Non-native phonetic learning is destabilized by exposure to phonological variability before and after training

Pamela Fuhrmeister[a] and Emily B. Myers[b]
Department of Speech, Language, and Hearing Sciences, University of Connecticut, 850 Bolton Road, Storrs, Connecticut 06269, USA
pamela.fuhrmeister@uconn.edu, emily.myers@uconn.edu

Abstract: Phonological variability is a key factor in many phonetic training studies, but it is unclear whether variability is universally helpful for learners. The current study explored variability and sleep consolidation in non-native phonetic learning. Two groups of participants were trained on a non-native contrast in one vowel context (/u/) and differed in whether they were also tested on an untrained context (/i/). Participants exposed to two vowels during the test were less accurate in perception of trained speech sounds and showed no overnight improvement. These findings suggest that introducing variability even in test phases may destabilize learning and prevent consolidation-based performance improvements.

[DDO]
Date Received: June 11, 2017 **Date Accepted:** October 18, 2017

1. Introduction

Learning to perceptually distinguish speech sounds in another language can be challenging for adults, especially when novel speech sounds are perceptually similar to speech sounds in the native language (Best *et al.*, 2001). For example, native English speakers often struggle to perceive differences in the dental and retroflex voiced stop consonants found in Hindi, due to the perceptual similarity of both dental and retroflex sounds to the alveolar /d/ sound. For this reason, a great deal of work has focused on developing optimal training paradigms to promote learning of these difficult phonetic contrasts. Sleep has been found to facilitate learning in a variety of linguistic domains (e.g., Davis and Gaskell, 2009; Fenn *et al.*, 2013, 2003), and recent studies have begun to explore the potential benefits of sleep in phonetic learning as well. In fact, many non-native speech sound learning studies from our lab consistently show performance improvements following sleep even in the absence of further training (Earle *et al.*, 2017; Earle and Myers, 2015a,b). For example, Earle *et al.* (2017) found that sleep duration predicted overnight improvement on discrimination and identification of a non-native Hindi contrast, and a similar phonetic training study by Earle and Myers (2015a) observed generalization to tokens produced by a novel talker following a period of sleep. However, in the case of non-native phonetic learning, learners may not benefit from sleep-mediated consolidation if, prior to sleep, they have extensive exposure to native language sounds that are perceptually similar to the trained non-native sounds. Specifically, in Earle and Myers (2015b), participants who were trained during the evening hours to learn a non-native phonetic contrast improved after sleep; however, those trained in the morning hours saw no such benefit from the overnight interval. The authors reasoned that this effect was a result of differential amounts of exposure to the participants' native language prior to sleep: participants trained in the evening had minimal exposure to their native language before sleep, while those trained in the morning had a day's worth of native language exposure prior to sleep. This suggests that learners may need to achieve a certain level of stability in their learning in order to protect against interference (in this case, from one's native language) and consolidate the information learned.

What is unclear, however, is what types of stimuli have the potential to interfere with learning or consolidation of non-native speech sounds and at what point in the learning or consolidation trajectory newly formed speech categories are most vulnerable to interference. Phonological variability is one potential source of interference.

---

[a]Author to whom correspondence should be addressed.
[b]Also at: Haskins Laboratories, 300 George Street, New Haven, CT 06511, USA.

Given that many non-native phonetic training studies probe for generalization to new vowel contexts in pre- and post-test materials, we asked whether the introduction of variability even in test blocks might affect learning of a novel speech contrast. Specifically, the current study seeks to determine whether exposure to phonological variability in non-native speech training paradigms unequivocally improves learning and whether exposure to phonological variability interacts with sleep-dependent consolidation. To test this, we trained two groups of participants on the Hindi dental-retroflex voiced stop contrast. Some studies have shown that variability in training is difficult for learners with poorer perceptual abilities (Amitay *et al.*, 2005; Antoniou and Wong, 2015; Perrachione *et al.*, 2011); therefore, in the current study, we chose to train all participants on stimuli presented in a single vowel context (/u/) and spoken by a single talker in order to increase the likelihood that all participants would benefit from training. Because previous work found overnight improvement for participants trained in the evening but not the morning hours (Earle and Myers, 2015b), we trained all participants in the evening and reassessed their performance the following morning. In one group, vowel-contextual variability was introduced by testing listeners on the dental and retroflex sounds presented in both /u/ and /i/ vowel contexts, whereas the other group heard the sounds presented only in the trained vowel context at test (Fig. 1). If any exposure to variability enhances non-native phonetic learning, we predict that extra exposure to the contrast in a different phonological context will result in category formation that is more stable and more able to benefit from consolidation during sleep. However, if incidental exposure to variability during testing results in learning that is weaker or less stable, we predict that exposure to variability even in testing only could prevent learners from demonstrating overnight improvement.

2. Methods

2.1 Participants

Sixty-eight participants (18 male, 43 female, ages 18–24) were recruited through the Psychology Department Participant Pool at the University of Connecticut and were either paid $10 per hour or given course credit for their participation. The study was advertised to monolingual, native speakers of American English only, and participants reported having typical hearing and vision and no history of speech or language disorders. All participants gave informed consent according to the University of Connecticut Institutional Review Board procedures.

2.2 Stimuli

Audio and visual stimuli were presented using E-Prime 2.0 [Psychology Software Tools (2012), Pittsburgh, PA]. Audio stimuli were presented at 75 dB sound pressure level (SPL) or a comfortable listening level using over-ear headphones (SONY MDR-7506, New York). Participants indicated their responses via mouse click. Audio stimuli consisted of five acoustically distinct recordings each of /d̪ug/, /ɖug/, /d̪ig/, and /ɖig/ recorded by a male, native Hindi speaker, and the sound files were rescaled to a mean amplitude of 70 dB SPL. This stimulus set was previously used in Earle and Myers (2015b) and the reader is referred to that article for a more complete description of stimulus creation.

2.3 Training and assessments

AX Discrimination Assessment. To assess discrimination of the trained speech sounds, participants completed an AX discrimination task consisting of two presentations of either "same" tokens (two tokens beginning with the speech sounds from the same

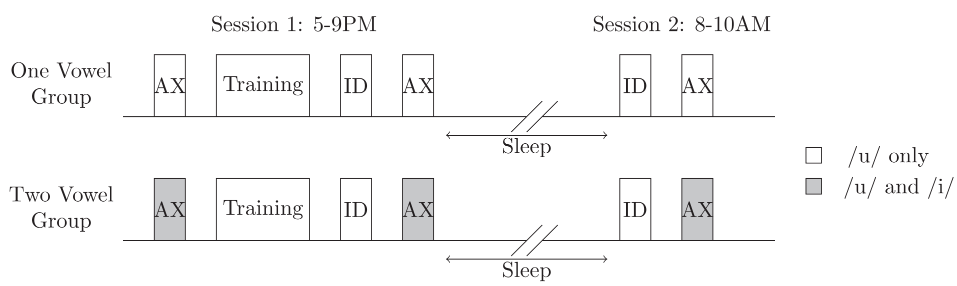

Fig. 1. Training and testing schedule for each group. The One Vowel group was exposed to the Hindi dental-retroflex contrast in one vowel context only (/u/) throughout the entire training protocol, while the Two Vowel group was exposed to the contrast in two vowel contexts (/u/ and /i/) during discrimination assessments only. Both training and identification assessments were identical for each group.

category but acoustically distinct exemplars) or "different" tokens (two tokens beginning with speech sounds from different categories) with a 1-s interstimulus interval separating the two. To measure the effects of variability introduced during testing, participants were assigned to one of two groups: one group was presented with the dental and retroflex tokens in only the /u/ vowel context (One Vowel group), and the other group was presented with /u/ and /i/ vowel contexts during the discrimination assessment at each time point (Two Vowel group). The stimuli consisted of the same tokens used in training, with the addition of an untrained vowel context for the Two Vowel group: (/d̪ug/, /ɖug/, /d̪ig/, and /ɖig/). Trials for each of the vowel contexts for the Two Vowel group were randomized.

Training. Participants in both groups were trained to categorize the Hindi dental and retroflex voiced stop consonants by means of a two-alternative forced choice identification task. Prior to training, participants were familiarized with the pairing of a novel visual stimulus (Fribbles)[1] that corresponded to each word. During familiarization, each visual stimulus was presented separately on the screen, and participants listened to five different exemplars of the corresponding word. For each trial during the training task, the two novel visual objects appeared side by side on the screen, and participants heard one of two minimal pair words beginning with either the dental or retroflex sound (/d̪ug/ or /ɖug/) and were asked to pair the word with the correct visual stimulus. Participants received visual feedback for each response ("Correct!" or "Incorrect"). Training consisted of a total of 200 trials with a 2-min break halfway through. Both groups of participants were exposed to only one vowel context (/u/) for all training trials.

Identification assessment. Assessment in identification consisted of 50 trials identical to the training procedure, except feedback was not given. Unlike the AX discrimination task, participants in both groups performed the identification task on only the trained vowel context.

2.4 Schedule

All participants came to the lab for two sessions: the first session took place in the evening hours (5–9 p.m.), and participants returned the following morning (8–10 a.m.) for reassessment. The evening session consisted of a baseline AX discrimination assessment, followed by the training task, and finally, post-training identification and AX discrimination assessments. The morning session included only identification and AX discrimination reassessments to measure retention of learning after the overnight interval (see Fig. 1).

3. Analyses and results

Eight of the original 68 participants were excluded from the analyses: two for not completing the experiment, three because of an experimenter error, two due to noncompliance with experimental tasks, and one participant was already able to discriminate the contrast at baseline. The remaining 60 participants (One Vowel group: $n = 31$, Two Vowel group: $n = 29$) are included in the following analyses (see Fig. 2).

3.1 Identification performance

To assess changes in identification performance of the trained vowel over time, percent accuracy scores were submitted to a 2×2 repeated measures analysis of variance

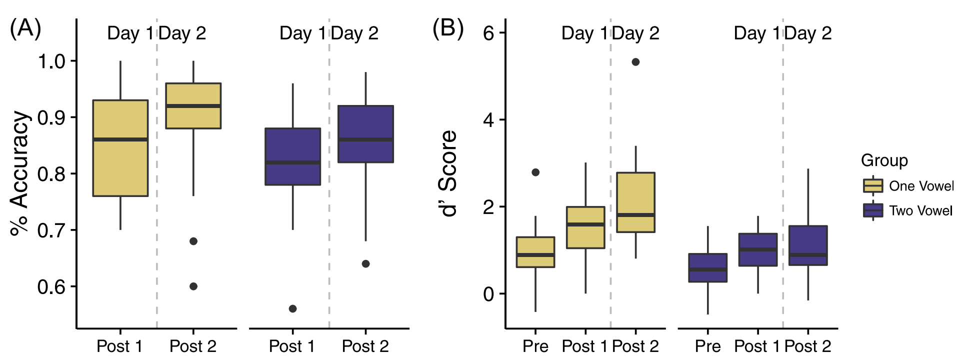

Fig. 2. (Color online) (A) Left panel shows percent accuracy in identification assessments on the first day (Post 1) and upon returning the next morning (Post 2). (B) Right panel shows d′ scores for the AX discrimination task, performed before training (Pre), immediately following training on the first day (Post 1), and upon returning to the lab the following morning (Post 2). Data in all analyses contain only the trained vowel context (/u/).

(ANOVA) with Group (One Vowel or Two Vowel) as the between-subjects factor and Time (Post-test 1 vs Post-test 2) as the within-subjects factor. The ANOVA revealed a significant main effect of group, $F(1,58) = 4.59$, $p = 0.04$, reflecting lower accuracy in the Two Vowel group, a significant main effect of Time, $F(1,58) = 5.84$, $p = 0.02$, and no interaction, $F(1,58) = 0.94$, $p = 0.34$. Because we were interested in how variability might differentially affect overnight consolidation, we carried out planned *t*-tests to measure overnight improvement in each group. Planned *t*-tests indicated overnight improvement for the One Vowel group from Post-test 1 ($M = 0.86$, $SE = 0.02$) to Post-test 2 ($M = 0.90$, $SE = 0.02$), $t(30) = -2.34$, $p = 0.03$, while the Two Vowel group did not improve overnight (Post-test 1: $M = 0.82$, $SE = 0.02$, Post-test 2: $M = 0.84$, $SE = 0.02$), $t(28) = -1.05$, $p = 0.30$. This finding should be interpreted with some caution, however, given the lack of a significant Time by Group interaction. In order to test for group differences in identification performance immediately after training, an additional planned *t*-test was carried out to compare the performance of both groups at Post-test 1. The planned *t*-test did not show a significant difference between the groups immediately after training, $t(58) = 1.56$, $p = 0.12$. This lack of a significant difference on the first post-test suggests that both groups were able to benefit comparably from training. However, the fact that the Two Vowel group did not improve after sleep indicates that the presence of the second vowel context in the discrimination tasks may have destabilized their learning of the contrast in the trained vowel context, which did not allow for overnight improvement even on the trained task (identification), although it was identical between the groups.

3.2 Discrimination performance

To assess changes in discrimination performance in the context of the trained vowel (/u/) over time, percent accuracy scores were converted to d′ scores to account for response bias (MacMillan and Creelman, 2004), and these scores were submitted to a 2×3 repeated measures ANOVA with Group as the between-subjects factor and Time (Baseline, Post-test 1, Post-test 2) as the within-subjects factor. The ANOVA revealed a significant main effect of Group, $F(1,58) = 19.86$, $p < 0.001$, with lower accuracy in general for the Two Vowel group; a significant main effect of Time, $F(2,116) = 30.59$, $p < 0.001$; and a significant interaction, $F(2,116) = 4.52$, $p = 0.01$. To further investigate this interaction, we carried out *post hoc* analyses to test for improvement of each group at each time point, using the Holm-Bonferroni method to correct for multiple comparisons. Paired *t*-tests indicate significant improvement for the One Vowel group from the Baseline measure ($M = 0.92$, $SE = 0.12$) to Post-test 1 ($M = 1.57$, $SE = 0.12$), $t(30) = -3.68$, $p = 0.004$, and significant overnight improvement from Post-test 1 to Post-test 2 ($M = 2.08$, $SE = 0.18$), $t(30) = -4.12$, $p = 0.001$. Thus, participants in this group were able to benefit from overnight consolidation. Like the One Vowel group, the Two Vowel group showed significant improvement from the Baseline measure ($M = 0.61$, $SE = 0.09$) to Post-test 1 ($M = 0.97$, $SE = 0.10$), $t(28) = -3.78$, $p = 0.001$; however, no such improvement was found from Post-test 1 to Post-test 2 ($M = 1.12$, $SE = 0.14$), $t(28) = -1.29$, $p = 0.83$. This suggests participants in the Two Vowel group were not able to improve as a result of sleep-dependent consolidation.

4. Discussion

In this experiment, we asked whether exposure to vowel-contextual variability in the assessments of a non-native phonetic training paradigm would affect learning and consolidation of the non-native contrast trained in a single vowel context. Despite the fact that the Two Vowel group received twice as much exposure to the contrast in the discrimination assessments, the Two Vowel group's performance was significantly poorer during discrimination of the speech sounds in the trained vowel context as compared to the One Vowel group. One interpretation of this pattern is that participants found it harder to perform the discrimination task when there was greater variability in the test set (two vowels compared to one). Strikingly, this pattern is not limited to the discrimination data: the Two Vowel group's identification of the contrast in the trained vowel context was significantly poorer than the One Vowel group even though these blocks were identical across groups. Taken together, this suggests that the vowel-contextual variability disrupted learning of the contrast in the trained vowel context. This difference in performance, especially on identification, lends support to our interpretation of the data, namely that exposure to variability introduced during test destabilized learning. Furthermore, only the One Vowel group showed evidence of overnight improvement on each task, while the Two Vowel group showed no such improvement after sleep, suggesting that skills that are more strongly learned are more able to benefit from overnight consolidation.

Our results are consistent with the idea that strength and stability of learning is a key factor in sleep-dependent consolidation for non-native speech sound learning. Indeed, perceptual learning studies in the visual domain support this notion. One such study by Shibata *et al.* (2017) found that participants who achieved more stability in their learning of a visual skill by overlearning (continuing to train on a skill after mastery) were not only protected against interference from a second task, but they also retained learning of the first task when assessed the following day as compared to a group that received the typical amount of training. A similar study found that participants who trained on a perceptual task until performance reached asymptote improved after a period of sleep, but those who stopped training prior to that point did not, which suggests a role for stability in sleep-mediated consolidation of newly learned skills (Hauptmann *et al.*, 2005). In our study, exposure to the contrast in an untrained vowel context may have destabilized learning, thus disrupting the overnight consolidation process.

Studies using high-variability training paradigms have often found benefits of exposing participants to variability during training (e.g., Bradlow *et al.*, 1999, 1997; Lively *et al.*, 1993, 1994; Logan *et al.*, 1991). Our results do not necessarily conflict with these findings; rather, they raise several questions about the learning trajectory in the acquisition of novel speech categories. In contrast to the present study, training tasks in previous high-variability training studies have typically included explicit feedback on the sounds presented by different talkers or in different phonological contexts (e.g., Logan *et al.*, 1991; Lively *et al.*, 1993). That is, if participants in our study had received feedback during training for the sounds in both vowel contexts, outcomes may have been more favorable. This is an empirical question that will need to be addressed with future research. In addition, training protocols in studies testing the efficacy of high-variability training have typically taken place over the course of several weeks (e.g., Logan *et al.*, 1991); therefore, it is reasonable to assume that this type of training is indeed beneficial in the long term when establishing new phonetic categories. However, in the short term, and in order for participants to capitalize on sleep-mediated consolidation, it may be optimal to limit variability in the exposure to the novel speech sounds that participants receive. Many perceptual training studies of non-native speech sounds test for generalization to novel talkers or phonological contexts; however, the current results suggest that even introducing variability at testing may impede overnight consolidation of a phonological contrast trained in a single vowel context, suggesting that it may be advisable to include such assessments only at the very end of the training protocol. This may be especially important in multi-day training studies, as our results suggest that vowel-contextual variability introduced during assessments leads to less stable learning that is not as able to benefit from consolidation.

We have shown that exposure to phonological variability during testing destabilizes learning; however, the question remains as to what strategies participants use during discrimination assessments. One possibility is that participants were engaged in some type of implicit learning during the discrimination task. For example, the presentation of the various tokens during the task may have allowed listeners to track the statistical distributions of the two categories (e.g., Maye and Gerken, 2000), and participants in the present study may have been able to utilize this information to improve their performance on both tasks. However, some studies have shown that learners typically rely on explicit strategies in early stages of the learning process (e.g., Chandrasekaran *et al.*, 2014); therefore, implicit learning may not be helpful to participants at this point. If it is the case that participants are learning from the discrimination task, we argue that the Two Vowel group had more opportunities to learn because they received twice as many tokens in the discrimination assessments and should therefore show more improvement. Notably, this is not what we found. We remain agnostic as to what type of learning strategy (if any) participants adopt during discrimination assessments and simply point out that our results are consistent with the notion that variability in different types of memory traces results in less retention of learning.

The optimal schedule for phonetic training has yet to be determined. On the one hand, the present results suggest that overnight consolidation of the limited set of instances trained here benefits from a lack of variability. On the other hand, we argue that the most robust test of category formation is listeners' ability to generalize to novel contexts (e.g., novel talkers or phonological contexts). Variability may play a key role in enabling such generalization. Sleep has additionally been shown to aid in abstraction and generalization of information after perceptual training on synthetic speech tokens (Fenn *et al*., 2013, 2003), in the lexical domain (e.g., Davis and Gaskell, 2009), and in non-native phonetic learning (Earle and Myers, 2015a). In fact, the study by Earle and Myers (2015a) found that sleep-mediated consolidation allowed participants to generalize to novel talkers after training on a closed set of tokens spoken by a single talker. In this sense, consolidation processes (in the absence of variability) may be sufficient for generalization. The division of labor of these two forces that aid generalization—variability in the input and consolidation processes—is the key to determining optimal methods for perceptual training. Our results highlight this tension and show that limiting variability in the early stages of learning may be advantageous for enabling sleep-mediated gains.

Acknowledgments

This work was supported by Grant No. R01 DC013064 to E.B.M., an NSF CAREER grant to E.B.M. (Grant No. 1554810), and Grant No. NSF IGERT DGE-1144399 to the University of Connecticut. The content is the responsibility of the authors and does not necessarily represent official views of the NIH, NIDCD, or NSF. The authors are grateful to Brianna Schlemmer for her help with data collection and two anonymous reviewers for helpful comments on a previous version of the manuscript.

References and links

[1]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/.

Amitay, S., Hawkey, D. J. C., and Moore, D. R. (**2005**). "Auditory frequency discrimination learning is affected by stimulus variability," Percept. Psychophys. **67**(4), 691–698.

Antoniou, M., and Wong, P. C. M. (**2015**). "Poor phonetic perceivers are affected by cognitive load when resolving talker variability," J. Acoust. Soc. Am. **138**(2), 571–574.

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," J. Acoust. Soc. Am. **109**(2), 775–794.

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

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**(4), 2299–2310.

Chandrasekaran, B., Yi, H.-G., and Maddox, W. T. (**2014**). "Dual-learning systems during speech category learning," Psychonomic Bull. Rev. **21**(2), 488–495.

Davis, M. H., and Gaskell, M. G. (**2009**). "A complementary systems account of word learning: Neural and behavioural evidence," Philos. Trans. R. Soc. B **364**, 3773–3800.

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**(1), EL91–EL97.

Earle, F. S., and Myers, E. B. (**2015b**). "Sleep and native language interference affect non-native speech sound learning," J. Exp. Psychol. **41**(6), 1680–1695.

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

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

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

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**(3), 1242–1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (**1994**). "Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories," J. Acoust. Soc. Am. **96**(4), 2076–2087.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**(2), 874–886.

Macmillan, N., and Creelman, C. (**2004**). *Detection Theory: A User's Guide* (Psychology Press, New York).

Maye, J., and Gerken, L. (**2000**). "Learning phonemes without minimal pairs," in *Proceedings of the 24th Annual Boston University Conference on Language Development*, Vol. 2, pp. 522–533.

Perrachione, T. K., Lee, J., Ha, L. Y. Y., and Wong, P. C. (**2011**). "Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design between individual differences and training paradigm design," J. Acoust. Soc. Am. **130**(1), 461–472.

Psychology Software Tools, Inc. [E-Prime 2.0]. (**2012**). https://www.pstnet.com (Last viewed October 27, 2017).

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," Nature Neurosci. **20**, 470–475.