Cognition, and Neuroscience, 33*(2), 196–210.

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F.-X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences of the United States of America, 102*(39), 14110–14115.

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2011). Noise on, voicing off: Speech perception deficits in children with specific language impairment. *Journal of Experimental Child Psychology, 110,* 362–372.

Appendix (p. 1 of 2)

Word Lists

	Target	Distractor		Target	Distractor		Target	Distractor
1	hat	hand	**4**	clock	crown	**7**	violin	vegetable
	boots	bee		grapes	gold		bath	bag
	watch	wasp		owl	oar		tree	tie
	axe	arch		ghost	globe		shirt	skis
	gate	glove		slide	square		seal	scarf
	skirt	sun		dice	drum		stick	sword
	shelf	shorts		broomstick	butter		donut	dollhouse
	shoe	ship		baby	bubble		pizza	puddle
	dancer	dolphin		suitcase	seesaw		bottle	blanket
	planet	puppy		dragon	doctor		ketchup	kitten
	towel	tongue		biscuit	balloon		salad	slipper
	noodles	nuggets		blindfold	boxer		garlic	grapefruit
	teacher	teacup		cherry	chicken		raincoat	rhino
	backpack	birdhouse		tiger	toilet		parrot	popcorn
	treasure	toothbrush		toaster	toothpaste		pebble	penguin
	swimsuit	sweatshirt		ladder	lemon		glasses	guitar
	shampoo	shower		cabbage	candle		perfume	pumpkin
	policeman	ponytail		garage	goldfish		atlas	armour
	dustbin	doorbell		sailor	sandbox		circle	celery
	peanuts	pushchair		cave	comb		blackboard	beanbag
	magician	marshmallow		cereal	submarine		scorpion	sharpener
	aeroplane	ambulance		bicycle	buffalo		radio	recorder
	sunglasses	screwdriver		apricot	astronaut		ladybird	letterbox
	tomato	triangle		dinosaur	dalmatian		sunflower	spiderweb
	fireworks	flowerpot		television	toilet paper		bellybutton	binoculars

Appendix (p. 2 of 2)

Word Lists

	Target	Distractor		Target	Distractor		Target	Distractor
2	teeth	toast	**5**	cow	can	**8**	wheelbarrow	watermelon
	bread	bus		duck	doll		house	heart
	tent	truck		shell	shed		train	toad
	door	dress		mouse	moon		spoon	swan
	sea	smile		phone	fist		orange	otter
	witch	wheel		kite	king		spider	sofa
	goat	glass		hedgehog	hotdog		diver	diamond
	flag	fish		crayon	camera		jelly	giant
	sponge	straw		tractor	trainers		puzzle	pasta
	jewel	jeans		beetle	bacon		fireman	fishbowl
	lion	letter		milkshake	melon		clover	cushion
	wallet	walrus		dentist	doughnut		curtain	cowboy
	seahorse	snowflake		earthworm	eagle		teapot	toolbox
	circus	spaceship		elbow	eyebrow		purple	paper
	breakfast	beehive		artist	anchor		plaster	puppet
	apple	ankle		turtle	turkey		pencil	pirate
	window	waffle		taxi	tissue		ice cream	iron
	staircase	sparkler		carrot	cupcake		beach ball	baboon
	starfish	strawberry		radish	robot		hairbrush	highchair
	burglar	bookcase		paintbrush	peacock		flower	fairy
	photograph	fire truck		mushroom	mermaid		library	licorice
	storybook	skeleton		flamingo	family		spaghetti	cinema
	telescope	trampoline		dragonfly	dandelion		chocolate	chimney
	motorbike	magazine		jellyfish	gingerbread		pineapple	pajamas
	peach	pig		alien	alarm clock		sandcastle	centipede
3	sink	saw	**6**	kiss	key	**9**	castle	canoe
	leaf	lime		whale	wand		ant	arm
	crab	cake		soap	sock		plate	pond
	fox	fork		pear	plug		bucket	beaver
	desk	dummy		snail	snake		fossil	footprint
	shark	shield		sheep	star		rainbow	racket
	sweets	swings		crisps	car		lipstick	lizard
	princess	poodle		mountain	monster		lunchbox	light bulb
	trophy	trumpet		laces	lorry		helmet	hammer
	pillow	pancake		flip-flops	finger		sandwich	sausage
	petal	panda		snowball	scissors		angel	acorn
	daisy	donkey		pretzel	playground		rattle	record
	badger	bagpipe		lettuce	lightning		skateboard	snowman
	rocket	rubber		moustache	mattress		rabbit	robin
	arrow	apron		football	feather		necklace	needle
	table	teabag		bathroom	brownie		cartoon	camel
	tadpole	teepee		yoyo	yoghurt		present	pocket
	lighthouse	leopard		zebra	zipper		giraffe	gokart
	horseshoe	hamster		jacket	juggler		rose	rat
	bracelet	bandage		scarecrow	sandals		dishwasher	domino
	hamburger	hospital		blueberry	basketball		waterfall	wheelchair
	rocking chair	rectangle		banana	bulldozer		crocodile	cucumber
	barbeque	butterfly		lemonade	lollipop		caterpillar	cauliflower
	bear	bat		elephant	eskimo		cat	clown
	triceratops	tarantula		raspberry	roller skate		chips	chess