

Research article

Sleep duration predicts behavioral and neural differences in adult speech sound learning



F. Sayako Earle^{a,b,*}, Nicole Landi^{c,d}, Emily B. Myers^{a,c,d}

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

^b Department of Communication Sciences and Disorders, University of Delaware, Newark, DE, USA

^c Department of Psychology, University of Connecticut, Storrs, CT, USA

^d Haskins Laboratories, New Haven, CT, USA

HIGHLIGHTS

- Sleep duration predicts improvement in measures of speech sound learning.
- Behavioral gains are associated with changes in ERP response magnitude overnight.
- Duration of sleep determines relative success of forming new perceptual categories.

ARTICLE INFO

Article history:

Received 24 June 2016

Received in revised form 12 October 2016

Accepted 22 October 2016

Available online 26 October 2016

Keywords:

Memory consolidation

Speech perception

Perceptual learning

ABSTRACT

Sleep is important for memory consolidation and contributes to the formation of new perceptual categories. This study examined sleep as a source of variability in typical learners' ability to form new speech sound categories. We trained monolingual English speakers to identify a set of non-native speech sounds at 8PM, and assessed their ability to identify and discriminate between these sounds immediately after training, and at 8AM on the following day. We tracked sleep duration overnight, and found that light sleep duration predicted gains in identification performance, while total sleep duration predicted gains in discrimination ability. Participants obtained an average of less than 6 h of sleep, pointing to the degree of sleep deprivation as a potential factor. Behavioral measures were associated with ERP indexes of neural sensitivity to the learned contrast. These results demonstrate that the relative success in forming new perceptual categories depends on the duration of post-training sleep.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Perceptual categories are the foundation for mental organization of the external world and underlie complex cognitive operations such as object recognition and linguistic communication. The capacity to acquire *new* perceptual categories is fundamental to learning. A core question in cognitive science is what kind of experiences facilitates category acquisition. Recent work has pointed to the importance of memory consolidation during sleep in certain types of perceptual learning. The present study asks whether sleep behavior can account for the variability that exists in the ability to acquire new speech sound categories.

Memory processes that are involved in capturing and retaining acoustic information aggregate one's experience with spoken language into functional acoustic-phonetic units, speech sounds such as /d/ or /a/. Variations in the quality of these speech sound representations have been linked to differences in language and reading ability [1,2]. While differences in linguistic input are often cited as the source of such perceptual variability [3], we propose that variance in the representation of speech sounds amongst typical learners can in part be attributed to differences in sleep. Sleep's importance in memory consolidation has been documented across many domains [4], and sleep appears to play a role in the organization of speech information as well. Prior work suggests that a period of sleep following perceptual training is associated with generalization of speech information to unfamiliar lexical contexts in mapping synthetic tokens to native categories [5]. In the acquisition of nonnative speech sounds, the effect of sleep appears to be in protecting phonetic features from conflicting acoustic

* Corresponding author at: University of Delaware, STAR Health Sciences Complex, 540 S College Ave, Ste. 210BB, Newark, DE 19713, USA.
E-mail address: fsearle@udel.edu (F.S. Earle).

information [6], and in the generalization of acoustic-phonetic features to an unfamiliar talker [7]. In these prior investigations, we have found the magnitude of the effect of offline consolidation on nonnative category acquisition to be highly variable between individuals. One possibility is that the magnitude of overnight gain is dependent on the amount of learning achieved during training. Another possibility, based on a study of word learning that showed that larger gains in learning were associated with longer sleep duration [8], is that the length of post-training sleep determines the degree of improvement. Non-rapid eye movement (NREM) sleep is considered critically important for the consolidation of hippocampal-encoded memory [9]. Specifically, it has been suggested that light sleep (NREM stages 1 and 2) is responsible for active potentiation of wake-state memory, in contrast with deep sleep's (NREM stage 3) role in maintaining homeostatic regulation [10].

The aim of the current study is to test the relationship between sleep behavior and non-native speech sound learning. At 8 P.M. on day 1, we trained monolingual English speakers with typical language and reading ability to identify voiced Hindi dental and retroflex consonants. The dental sound is produced with the tongue tip placed behind the teeth, and the retroflex with the tongue curled behind the alveolar ridge, and both sounds perceptually assimilate to the English (alveolar)/d/ for most monolingual speakers of English. Participants were trained by the experimenter to use a sleep-monitoring headband [11] prior to leaving the lab on the first day (see [12] for reliability of the device), wore the device overnight, and returned at 8 A.M. the next morning for reassessment. We measured learning of the non-native contrast through two tasks: discrimination and identification. In discrimination, the listener compares two tokens to determine if the speech features are categorically similar or dissimilar, but this skill does not require awareness of category identity. In contrast, identification requires an explicit recall of a category label for tokens falling within a specific distribution of auditory stimuli. Therefore, the changes in performance on these perceptual tasks reflect different types of perceptual learning that are both commonly used to describe the representational quality of speech sounds. We tracked performance in each task on days 1 and 2 (see Fig. 1 for scheduling of tasks). We predicted that sleep duration would be positively associated with gains in behavioral performance.

It is possible that discrimination and identification tasks underestimate participants' ability to perceive differences among the trained sounds, or that motivational factors that vary as a function of fatigue might affect task performance. Thus, in order to obtain a task-independent measure of changes to perceptual sensitivity, we also collected electrophysiological recordings during an oddball paradigm to determine each participant's mismatch negativity (MMN) response to the trained contrast immediately before training on day 1, and after behavioral reassessment on day 2. The MMN component of the event-related potential (ERP) reflects pre-attentive responses to stimulus change, and has been used to measure phonological processing in both typical and atypical populations [13]. We predicted that overnight changes in the MMN response would be positively associated with changes in behavioral performance.

2. Materials and methods

2.1. Participants

All participants provided informed written consent in accordance with our human subjects research protocol approved by the University of Connecticut Institutional Review Board prior to participation. The data presented in this manuscript is from a subset

from a larger dataset collected from monolingual, native speakers of American English, 18–24 years of age, who grew up in a household with only native speakers of American English. The participants included in the current report met additional criteria as follows. Participants reported no neurological, socio-emotional, or attention disorder, and a history of typical language, reading, and cognitive development. In addition, participants passed a pure tone hearing screening and obtained scores at or above the 25th percentile on the following standardized assessments of reading and nonverbal cognition at the time of the study: *Woodcock Reading Mastery Test-Revised* [14], *Test of Word Reading Efficiency* [15], *Wechsler Abbreviated Scale of Intelligence* [16]. These participants were further identified as 'typical' in language by the Modified Token Test and the 15-word spelling test, which are non-standardized measures devised to identify adults with a history of language impairment [17]. Twenty-seven participants (Mean: 20 years, 5 months; SD: 16 months; 17 female, 25 right-handed) met all criteria and completed the EEG recording and the perceptual training; of these, data from two participants were excluded from the current analyses for the reason that behavioral performance immediately after training was greater than 3 standard deviations above the mean; this brought their scores close to task ceiling (equal to or greater than 95% accurate) and as such, a meaningful measure of overnight improvement on perceptual tasks could not be obtained. Data on the remaining 25 participants contributed behavioral/EEG measures. For analyses involving sleep data, we excluded data points that revealed missing segments due to signal dropout during the recording period ($n = 7$). The remaining 18 (Mean: 20 years, 4 months; SD: 17 months; 11 female, 16 right-handed) completed sleep data collection.

Standardized assessments were administered and scored by the first author or one other graduate student, and rescored by one of two trained undergraduate students. Any discrepancies in scoring were flagged by the second scorer and resolved by the first author.

2.2. Stimuli

2.2.1. Nonnative phonetic training/assessment

A set of 10 naturally spoken tokens (5 each: /d̪ʌg/ [dental], /d̪ʌg/[retroflex]) were produced by a male, native speaker of Hindi in a sound attenuated audiology booth and recorded by an Edirol digital recorder [18]. Praat software [19] was used to cut tokens to the burst onset and to adjust mean amplitudes to 70 dB. For training, two novel visual objects [20] were used, each corresponding to one 'word' within the minimal pair: /d̪ʌg/-/d̪ʌg/ used during training.

2.2.2. Mismatch negativity (MMN) response

Two tokens from the above training set (one dental and one retroflex) were further modified in Praat to equate burst, vowel, and token duration to total 250 ms from onset to offset. The dental token was used as the frequently occurring Standard, and the retroflex token was used as the infrequently occurring Deviant.

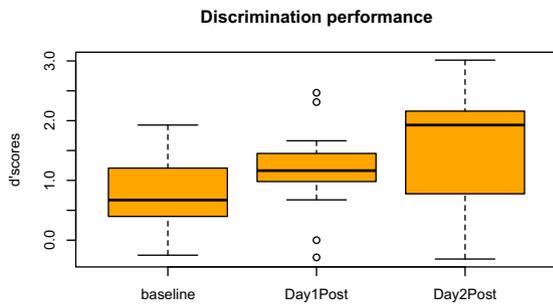
2.3. Procedure

Participants were scheduled for the experiment on two consecutive days: from 7 to 9 P.M. on day 1, and 8 to 10 A.M. on day 2 (Fig. 1). At the end of the first session, participants were instructed on the use of the sleep-monitoring headband [11,12], and were sent home with the device. Participants returned at 8 A.M. the next day and were first reassessed behaviorally on their perception of the non-native contrast. This was followed by a second EEG/ERP recording session, and then by administration of the remaining language and cognitive assessments.

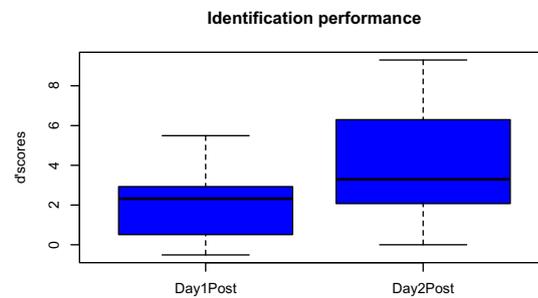
(A)

Day 1 (7-9 PM)			Day 2 (8-10AM)		
7-7:30PM	7:30-8PM	8-9PM	8-8:30AM	7:30-8PM	7-7:30PM
Enrollment Language Assessments	EEG/ERP	Discrimination baseline Identification Training Identification Day 1 Post Discrimination Day 1 Post	Identification Day 2Post Discrimination Day 2Post	EEG/ERP	Cognitive/ Reading assessment

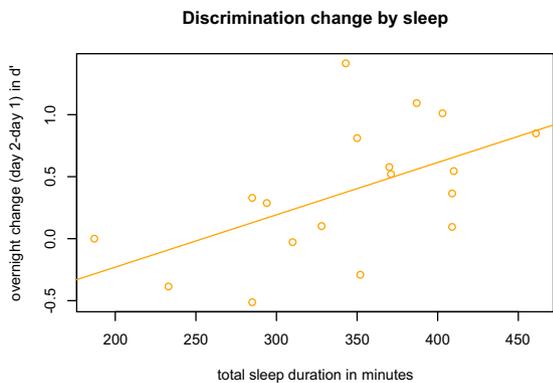
(B)



(C)



(D)



(E)

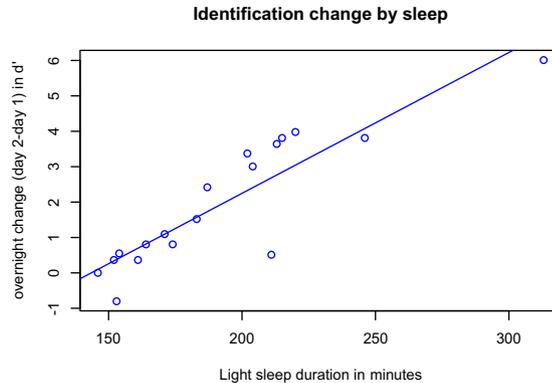


Fig. 1. Perceptual performance over time and relationships with sleep duration.

(A) Schedule of tasks in the study protocol.

(B, C) Mean perceptual task performance by time point: before training on day 1 (baseline), immediately following training on day 1 (day 1 post), and 12 h later (day 2 post). Performance is measured by d' scores (y-axis).

(D, E) Relationships between individual differences in sleep duration and the overnight change to perceptual task performance (day 2–day 1 d' scores on y-axis).

2.3.1. Nonnative phonetic training/assessment

On day 1, participants completed a baseline assessment of discrimination, training using an identification task, identification post-training assessment, and a post-training discrimination assessment. On day 2, participants completed a post-training identification assessment, followed by a post-training discrimination assessment. E-Prime 2.0 software [21] was used for stimulus presentation and recording participant response. Participants were presented the auditory stimuli through SONY MDR-7506 Hi-Fi digital Sound Monitor headphones, at an average listening level of 70 dB SPL.

During training, participants were first exposed to a familiarization sequence of each auditory token presented with the corresponding picture. They were then instructed to choose which of two pictures of novel objects corresponded to each of the two 'words.' Participants completed 200 trials of this task with a 2-min break halfway through, and were provided with written feedback

('correct' or 'incorrect') after every trial. During each post-training assessment, participants completed 50 trials of the same identification task without feedback. For the discrimination task, participants were presented two tokens in sequence (stimulus onset asynchrony [SOA] of 1 s), and instructed to indicate if the sound at the beginning of the two words belonged to the same or different speech sound category. During each assessment, participants completed 128 (64 'same' and 'different') discrimination trials without feedback.

Performance accuracy for each task was converted to d' scores [22]. In order to avoid infinite values, proportion accuracy scores of 0 or 1 were adjusted to a ceiling z -value of 4.65 [23].

2.3.2. ERP recording and analysis

E-Prime 2.0 software was used for stimulus presentation. Stimuli were presented over two wall-mounted free-field speakers approximately two feet in front and two feet above the participants

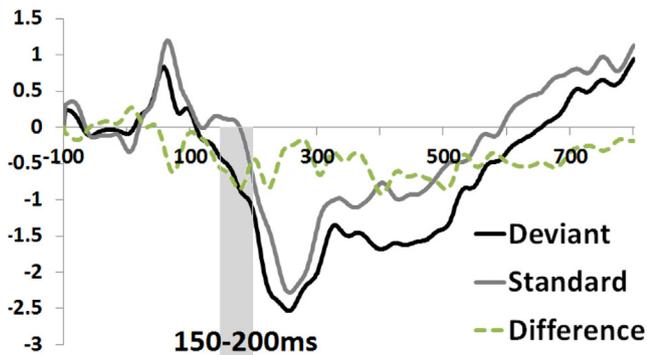
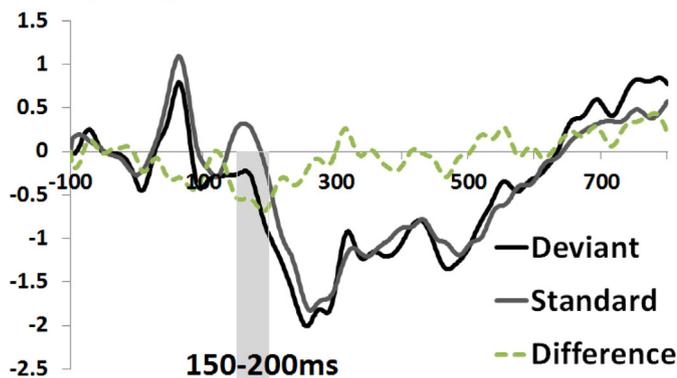
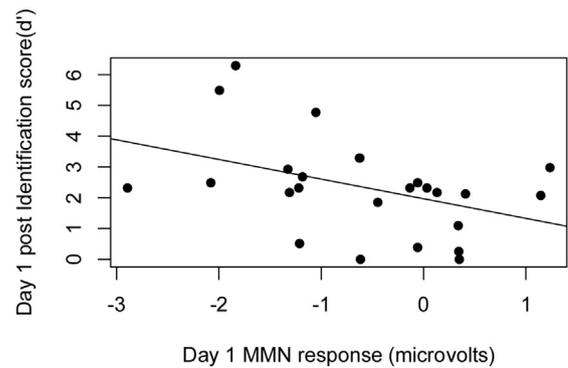
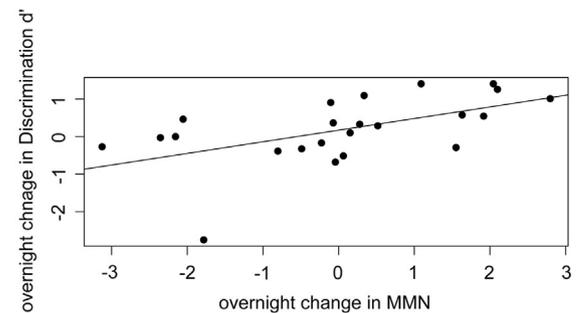
(A) Day 1 response waveforms**(C) Day 2 response waveforms****(B) Day 1 Identification score and MMN****(D) Overnight change in behavior/MMN**

Fig. 2. ERP responses to the target contrast and relationships with behavioral performance.

(A, C) ERP waveforms for each stimulus condition: Standard, Deviant, and Difference (Deviant minus Standard) recorded at the Fz electrode. Amplitude (y-axis) is in microvolts. (B, D) Relationships between MMN amplitude (area under the curve 150–200 ms of the difference waveform) and behavior.

at an average listening level of 75 dB SPL. Participants watched a muted nature documentary during the entire recording. In each recording session, participants were presented with a train of 667 stimuli (15%, or 100, deviant tokens) with an SOA of 650 ms in two blocks. Stimuli were arranged in a pseudorandom order such that deviant tokens were presented no fewer than 3, and no more than 8, standard tokens apart.

The EEG was recorded continuously using a 128-channel Hydrocel geodesic sensor net [24] with a Cz reference. Data were sampled at 500 Hz/channel with hardware filter settings 0.1–100 Hz. Using Netstation software [24], the recordings were bandpass filtered (0.3–30 Hz) and segmented (–100 to 800 ms from stimulus onset) to obtain ERPs. Segments containing artifacts (defined as min-max voltage $>200 \mu\text{V}$ over a moving average of 80 ms for eye blinks, and min-max voltage $>100 \mu\text{V}$ over a moving average of 80 ms for eye movements) were removed and bad channels (defined as channels with greater than 40% of segments marked bad) were spline interpolated before averaging. The data was re-referenced to the average reference and baseline corrected to 100 ms pre-stimulus presentation. Across participants, the number of segments retained for statistical analyses are as follows: for day 1, 72% deviant, 74% standard; for day 2, 71% deviant, 71% standard.

Only Standard tokens that immediately preceded Deviant tokens were included in statistical analyses. Area under the curve from 150 to 200 ms post-stimulus onset (where the MMN has been identified for auditory stimuli with only a subtle difference [25]) was calculated for the response waveforms at the electrode corre-

sponding to Fz in the 10/20 system for each condition separately, and subjected to further statistical analysis. Individual MMN values were computed by calculating the difference between the Deviant and Standard waveforms for the specified time window.

3. Results

3.1. Association between behavior and sleep

We tracked performance in each perceptual task (discrimination and identification) on days 1 and 2 (Fig. 1). An initial repeated measures analysis of variance (ANOVA) with time as the within-subjects factor, and subsequent paired samples *t*-tests, determined that participants improved in discrimination performance at each time point (Time main effect: $F(2,34) = 12.09$; $p < 0.001$, $\eta^2 = 0.42$; baseline to day 1 post, $t(17) = 2.74$, $p = 0.014$; day 1 to day 2 post, $t(17) = 3.02$, $p = 0.008$; all pairwise comparisons significant at 0.05 after Bonferroni correction). A paired samples *t*-test determined that participants improved in identification performance between days 1 and 2 ($t(17) = 4.51$, $p < 0.001$). The increase in performance from day 1 to day 2 in the absence of further training replicates previous findings showing that an intervening period of sleep facilitates improvement on perceptual tasks [5].

The average sleep duration across participants was 351 min (SD 45 min, range [285,410]). Different stages of sleep are associated with consolidation of different types of memory [26]; therefore, the two perceptual tasks likely differ in their respective relation-

ships to sleep as well, depending on the type of memory on which performance relies. Because duration of sleep stages are highly collinear, total sleep, rapid eye movement (REM) sleep, light sleep (NREM stages 1 and 2), and deep sleep (NREM stages 3 and 4) durations were entered stepwise as predictors into a multiple regression model using SPSS software [27] with performance gain (the difference between day 2 and day 1 posttest scores) as the dependent variable for each task. Total sleep duration accounted for the largest proportion of variance in overnight discrimination gain ($F(1,16)=6.94$, $p=0.018$, $r^2=0.302$), and light sleep duration accounted for the largest proportion of variance in identification gain ($F(1,16)=8.01$, $p<0.001$, $r^2=0.800$).

The significant MMN response on day 1 ($F(1,23)=4.96$, $p=0.007$, $\eta^2=0.27$) was found to be associated with identification performance on day 1 ($F(1,23)=4.84$, $p=0.038$, $r^2=0.17$). While the MMN response on day 2 was marginal ($F(1,23)=3.98$, $p=0.057$, $\eta^2=0.14$), we found that individual changes in the MMN between days were significantly correlated with the magnitude of behavioral performance gain on the discrimination task ($F(1,23)=6.10$, $p=0.021$, $r^2=0.21$), with a larger increase in MMN associated with improved discrimination performance from day 1 to day 2 (Fig. 2).

4. Discussion

The present findings lend strong support to the hypothesis that some differences in perceptual speech category learning are attributable to differences in sleep behavior. The observation that light sleep duration is associated with category identification performance is consistent with the argument that light sleep plays a dominant role in the consolidation of hippocampal memory [10].

The finding that discrimination performance is associated with total sleep duration allows for several possibilities for interpretation. For one, category discrimination may be facilitated by a sleep-mediated process (or a combination of processes) that are not confined to a particular stage of sleep. A second possibility is that improvement on perceptual discrimination may be facilitated by a period of a lack of conflicting speech input (i.e. silence) that is indirectly provided by periods of sleep. This interpretation would be consistent with our previous findings [6] that discrimination performance is vulnerable to degradation following exposure to conflicting phonetic information, while identification performance is not. Finally, all participants obtained less than 7 h of sleep between sessions, raising the possibility that sleep deprivation may have played a role in performance, particularly in those who showed the least amounts of performance gain. We note that performance in category discrimination, because the task was not explicitly trained, is consequently an indirect measure of learning. This may explain the discrepancy in the magnitude of training-induced gain between the two tasks.

Future investigations will address the need to disambiguate the subtle distinctions between different perceptual tasks as a measure of perceptual learning. Nonetheless, the relationship between learning and sleep duration suggests that, among typical individuals with comparable linguistic experience, sleep duration accounts for differences in the encoding of new speech features, necessary in the formation of new perceptual categories. Furthermore, electrophysiological results support the interpretation that the observed changes in perceptual performance reflect changes to neural structure and/or physiological changes.

5. Conclusions

This study demonstrated that relative success in the acquisition of new speech sounds is determined by post-exposure sleep duration. Over time, habitual differences in sleep habits may lead

incrementally to differences in the quality of the resultant linguistic units. This work prompts further investigation into the relationship between habitual sleep behaviors and language function.

Funding

This work was supported by NIH NIDCD grants R01 DC013064 to EBM, F31 DC014194 to FSE, and NIH NICHD grant P01 HD001994 (Rueckl, PI). FSE was supported by an ASH Foundation scholarship, and the Fund for Innovation in Science Education at the University of Connecticut. The content is the responsibility of the author and does not necessarily represent official views of our funding sources.

Acknowledgments

F.S.E. conceived the study, and together with E.B.M. designed the experiment. F.S.E. carried out the data collection and analysis with assistance from N.L. in analysis and interpretation of EEG data; F.S.E., E.B.M., and N.L. co-wrote the paper. We thank Nina Gumkowski for EEG technical support, Stephanie Del Tufo for assistance in participant classification, and Megan Speed and Jessica Joseph for rescoring standardized assessments.

References

- [1] K. Nation, M.J. Snowling, Individual differences in contextual facilitation: evidence from dyslexia and poor reading comprehension, *Child Dev.* 69 (4) (1998) 996–1011.
- [2] M.F. Joanisse, F.R. Manis, P. Keating, M.S. Seidenberg, Language deficits in dyslexic children: speech perception, phonology, and morphology, *J. Exp. Child Psychol.* 77 (1) (2000) 30–60.
- [3] D. Hartas, Inequality and the home learning environment: predictions about seven-year-olds' language and literacy, *Br. Educ. Res. J.* 38 (5) (2012) 859–879.
- [4] B. Rasch, J. Born, About sleep's role in memory, *Physiol. Rev.* 93 (2) (2013) 681–766.
- [5] K.M. Fenn, H.C. Nusbaum, D. Margoliash, Consolidation during sleep of perceptual learning of spoken language, *Nature* 425 (6958) (2003) 614–616.
- [6] F.S. Earle, E.B. Myers, Sleep and native language interference affect non-native speech sound learning, *J. Exp. Psychol.: Hum. Percept. Perform.* 41 (6) (2015) 1680–1695.
- [7] F.S. Earle, E.B. Myers, Overnight consolidation promotes generalization across talkers in the identification of nonnative speech sounds, *J. Acoust. Soc. Am.* 137 (1) (2015) EL91–EL97.
- [8] O. Lahl, C. Wispel, B. Willigens, R. Pietrowsky, An ultra short episode of sleep is sufficient to promote declarative memory performance, *J. Sleep Res.* 17 (1) (2008) 3–10.
- [9] S. Diekelmann, J. Born, The memory function of sleep, *Nat. Rev. Neurosci.* 11 (2) (2010) 114–126.
- [10] L. Genzel, M.C. Kroes, M. Dresler, F.P. Battaglia, Light sleep versus slow wave sleep in memory consolidation: a question of global versus local processes? *Trends Neurosci.* 37 (1) (2014) 10–19.
- [11] Zeo, Inc., Boston, MA.
- [12] J.R. Shambroom, S.E. Fabregas, J. Johnstone, Validation of an automated wireless system to monitor sleep in healthy adults, *J. Sleep Res.* 21 (2) (2012) 221–230.
- [13] R. Näätänen, P. Paavilainen, T. Rinne, K. Alho, The mismatch negativity (MMN) in basic research of central auditory processing: a review, *Clin. Neurophysiol.* 118 (12) (2007) 2544–2590.
- [14] R.W. Woodcock, Woodcock Reading Mastery Tests – Revised/Normative Update (WRMT-R/NU), American Guidance Service, Circle Pines, MN, 1998.
- [15] J.K. Torgesen, R. Wagner, C. Rashotte, Test of Word Reading Efficiency (TOWRE), Pro-Ed., Austin, TX, 1999.
- [16] D. Wechsler, Wechsler Abbreviated Scale of Intelligence, 3rd ed., The Psychological Corporation, San Antonio, TX, 1999.
- [17] L.J. Fidler, E. Plante, R. Vance, Identification of adults with developmental language impairments, *Am. J. Speech Lang. Pathol.* 20 (1) (2011) 2–13.
- [18] Roland corporation, Los Angeles, CA.
- [19] P. Boersma, D. Weenink, 2011. Praat: Doing phonetics by computer [Computer program]. Version 5.3. 03.
- [20] "fribbles"; Stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University, <http://www.tarrlab.org/>.
- [21] Psychology Software Tools, Pittsburgh, PA.
- [22] N.A. Macmillan, C.D. Creelman, Detection Theory: A User's Guide, Cambridge University Press, Cambridge, England, 1991.
- [23] P. Iverson, P.K. Kuhl, Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/, *J. Acoust. Soc. Am.* 99 (2) (1996) 1130–1140.

- [24] Electrical Geodesics, Inc., Eugene, OR, USA.
- [25] M. Sallinen, J. Kaartinen, H. Lyytinen, Processing of auditory stimuli during tonic and phasic periods of REM sleep as revealed by event-related brain potentials, *J. Sleep Res.* 5 (4) (1996) 220–228.
- [26] S.M. Fogel, C.T. Smith, K.A. Cote, Dissociable learning-dependent changes in REM and non-REM sleep in declarative and procedural memory systems, *Behav. Brain Res.* 180 (1) (2007) 48–61.
- [27] IBM Corp. Released, IBM SPSS Statistics for Windows, Version 22.0, IBM Corp., Armonk, NY, 2013.