Multi-session training in the evening schedule exhibits enhanced speech learning

Article *in* The Journal of the Acoustical Society of America · February 2024

DOI: 10.1121/10.0024934

5 authors, including:

Akshay Maggu
University of Connecticut

Bhamini Sharma
The Hong Kong Polytechnic University

JASA ARTICLE

Multi-session training in the evening schedule exhibits enhanced speech learning

Akshay R. Maggu,[1,2,3,a)] Bhamini Sharma,[4] Mansi S. Roy,[3] Tanaya Rowell,[5] and Lydia Seiling[6]
[1]Department of Speech, Language, and Hearing Sciences, University of Connecticut, Storrs, Connecticut 06269, USA
[2]Institute for Brain and Cognitive Sciences, University of Connecticut, Storrs, Connecticut 06269, USA
[3]Department of Speech-Language-Hearing Sciences, Hofstra University, Hempstead, New York 11549, USA
[4]Department of Communication Arts, Sciences, and Disorders, City University of New York – Brooklyn College, Brooklyn, New York 11210, USA
[5]Department of Biology, Hofstra University, Hempstead, New York 11549, USA
[6]Department of Health and Human Services, Hofstra University, Hempstead, New York 11549, USA

ABSTRACT:
Speech learning can be influenced by a variety of factors. A growing body of literature suggests a significant influence of sleep on speech learning, i.e., those trained in the evening outperform those trained in the morning most probably due to consolidation of learning that happens during the sleep for the evening group. Since, learning, in general, may be a process that spans multiple sessions, in the current exploratory study, we aimed at investigating the effect of a multi-session training paradigm on the learning performance of the morning vs evening group. We compared young adults who were trained in the morning (8–10 am; $n = 16$) with those who were trained in the evening (6–8 pm; $n = 16$) on a Hindi dental-retroflex pseudoword-picture association training paradigm. Overall, we found that the evening group learned to a larger extent both for the identification (on trained items) and discrimination (on untrained items) tasks. The current findings, even with a multi-session paradigm, are consistent with the previous findings that support enhanced performance by training in the evening. These findings may have clinical implications toward scheduling of speech therapy.
https://doi.org/10.1121/10.0024934

(Received 27 August 2023; revised 28 January 2024; accepted 29 January 2024; published online 13 February 2024)
[Editor: Melissa Michaud Baese-Berk] Pages: 1336–1342

I. INTRODUCTION

Speech and language learning is a complex process that can be affected by a variety of subject-internal (e.g., musical experience, cognitive abilities, general perceptual aptitude) and subject-external (e.g., amount of stimulation, stimuli variability, multiple sessions) factors (Allen, 2013; Fuhrmeister and Myers, 2020; Qin *et al.*, 2021; Wade *et al.*, 2007; Wong and Perrachione, 2007). In the past decade, a growing body of scientific literature (Batterink *et al.*, 2014; Drouin *et al.*, 2023; Earle and Arthur, 2017; Earle and Myers, 2015a,b; Fenn *et al.*, 2003; Fuhrmeister *et al.*, 2020; Fuhrmeister and Myers, 2020; Holz *et al.*, 2012; Qin and Zhang, 2019) has uncovered the role of sleep as a significant subject-internal factor that can affect speech and language learning. More specifically, from studies based on training non-native speech sounds, it has been found that those who are trained in the evening learn better as compared to those trained in the morning (Drouin *et al.*, 2023; Earle and Myers, 2015a,b; Fuhrmeister *et al.*, 2020; Holz *et al.*, 2012; Qin and Zhang, 2019) due to consolidation that happens during sleep. As most of the previous relevant studies comparing the morning vs evening training groups are based on training within a single session, the effect of a combination of an extended training dosage with training sessions spanning across multiple days is currently unknown. In the current exploratory study, using a 5-session artificial language training paradigm (Antoniou and Wong, 2016; Chen *et al.*, 2015; Maggu *et al.*, 2018; Maggu *et al.*, 2019; Wong and Perrachione, 2007), we examined the effects of multi-session training on the learning (of trained items) and generalization (to untrained items) for the morning vs evening groups.

Recent studies on non-native speech sound learning have revealed a significant effect of sleep on learning. More specifically, from a theoretical perspective of the two-stage complementary systems (CLS) model, learning causes initial episodic encoding of information that is being taught within a particular context, but overnight sleep facilitates a consolidation process that helps in the abstraction of information beyond the context, via a declarative memory system, leading to generalization to novel stimuli. This has been found to be true for both tone (Qin and Zhang, 2019) and non-tone language stimuli (Drouin *et al.*, 2023; Earle and Myers, 2015a). One of the most popular paradigms that has yielded replicable findings in this regard consists of two

[a)]Present address: 202 Phillips Communication Sciences Building, 2 Alethia Drive, Storrs, CT 06269, USA. Email: akshay.maggu@uconn.edu

https://doi.org/10.1121/10.0024934

subject sample groups i.e., those trained in the morning vs those trained in the evening (Earle and Myers, 2015a,b). These groups are evaluated using a discrimination task pre- and post-training with untrained stimuli and using an identification task for the trained items right after the training task. A second post-training evaluation is conducted at a 12-h interval from the first post-training evaluation which is on the same day evening for the morning-training-group but on the next day morning for the evening-training-group. In other words, even though there is an equal duration gap (i.e., 12 h) between the two post-training evaluations within the groups, the evening group gets nighttime sleep between the two post-training evaluations while the morning group does not. A key finding from these studies has been that those trained in the evening benefit from overnight consolidation while those in the morning do not (Earle and Myers, 2015a,b; Qin and Zhang, 2019). Another reason why the morning group lags in performance is potentially due to interference caused by exposure to the native language between their two post-training evaluations. This has been experimentally demonstrated by Earle and Myers (Earle and Myers, 2015b) who when exposed the speakers of American English, who were just trained for Hindi contrasts, to English alveolar plosive sounds (e.g., /d/), found reduced performance on the post-training evaluation for Hindi contrasts.

We know that training in the evening before the time of sleep leads to enhanced learning as compared to training in the morning. However, most of the studies in the literature (Drouin *et al.*, 2023; Earle and Myers, 2015a,b; Holz *et al.*, 2012; Qin and Zhang, 2019) have examined these effects via a one-session training paradigm where either those trained in the morning were evaluated right after training, on the same day, or the next day morning. It has been found that training in the morning does not yield as much learning outcome as compared to training in the evening either due to lack of consolidation at night and/or probably due to increased interference by the native language during the daytime (Earle and Myers, 2015b). However, learning, in general, may not be a "one-session process," especially in the clinical fields of psychology and speech language pathology, and thus, a repetitive multi-session therapy program is usually normative for clinics. Research in the area of auditory perceptual learning has shown that multiple sessions of shorter duration are more beneficial than a single session of longer duration (Molloy *et al.*, 2012). In a study of veterans with hearing loss who were provided with auditory training, it was found that multiple regular sessions led to enhanced improvement in their auditory perception (Chisolm *et al.*, 2013). Thus, given that learning typically involves multiple sessions, it becomes imperative to understand how multiple sessions of training interact with factors such as sleep-mediated consolidation in the morning vs evening groups.

In the current study, we compared young adults who were trained in the morning (8–10 am) vs those who were trained in the evening (6–8 pm) on a 5-session Hindi voiceless dental and retroflex pseudoword-picture association paradigm. More specifically, in this exploratory study, we evaluated the effect of a multi-session training paradigm (with one session per day) on the learning performance of morning vs evening groups. If the combination of sleep-mediated consolidation (a subject-internal factor) and multiple sessions (a subject-external factor) turned out to be additive, we predicted to see enhanced learning and generalization in the evening group as compared to the morning group, a finding similar to the existing findings on one-session paradigm (Earle and Myers, 2015b,a). On the other hand, if the combination of sleep-mediated consolidation and multiple sessions did not turn out to be additive, we did not expect to see an enhanced performance on the learning and generalization tasks in the evening group.

II. METHOD

A. Participants

In the current study, 32 young adults (mean age: 18.5, standard deviation, SD: 0.92; 6 males, 25 females, 1 non-binary) with normal hearing sensitivity were recruited via posting flyers and/or via the SONA system of Hofstra University. Another prerequisite for them was to be native speakers of North American English and have no knowledge of any South-Asian languages (e.g., Hindi, Tamil, Kannada, Bengali, etc.). They were either paid $10/h or were granted SONA course credits. This study was approved by the Institutional Review Board of Hofstra University. The participants were randomly allocated to the morning group ($n = 16$; mean age: 18.38, SD: 0.72; 3 males, 13 females) or the evening group ($n = 16$, mean age: 18.62, SD: 0.56; 3 males, 12 females, 1 non-binary). There were eight other participants who were originally a part of this study but had to be excluded as three of them could not pass the hearing testing and five of them could not attend the five consecutive training sessions and/or discontinued without finishing the study.

A priori ***power analysis:*** We decided on the number of participants for the current study based on the *a priori* power analysis conducted in our previous study (Maggu *et al.*, 2019) which essentially employed the same methods and stimuli as in the current study. In our previous study (Maggu *et al.*, 2019), it was estimated that in order to achieve 80% power with an alpha of 0.05 with the minimum effect size (0.57) of our dependent variable of interest, 14 subjects per training group were sufficient.

B. Hearing testing

All participants were evaluated for pure tone thresholds between 250 and 8000 Hz along with testing inter-octave frequencies using a GSI-18 screening audiometer (Grasen Stadler, Eden Prairie, MN) in an acoustically padded room. Pure tones were routed to each ear separately via TDH-39 supra-aural headphones. All participants had pure tone thresholds of ≤ 25 dB hearing level (HL) for all octave and inter-octave frequencies between 250 and 8000 Hz. Figure 1 depicts the audiograms for the participants for their right [Fig. 1(A) and left [Fig. 1(B)] ears.

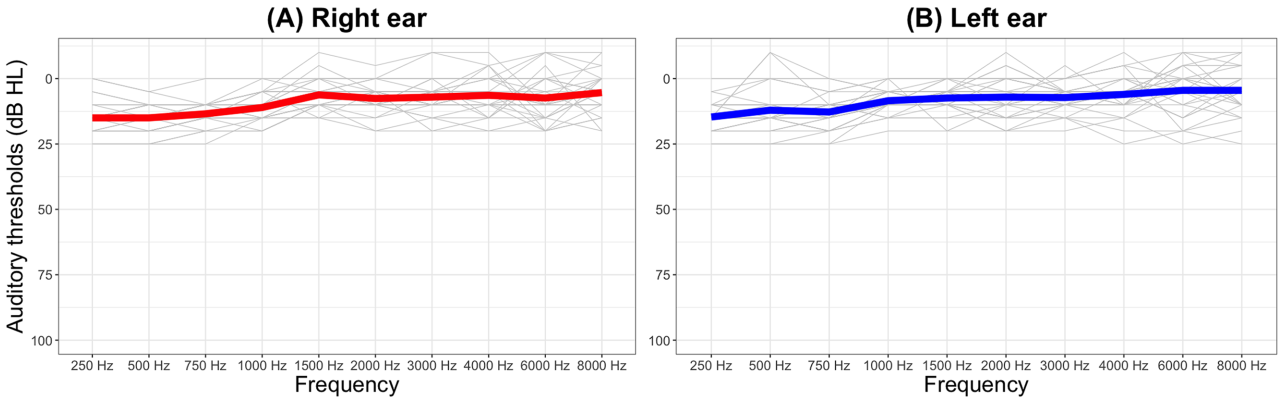

FIG. 1. (Color online) Pure tone thresholds for frequencies from 250 to 8000 Hz, including inter-octave frequencies for the (A) right ear; and (B) left ear. Solid red and blue lines are average pure tone thresholds for the right and left ears, respectively.

C. Musical experience

As it has been found that musical experience is a significant facilitator for speech learning (Maggu *et al.*, 2018; Wong and Perrachione, 2007), we wanted to ascertain that our study did not differ in their musical proficiency. Subjects were asked to report their musical proficiency on a Likert scale ranging from 0 to 7 with "0" being "not at all proficient" and "7" being "very proficient." Morning (mean = 2.31, SD = 2.3) and evening groups (mean = 2.88, SD = 2.25) did not significantly differ in their musical proficiency [$t(30) = -0.69$, $p = 0.489$].

D. Cognitive testing

It has been found that cognitive abilities are associated with speech perception abilities (Ingvalson *et al.*, 2015; Ingvalson *et al.*, 2017; Rudner *et al.*, 2009; Yoo and Bidelman, 2019). For example, those with enhanced working memory may perform better on speech perception as compared to those with poor working memory. We wanted to ascertain that our study groups did not differ in their cognitive abilities. Cognitive testing involved testing for executive function and working memory. It was conducted via the NIH cognition toolbox battery (Weintraub *et al.*, 2013; Weintraub *et al.*, 2014) via an iPad (9th generation; 15.4.1). For testing the executive function, the Flanker subtest of the battery was used and for testing the working memory, the list sorting working memory subtest of the battery was used. It was found that there was no significant difference between the two groups on executive function [morning: mean = 94.06, SD = 12.64; evening = 99.06, SD = 20.26; $t(30) = -0.83$, $p = 0.409$] and working memory [morning: mean = 112.81, SD = 18.22; evening = 108.06, SD = 29.16; $t(30) = 0.55$, $p = 0.584$].

E. Speech testing and training

1. Stimuli

Speech stimuli that were used for both training and testing consisted of Hindi voiceless dental-retroflex sounds from our previous study (Maggu *et al.*, 2019). These sounds included consonants /t̪, ʈ/ produced with eight vowels (/a/, /i/, /u/, /o/, /e/, /ae/, /ə/, /ɒ/) by a phonetically trained native Hindi speaker. Recordings were made using a Shure SM10A microphone and Praat (Boersma and Weenink, 2010) with a 44 100 Hz sampling rate and 16-Hz sampling depth. Picture stimuli that were used in the training task of the current study consisted of black and white pictures of familiar objects, used in the previous studies (Antoniou and Wong, 2016; Chen *et al.*, 2015; Maggu *et al.*, 2018; Maggu *et al.*, 2019; Wong and Perrachione, 2007). The pictures of familiar objects were chosen for this study in order to minimize the need for learning novel objects and novel "words" at the same time. The speech stimuli were resynthesized using Praat to match duration, intensity, and pitch to ascertain that the dental and retroflex tokens had all acoustic parameters the same and differed mainly on the consonant-vowel (CV) transition of the third formant—a feature that is contrastive between dental and retroflex plosives. The stimuli pairs were assigned to either Set-A or Set-B. Set-A consisted of sounds /t̪, ʈ/ in /Ca/, /Ci/, /Cu/, /Co/ context, and Set B had these sounds in /Ce/, /Cae/, /Cə/, /Cɒ/ context. Since the Set-A and Set-B sounds exhibited equivalent performance in our previous study (Maggu *et al.*, 2019), the participants in the current study were tested on Set-A and trained on Set-B. All stimuli in all the tasks of the current study were routed to the participants via Sennheiser 280 headphones (Sennheiser, Wedemark, Germany).

2. Pre- and post-training evaluation

At the start and end of the training paradigm, a self-paced AX (same/different) discrimination testing was conducted to assess the participants' ability to perceive the contrast. Sounds from Set-A were used for these evaluations. Dental-retroflex pairs were presented, and the subjects had to press the appropriate button on the keyboard to indicate "same" or "different." The subjects were provided relevant instructions on the computer screen. After they pressed "space" on the keyboard, a "+" appeared for 500 ms, followed by the stimuli pairs with 1000 ms interstimulus interval between the sounds in the pairs. If they perceived any difference between the sound pairs, they were expected to

JASA https://doi.org/10.1121/10.0024934

press the button corresponding to "different." On the other hand, if they perceived no difference between the sound pairs, they were expected to press the button corresponding to "same." The rundown of the evaluation was programmed using E-Prime 3.0. Subjects were presented with a total of 64 pairs (4 token pairs $\times$ 4 repetitions $\times$ 2 "same" sequences $\times$ 2 "different" sequences). Before the actual evaluation, subjects were presented with practice items that were not part of the evaluation, to familiarize the subjects regarding the task. Feedback was provided for the practice items to enable learning of the task. The evaluation took about 5 min to complete.

3. Training

Training consisted of a pseudoword-picture association task as in the previous studies (Antoniou and Wong, 2016; Chen *et al.*, 2015; Maggu *et al.*, 2018; Maggu *et al.*, 2019; Wong and Perrachione, 2007). Training was conducted on five consecutive days between 8 and 10 am for the morning group and 6 and 8 pm for the evening group. In each training session, 64 pseudoword-picture pairs [2 contrast-type (dental/retroflex) $\times$ 4 tokens $\times$ 8 repetitions] were presented with 2300 ms stimulus onset asynchrony between the items. The participants were instructed to pay attention to the presented pseudoword-picture pairs to learn these associations. For example, every time the participants heard the sound /t̪a/, it was accompanied by the picture of "shirt" and so on [Fig. 2(A)]. The instructions were as follows: "You will see some pictures and hear the word associated with those pictures in a new language. You do not need to press anything to go to the next picture—it will advance automatically. Press any key to continue with the instructions." These were followed with another slide of instructions: "After you have finished learning these word-picture pairings, you will be given a test to see how well you learned the new language. Press any key to begin training." After the training, these instructions were provided: "You have finished the training phase. Now you will be tested to see how well you have learned these new words. Press any key to continue with instructions." After they pressed any key to continue, the subjects were evaluated for what they learned in that training session via an 8-alternate forced choice identification (AFC) task. The instructions were as follows: "Next you will see 8 pictures on the screen. You will also hear a word in the new language you just learned. Please select the correct picture by using the keyboard to enter the appropriate number (1–8). You will not be timed. Press any key to begin." They were expected to select the appropriate picture for the presented pseudoword based on their learning in the pseudoword-picture association task. For example, for a sound /t̪ae/, the participants were expected to press the key "2" on the keyboard that corresponded to the picture of a soccer ball—an association that was provided to them in the training [Fig. 2(B)]. The subjects did not receive any feedback on their responses to this task. The order of stimuli was randomized across sessions and subjects in both training and identification tasks. Each session lasted for 13–15 min in duration.

Pre-training evaluation was conducted on the first day followed by which the first session of training took place. Post-training evaluation was conducted on the fifth day after the fifth session of training was conducted.

III. RESULTS

A. Training

Identification data obtained from the 8-AFC task, which was conducted following each training session, were used to

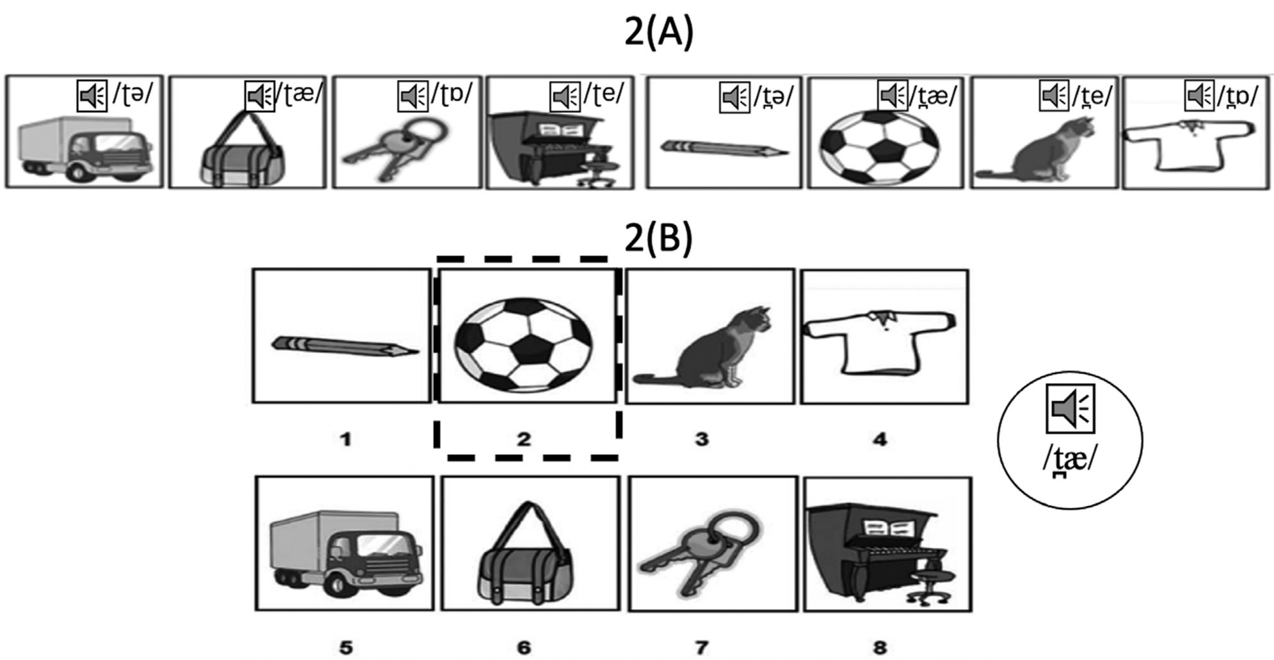

FIG. 2. (A) Association of pictures with the pseudoword in the training paradigm; (B) 8-AFC identification task where the subjects were expected to match the relevant picture to the heard sound based on what they learned during the training. In the previous example, a subject would choose picture #2 for the pseudoword /t̪ae/.

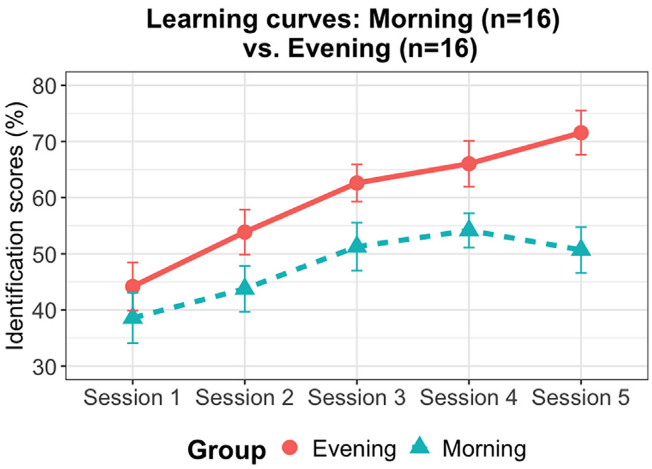

FIG. 3. (Color online) Comparison of the learning curves for morning and evening groups across the five training sessions. Error bars indicate ± 1 (Standard Error of the Mean) SEM.

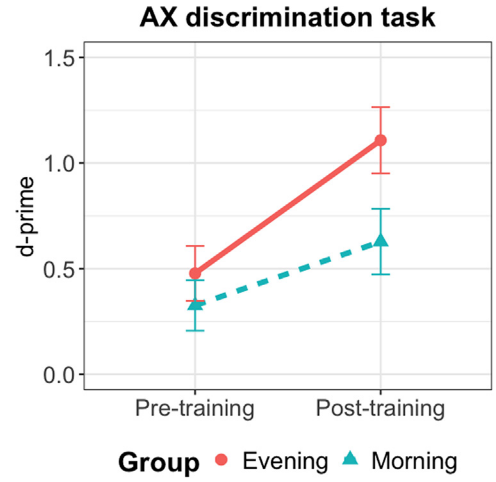

FIG. 4. (Color online) Comparison of the morning and evening groups on discrimination of the dental-retroflex contrast before (i.e., pre) and after (i.e., post) training. Error bars indicate ±1 SEM.

plot learning curves. Figure 3 contains average learning curves for the morning and evening sessions. For the morning group, identification scores ranged between 40% and 50% and for the evening group, identification scores ranged between 40% and 70% across the five sessions of the training. Then, 2 (Morning and Evening groups) × 2(first and last sessions) mixed analysis of variance (ANOVA) was conducted on the training data, using R studio version 3.6.3 (RStudio Team, 2019). Statistical analysis revealed main effects of Group [$F(1,30) = 7.26$, $p = 0.011$] and Session [$F(1,30) = 47.43$, $p < 0.001$] and a significant Group × Session interaction [$F(1, 30) = 7.44$, $p = 0.011$]. A simple effect analysis revealed that the two groups did not differ in the first session [$F(1,30) = 0.73$, $p = 0.39$] but significantly differed on the fifth session [$F(1,30) = 16.02$, $p < 0.001$].

B. Pre- vs post-training evaluation

An AX discrimination task using Set-A of the stimuli was used for examining the extent of generalization due to training with Set-B stimuli. D-prime was calculated using a dprime() function-based pipeline (Tichko, 2021) on R studio (RStudio Team, 2019). Overall, both morning and evening groups improved after training on the AX discrimination task. However, the evening group exhibited more improvement as compared to the morning group (Fig. 4). A 2(Group) × 2(Evaluation) mixed ANOVA revealed a main effect of Evaluation [$F(1,30) = 15.35$, $p < 0.001$] and a significant Group × Evaluation interaction [$F(1,30) = 4.23$, $p = 0.047$]. Further, a simple effect analysis revealed that the two groups differed only on the post-training evaluation [$F(1,30) = 5.42$, $p = 0.024$] but not on the pre-training evaluation [$F(1,30) = 0.41$, $p = 0.527$].

IV. DISCUSSION

The current exploratory study was aimed at examining the effect of multiple sessions of training on the learning performance of morning vs evening training groups. More especially, we investigated this research question by comparing a morning-training-group (8–10 am; $n = 16$) with an evening-training-group (6–8 pm; $n = 16$) on a pseudoword-picture association training containing non-native speech contrast that spanned across five consecutive days. Overall, we found that both the morning and the evening groups continued to learn across the five consecutive days of training, but the morning group lagged behind the evening group in their learning performance. In addition to the difference in learning performance between the groups, we found that the morning group also performed worse than the evening group on the generalization to novel dental-retroflex contrasts (i.e., untrained stimuli).

The current findings are consistent with the existing research that has found that those trained in the evening exhibit more improvement as compared to those who are trained in the morning (Drouin *et al.*, 2023; Earle and Myers, 2015a,b; Holz *et al.*, 2012; Qin and Zhang, 2019). One of the key differences between the results of the current and the previous study (Earle and Myers, 2015b) was that in the current study, learning during the evening resulted in enhanced discrimination of dental-retroflex contrasts with novel vowel contexts. Previous studies (Earle and Myers, 2015a,b; Fuhrmeister *et al.*, 2020; Fuhrmeister and Myers, 2020) have found that the tasks requiring a declarative recall (e.g., picture-pseudoword identification) are benefitted by the sleep-mediated consolidation. This benefit was found to extend to discrimination tasks of trained items but not to untrained items (Earle and Myers, 2015a,b) unlike in the current study. However, a critical difference between the existing studies and the current study is that the existing studies used a one-session training paradigm with the follow-up testing to be done 12 h after the training, i.e., on the same night for the morning group and the next morning for the evening group. In comparison, in the current study, the training and identification testing was conducted every 24 h for five consecutive days. It can be speculated that the repeated training sessions might have interacted with the sleep-mediated consolidation mechanisms to extend the benefits toward discrimination of the dental-retroflex contrast with novel vowel contexts. Generalization learning to novel dental-retroflex contrasts as a result of this multisession training is consistent with our previous study (Maggu *et al.*, 2019). Further, there are several other methodological differences between the current research and the previous studies that could have contributed to the difference between the findings of the current and the aforementioned studies. First, the stimuli in the current study were Hindi voiceless dental and retroflex sounds (e.g., /t̪, ʈ/) while the previous studies contained Hindi pre-voiced dental and retroflex sounds (e.g., /d̪, ɖ/). As it is known that the pre-voiced sounds are typologically more marked/complex as compared to voiceless sounds (Itô and Mester, 1998; Kager *et al.*, 2007; Lombardi, 1999), this could have been a possible contributor towards the lack of generalization to novel contrasts in the previous studies. Second, the current study trained in multiple phonological environments by using four different vowels with the dental and retroflex sounds. In contrast, most of the previous studies are based on training using one (Earle and Myers, 2015a,b; Fuhrmeister *et al.*, 2020) or two vowels (Fuhrmeister and Myers, 2020) with the dental and retroflex sounds. Further, in the current study, as in the previous studies from our lab (Maggu *et al.*, 2018; Maggu *et al.*, 2019), we have used pictures of known objects, unlike the previous studies that have used pictures of novel objects. Further, no feedback was provided during the training in the current study while feedback has been used in the previous studies.

The findings of the current study are of relevance both from theoretical and clinical standpoints. By visual examination of the learning data, comparable performance can be observed for the morning group in their final training session and the evening group on their first training session. In other words, repeating the training sessions did not improve the performance of the morning group. This could have been due to the interaction between multiple sessions and sleep-mediated memory consolidation that possibly strengthened the encoding/learning for those trained in the evening (see discussions in Fuhrmeister *et al.*, 2020; Shibata *et al.*, 2017). From the standpoint of clinics where the therapy is conducted on a regular basis within a set schedule, these findings are relevant for optimizing the therapy schedule for the benefit of both patients and clinicians alike.

While this study compared the effect of repeating the sessions in the morning vs the evening over five consecutive days, on non-native speech contrast discrimination, the dosage per session was kept equal across the training groups. In the future, it will be interesting to examine the effects of increasing the dosage per session in the morning group as compared to the evening group. For example, by doubling the training items in the morning sessions, it would be interesting to see whether the morning group can catch up to the performance of the evening group. Findings from such research will be beneficial in optimizing the therapy by understanding the amount of effort/time that will be needed depending on the therapy schedule.

A. Limitations

Previous studies (Earle *et al.*, 2017; Qin and Zhang, 2019) have administered questionnaires to collect information regarding participants' sleep (e.g., sleep duration, wake-up and fall-asleep times, quality of sleep, etc.) for the time that they were enrolled in the experiment. In the current study, such measurements were not collected across the five days of the training. As self-report sleep-related data have been suggested to be less precise and further suggestions have been made to objectively record sleep duration and quality for more precise measurements, future studies could employ sleep-monitoring headband-like devices to record sleep-related aspects in a more accurate manner (Qin and Zhang, 2019). Also, future studies may consider employing mixed effect models to account for by-participant and by-item variabilities.

ACKNOWLEDGMENTS

The current work was supported by the Faculty Research Development Grant program and the Rabinowitz Honors College program at Hofstra University to AM.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts of interest to disclose.

Ethics Approval

This study was approved by the Institutional Review Board of Hofstra University.

DATA AVAILABILITY

Data will be made available on request.

Allen, M. M. (**2013**). "Intervention efficacy and intensity for children with speech sound disorder," J. Speech. Lang. Hear. Res. **56**, 865–877.
Antoniou, M., and Wong, P. C. M. (**2016**). "Varying irrelevant phonetic features hinders learning of the feature being trained," J. Acoust. Soc. Am. **139**, 271–278.
Batterink, L. J., Oudiette, D., Reber, P. J., and Paller, K. A. (**2014**). "Sleep facilitates learning a new linguistic rule," Neuropsychologia **65**, 169–179.
Boersma, P., and Weenink, D. (**2010**). "Praat: Doing phonetics by computer (version 5.1.44) [computer program]," http://scholar.google.com/scholar?cluster=834095044295694369&hl=en&oi=scholarr (Last viewed September 21, 2015).
Chen, Z., Wong, F. C., Jones, J. A., Li, W., Liu, P., Chen, X., and Liu, H. (**2015**). "Transfer effect of speech-sound learning on auditory-motor processing of perceived vocal pitch errors," Sci. Rep. **5**, 13134.
Chisolm, T. H., Saunders, G. H., Frederick, M. T., McArdle, R. A., Smith, S. L., and Wilson, R. H. (**2013**). "Learning to listen again: The role of compliance in auditory training for adults with hearing loss," Am. J. Audiol. **22**, 339–342.
Drouin, J. R., Zysk, V. A., Myers, E. B., and Theodore, R. M. (**2023**). "Sleep-based memory consolidation stabilizes perceptual learning of noise-vocoded speech," J. Speech Lang. Hear. Res. **66**, 720–734.

Earle, F. S., and Arthur, D. T. (**2017**). "Native phonological processing abilities predict post-consolidation nonnative contrast learning in adults," J. Acoust. Soc. Am. **142**, EL525–EL531.

Earle, F. S., Landi, N., and Myers, E. B. (**2017**). "Sleep duration predicts behavioral and neural differences in adult speech sound learning," Neurosci. Lett. **636**, 77–82.

Earle, F. S., and Myers, E. B. (**2015a**). "Overnight consolidation promotes generalization across talkers in the identification of nonnative speech sounds," J. Acoust. Soc. Am. **137**, EL91–EL97.

Earle, F. S., and Myers, E. B. (**2015b**). "Sleep and native language interference affect non-native speech sound learning," J. Exp. Psychol. Hum. Percept. Perform. **41**, 1680–1695.

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

Fuhrmeister, P., and Myers, E. B. (**2020**). "Desirable and undesirable difficulties: Influences of variability, training schedule, and aptitude on nonnative phonetic learning," Atten. Percept. Psychophys. **82**, 2049–2065.

Fuhrmeister, P., Smith, G., and Myers, E. B. (**2020**). "Overlearning of nonnative speech sounds does not result in superior consolidation after a period of sleep," J. Acoust. Soc. Am. **147**, EL289–EL294.

Holz, J., Piosczyk, H., Landmann, N., Feige, B., Spiegelhalder, K., Riemann, D., Nissen, C., and Voderholzer, U. (**2012**). "The timing of learning before night-time sleep differentially affects declarative and procedural long-term memory consolidation in adolescents," PLoS One **7**, e40963.

Ingvalson, E. M., Dhar, S., Wong, P. C. M., and Liu, H. (**2015**). "Working memory training to improve speech perception in noise across languages," J. Acoust. Soc. Am. **137**, 3477–3486.

Ingvalson, E. M., Lansford, K. L., Fedorova, V., and Fernandez, G. (**2017**). "Cognitive factors as predictors of accented speech perception for younger and older adults," J. Acoust. Soc. Am. **141**, 4652–4659.

Itô, J., and Mester, A. (**1998**). "Markedness and word structure: OCP effects in Japanese," M.S. thesis, University of California Santa Cruz, Santa Cruz, CA.

Kager, R., Van der Feest, S., Fikkert, P., Kerkhoff, A., and Zamuner, T. (**2007**). "Representations of [voice]: Evidence from acquisition," in *Voicing Dutch*, edited by J. van de Weijer and E. J. van der Torre (John Benjamins, Amsterdam), pp. 41–80.

Lombardi, L. (**1999**). "Positional faithfulness and voicing assimilation in Optimality Theory," Nat. Lang. Linguist. Theory **17**, 267–302.

Maggu, A. R., Kager, R., Xu, S., and Wong, P. C. M. (**2019**). "Complexity drives speech sound development: Evidence from artificial language training," J. Exp. Psychol. Hum. Percept. Perform. **45**, 628–644.

Maggu, A. R., Wong, P. C., Liu, H., and Wong, F. C. (**2018**). "Experience-dependent Influence of Music and Language on Lexical Pitch Learning Is Not Additive," in *Proceedings of Interspeech 2018*, September 2–6, Hyderabad, India.

Molloy, K., Moore, D. R., Sohoglu, E., and Amitay, S. (**2012**). "Less is more: Latent learning Is maximized by shorter training sessions in auditory perceptual learning," PLoS One **7**, e36929.

Qin, Z., Gong, M., and Zhang, C. (**2021**). "Neural responses in novice learners' perceptual learning and generalization of lexical tones: The effect of training variability," Brain Lang. **223**, 105029.

Qin, Z., and Zhang, C. (**2019**). "The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts," PLoS One **14**, e0221498.

RStudio Team (**2019**). "RStudio: Integrated Development for R," http://www.rstudio.com/ (Last viewed November 14, 2023).

Rudner, M., Foo, C., Rönnberg, J., and Lunner, T. (**2009**). "Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids," Scand. J. Psychol. **50**, 405–418.

Shibata, K., Sasaki, Y., Bang, J. W., Walsh, E. G., Machizawa, M. G., Tamaki, M., Chang, L.-H., and Watanabe, T. (**2017**). "Overlearning hyperstabilizes a skill by rapidly making neurochemical processing inhibitory-dominant," Nat. Neurosci. **20**, 470–475.

Tichko, P. (**2021**). "*pipeline to calculate D-prime in R*," https://ptichko.github.io/2021/02/27/Pipeline-To-Calculate-D-Prime-in-R.html (Last viewed November 9, 2023).

Wade, T., Jongman, A., and Sereno, J. (**2007**). "Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds," Phonetica **64**, 122–144.

Weintraub, S., Bauer, P. J., Zelazo, P. D., Wallner-Allen, K., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Slotkin, J., Blitz, D. L., Carlozzi, N. E., Havlik, R. J., Beaumont, J. L., Mungas, D., Manly, J. J., Borosh, B. G., Nowinski, C. J., and Gershon, R. C. (**2013**). "I. NIH toolbox cognition battery (CB): Introduction and pediatric data," Monographs Soc. Res. Child. **78**, 1–15.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Slotkin, J., Carlozzi, N. E., Bauer, P. J., Wallner-Allen, K., Fox, N., Havlik, R. J., Beaumont, J. L., Mungas, D., Manly, J. J., Moy, C., Conway, K., Edwards, E., Nowinski, C. J., and Gershon, R. C. (**2014**). "The cognition battery of the NIH toolbox for Assessment of neurological and behavioral function: Validation in an adult sample," J. Int. Neuropsychol. Soc. **20**, 567–578.

Wong, P. C. M., and Perrachione, T. K. (**2007**). "Learning pitch patterns in lexical identification by native English-speaking adults," Appl. Psycholinguist. **28**, 565–585.

Yoo, J., and Bidelman, G. M. (**2019**). "Linguistic, perceptual, and cognitive factors underlying musicians' benefits in noise-degraded speech perception," Hear. Res. **377**, 189–195.