



OPEN

# Transcutaneous cervical vagus nerve stimulation enhances second-language vocabulary acquisition while simultaneously mitigating fatigue and promoting focus

Toshiya Miyatsu<sup>1</sup>✉, Vanessa Oviedo<sup>1,2</sup>, Jajaira Reynaga<sup>1,2</sup>, Valerie P. Karuzis<sup>3</sup>, David Martinez<sup>3</sup>, Polly O'Rourke<sup>3</sup>, Melissa Key<sup>4,5</sup>, Lindsey McIntire<sup>4,5</sup>, William Aue<sup>4</sup>, Richard McKinley<sup>4</sup>, Peter Pirolli<sup>1</sup> & Timothy Broderick<sup>1</sup>

Transcutaneous vagus nerve stimulation (tVNS) is a promising technique for enhancing cognitive performance and skill acquisition. Yet, its efficacy for enhancing learning rate and long-term retention in an ecologically valid learning environment has not been demonstrated. We conducted two double-blind sham-controlled experiments examining the efficacy of auricular tVNS (taVNS: Experiment (1) and cervical tVNS (tcVNS: Experiment (2), on a 5 day second-language vocabulary acquisition protocol among highly selected career linguists at the US Department of Defense's premier language school. tcVNS produced accelerated recall performance during training (Day 2–4), benefits of which were maintained across a 24 h retention interval with no stimulation at the final test. Consistent with prior work, tcVNS also produced fatigue-mitigating and focus-promoting effects as measured by the Air Force Research Laboratory Mood Questionnaire. Based on the current and the previous findings supporting tVNS' efficacy on performance, training enhancement, and fatigue mitigation, we believe tcVNS to be an effective learning acceleration tool that can be utilized at language-teaching and other institutions focused on intensive training of cognitive skills.

Cognitive performance and skill acquisition are central to the mission of institutions in a wide variety of sectors, such as education, commercial, and military. As such, there is a long history of research aimed at discovering and refining behavioral, pharmacological, and technological techniques that can enhance cognition<sup>1</sup>. Enabled by recent technological advances and demonstration of safety, there are surging interests in applying peripheral nerve stimulation for these purposes<sup>2,3</sup>. For example, the vagus nerve can now be stimulated non-invasively over the skin with minimal side effects<sup>3</sup>. As will be reviewed later, vagus nerve stimulation has been shown to enhance performance across a variety of learning tasks. Yet, its effect on ecologically valid long-term learning where the learning events take place over multiple days and the learning is accessed after a delay (e.g., 24 h) is sorely lacking. To fill this gap, we investigated the efficacy of two prominent types of non-invasive vagus nerve stimulation on learning throughout a five consecutive days of second-language vocabulary acquisition protocol among career linguists under training.

Transcutaneous vagus nerve stimulation (tVNS) is hypothesized to enhance cognitive performance and skill acquisition by eliciting the release of key neurotransmitters that modulate cognitive performance and learning. Both of the two common forms of tVNS, transcutaneous auricular VNS (taVNS) and transcutaneous cervical VNS (tcVNS), activate the afferent fiber of the vagus nerve. The electrical stimulation is hypothesized to travel through the afferent fibers and reaches the nucleus tractus solitarius (NTS) in the brainstem. From the NTS, several neural pathways are engaged, including projection to the locus coeruleus (LC), the primary source of

<sup>1</sup>Florida Institute for Human and Machine Cognition, Pensacola, USA. <sup>2</sup>University of California Santa Cruz, Santa Cruz, USA. <sup>3</sup>University of Maryland Applied Research Lab for Intelligence & Security, College Park, USA. <sup>4</sup>Air Force Research Laboratory, Wright-Patterson AFB, USA. <sup>5</sup>DCS Corp., Alexandria, USA. ✉email: tmiyatsu@ihmc.org

norepinephrine (NE) in the brain<sup>4,5</sup>. Activation of LC results in noradrenergic projections to brain regions critical for cognitive performance and learning, such as the hippocampus, the amygdala, the thalamus, and the prefrontal cortex. Additionally, the NTS also connects to other areas involved in cognitive and autonomic regulation, such as the nucleus basalis which is involved in acetylcholine release<sup>6</sup>, the dorsal raphe nucleus (DRN) which is involved in serotonin release<sup>7</sup>, and the parabrachial nucleus (PBN) which plays a role in processing sensory information and modulating autonomic functions through its innervations to GABA-rich regions, such as the thalamus and hypothalamus<sup>8</sup>.

The cascading neural activation outlined above could benefit learning in at least two ways. First, when applied prior to a learning event, the increase in the key neurotransmitters could bring about enhanced arousal and attention<sup>9</sup> while encoding the to-be-learned information. Second, when applied following a learning event, the increased activation of the LC-NE system could enhance the consolidation of the learned information, a well-established phenomenon in the clinical literature<sup>10,11</sup>. To foreshadow, we tested a procedure that delivered tVNS both prior to and following a learning event (i.e., priming and consolidation stimulation) to take advantage of both of these potential mechanisms.

It is well-established that both taVNS and tcVNS can enhance cognitive performance and learning in some situations. taVNS has been shown to enhance processing speed and response selection<sup>12</sup>, inhibitory control<sup>13,14</sup>, post-error slowing<sup>15</sup>, conflict-triggered adjustment of cognitive control<sup>16</sup>, divergent thinking<sup>17</sup>, fear extinction<sup>18,19</sup>, emotion recognition from facial expressions<sup>20,21</sup>, and attention to social cues<sup>22</sup>. Pertinent to the current study, taVNS has also been shown to enhance learning of face-name pairings, an associative learning task similar to second-language vocabulary learning<sup>23</sup>, as well as learning of tones of Mandarin<sup>24,25</sup>. tcVNS has been shown to enhance vigilance and multitasking under sleep deprivation<sup>26</sup>, visual target identification reaction time and retention<sup>27</sup>, non-verbal abstract reasoning<sup>28</sup>, and emotional image recognition<sup>29</sup>. Yet, whether the noradrenergic and/or cholinergic facilitation of cognition translates to long-term learning in an applied context is unclear<sup>29</sup>. Further, not all published studies have demonstrated positive effects of tVNS on memory performance<sup>30–32</sup>.

In the current study, we tested the efficacy of auricular (Experiment 1) and cervical tVNS (Experiment 2) on a second-language vocabulary acquisition protocol across 5 consecutive days among career linguists under training at the US military's premier language school, the Defense Language Institute (DLI). On each of the training days, participants received active or sham stimulation before and after (i.e., priming and consolidation stimulation) the learning session of the day which involved trying to recall and getting feedback on 100 target vocabulary (See Fig. 1). Of note, these officers were highly selected based on their performance on a standardized test measuring language-learning aptitude (see Methods) and are tasked to complete a challenging 64-week curriculum with 50% passing rate to be a professional linguist capable of carrying out military missions. As the vocabulary size has been consistently implicated to be the single most significant factor in second-language learning accounting for 50–70% of proficiency gains<sup>33–35</sup>, and vocabulary learning of up to 30 words per day is a particularly challenging aspect of DLI's curriculum<sup>36</sup>, methods for accelerating vocabulary acquisition is highly sought after.

Experiment 1 enrolled students who were studying Mandarin or Farsi, and a taVNS device was used while Experiment 2 enrolled students who were studying Arabic and a tcVNS device was used. Because the design and analysis for Experiments 1 and 2 are mostly identical, we present the results and methods of the two experiments in the same sections below for brevity.

## Results

### Language learning task performance

We took a mixed-effects modeling approach to leverage the information unique to the participants and the studied items (see Methods). We built separate models for Session 2–4 (learning models assessing the tVNS effect during the training) and Session 4–5 (retention models assessing whether the rate of retention differed between conditions). Below, we report the results on the vocabulary learning task for Mandarin learners in Experiment 1 (taVNS) and Arabic learners in Experiment 2 (tcVNS). As the Farsi data for Experiment 1 were collected to supplement the Mandarin data during a recruitment difficulty due to Covid-19 pandemic, there was not a sufficient number of participants to analyze the Farsi learners separately. We elected to report the Experiment 1 analysis combining Mandarin and Farsi learners in the Supplemental Materials (Table S2) for brevity. The statistical significance of all the reported comparisons below remained the same with the Farsi learners excluded or included.

#### Learning models: recall accuracy

Two separate logistic mixed-effects models for Experiment 1 and 2 were built to predict item-level accuracy (0 or 1) over learning sessions 2, 3, and 4. The maximum model structure provided to buildmer was: Accuracy ~ Session × Treatment + Baseline Recall Score + (Session | Participant) + (Session × Treatment | Item).

**Experiment 1—Mandarin recall accuracy during learning.** Figure 2a (left panel) plots participants' raw mean recall accuracy across 5 sessions. The final model is presented in Table 1. There was a significant positive effect of session, indicating that overall, participants improved on recall accuracy across the learning sessions. Baseline recall score was also a significant positive predictor, such that participants who performed higher on baseline recall accuracy were associated with higher recall accuracy performance across sessions 2–4. There was no significant effect of tVNS treatment, or significant interaction between tVNS treatment and session.

**Experiment 2—Arabic recall accuracy during learning.** Figure 2a (right panel) plots participants' raw mean recall accuracy across 5 sessions. The final model is presented in Table 1. There is a significant positive effect of tVNS treatment, indicating that participants who received tVNS performed better overall. Figure 2b plots the model estimate of this significant tVNS effect on recall. There was also a significant positive effect of session,

## a Experiment 1 Device & Control



### Xen by Neuvana

After calibration, control participants were told that they would receive sub-threshold stimulation, but actually received no stimulation

## b Experiment 2 Device & Control



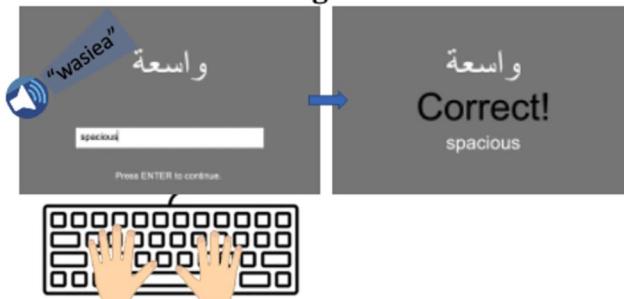
### gammaCore by electroCore

A sham device that emitted sound and vibrations similar to the active device but gave no stimulation

## c Experiment 1 & 2 Protocol

DAY 1: BASELINE TEST	DAY 2: TRAINING	DAY 3: TRAINING	DAY 4: TRAINING	DAY 5: FINAL TEST
• Consent • Eligibility Screening • Mood Questionnaires • Baseline Language Learning Test (Recall only)	• Mood Questionnaire • <u>Priming Stimulation</u> • Language Learning task • <u>Consolidation Stimulation</u> • Language Learning Test (Recall & Recognition) • Mood Questionnaires	• Mood Questionnaire • <u>Priming Stimulation</u> • Language Learning task • <u>Consolidation Stimulation</u> • Language Learning Test (Recall & Recognition) • Mood Questionnaires	• Mood Questionnaire • <u>Priming Stimulation</u> • Language Learning task • <u>Consolidation Stimulation</u> • Language Learning Test (Recall & Recognition) • Mood Questionnaires	• Mood Questionnaire • <u>Priming Stimulation</u> • Language Learning task • <u>Consolidation Stimulation</u> • Language Learning Test (Recall & Recognition) • Mood Questionnaires
				• Mood Questionnaire • Final Language Learning Test (Recall & Recognition) • Mood Questionnaire • Debriefing

## d Feedback Learning Trial



## e Recognition Trial



**Figure 1.** Methods. (a) The auricular tVNS device used for Experiment 1: Xen by Neuvana. (b) The cervical tVNS device used for Experiment 2: gammaCore by electroCore. (c) Schematic illustration of the 5 day experimental protocol. (d) Example of a feedback learning trial. (e) Example of a recognition test trial.

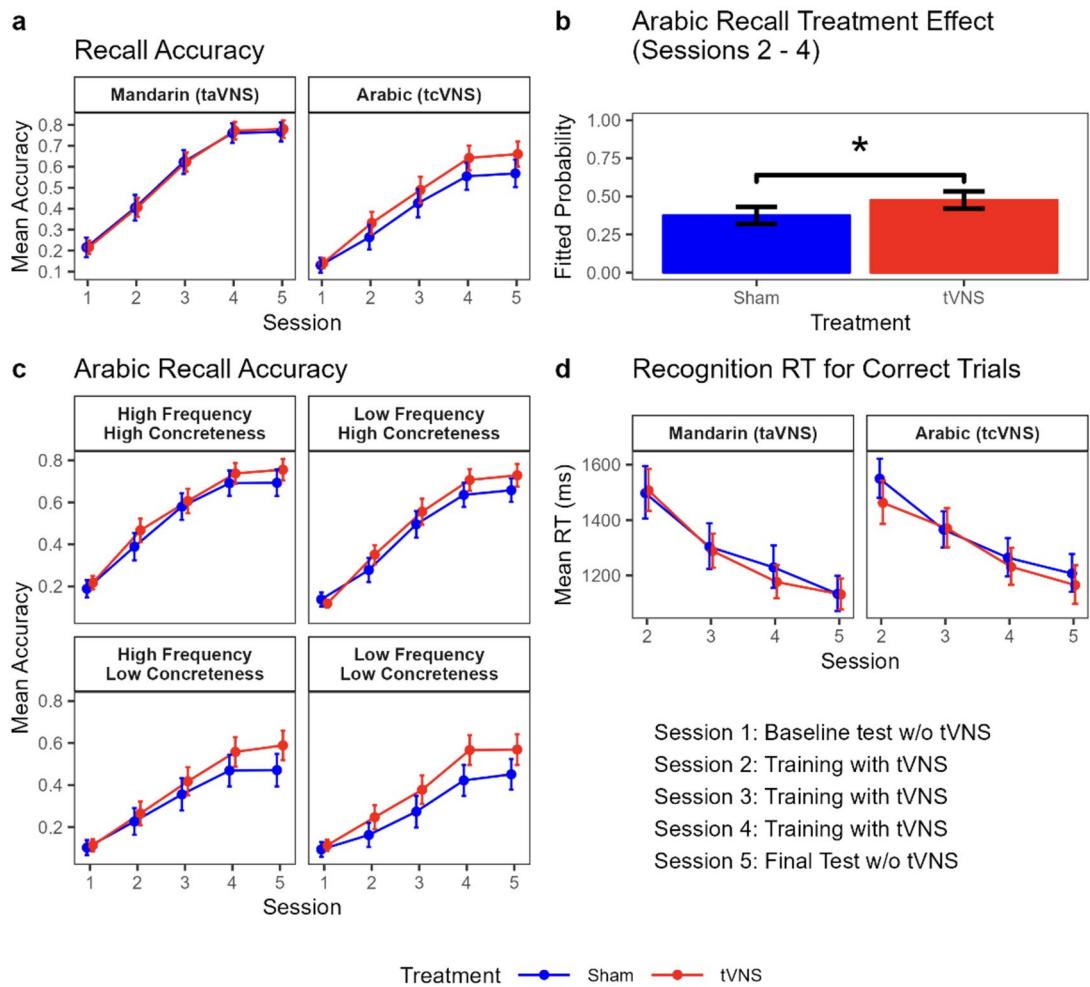
indicating that overall, participants improved across the sessions. There was no significant interaction between tVNS treatment and session. Figure 2c shows the mean performance across 5 days broken down by the item type. At a first glance, it appears that the benefits of tCvNS are more prominent among more difficult items (i.e., low word frequency and low concreteness). We elect to report the analyses detailing this interaction between the effect of tVNS and word type on vocabulary learning in a later paper that is more focused on the linguistic aspects of the study.

### Learning models: recognition reaction time

Two separate linear mixed-effects models were built to predict item-level log reaction time to correct trials across sessions 2 through 4, the learning sessions. The maximal model structure provided to buildmer was:  $\log(\text{RT}) \sim \text{Session} \times \text{Treatment} + (\text{Session} | \text{Participant}) + (\text{Session} \times \text{Treatment} | \text{Item})$ . For these models, no baseline covariate was included, as recognition was not assessed during session 1. The raw mean reaction time was logarithmically transformed (natural log) to make the distribution more symmetrical.

**Experiment 1 – Mandarin recognition log reaction time during learning.** Figure 2d (left panel) plots participants' mean raw reaction time for the correct trials during the recognition test across 4 sessions. Table 1 displays the results of the final model. There was a significant reduction in reaction time to correct trials across learning sessions, across all participants, as indicated by the negative effect of session. There was not a significant effect of tVNS treatment or a significant interaction between session and tVNS treatment.

**Experiment 2 – Arabic recognition log reaction time during learning.** Figure 2d (right panel) plots participants' mean raw reaction time for the correct trials during the recognition test across 4 sessions. The model results (Table 1) include a significant negative effect of session, such that participants got faster to correct trials across training sessions. No significant effects of tVNS treatment or session by tVNS treatment interaction were found.



**Figure 2.** Language learning results. (a) Mean recall performance in Experiment 1 Mandarin participants and Experiment 2 Arabic participants across 5 sessions (i.e., Day 1 through 5). The error bars denote  $\pm 1$  standard error. (b) Model-estimated treatment effect for Arabic Recall Accuracy. The error bars denote 95% confidence intervals. (c) Mean recall performance in Experiment 2 Arabic participants broken down by the item type. (d) Mean raw reaction time for the correct trials during the recognition test in Experiment 1 Mandarin participants and Experiment 2 Arabic participants across 4 sessions (i.e., Day 2 through 5). Note that there was no baseline recognition test on Day 1. The error bars denote  $\pm 1$  standard error.

#### Learning models: summary

As expected, participants' performance improved across training sessions across both measures (recall accuracy and recognition log RT). Of the four models run above, only one of them showed a significant effect of tVNS treatment – Arabic recall accuracy. This model showed that for Experiment 2, participants receiving cervical vagus stimulation performed significantly better than participants in the sham condition. No tVNS effect or interaction was found for auricular tVNS groups in Experiment 1.

#### Retention models (sessions 4 through 5)

A series of models were run to examine performance from sessions 4 to 5. Because session 4 was the last day that participants received training on the vocabulary and tVNS, these models examine whether the participants in the tVNS and the sham conditions differed in the rate of retention of the learned vocabulary information.

#### Retention models: recall accuracy

Two separate logistic mixed-effects models were built to predict item-level accuracy (0 or 1) across session 4 and 5. The maximum model structure provided to buildmer was: Accuracy ~ Session  $\times$  Treatment + (Session | Participant) + (Session  $\times$  Treatment | Item).

**Experiment 1 – Mandarin recall accuracy during retention.** As seen in Table 2, there was no effect of tVNS condition or interaction between session and tVNS treatment. There was also no significant effect of session, indicating that overall, performance on recall accuracy did not significantly decrease the day after training.

Model	Fixed effects	Estimate	SE	z-value	p-value
Mandarin recall	(Intercept)	-0.544	0.226	-2.406	0.016*
	Baseline recall score	9.046	0.653	13.843	<0.001*
	Session	1.520	0.099	15.295	<0.001*
	Treatment	-0.116	0.226	-0.514	0.607
	Session $\times$ treatment	-0.089	0.124	-0.716	0.474
Arabic recall	(Intercept)	-1.822	0.252	-7.226	<0.001*
	Baseline recall score	14.509	1.052	13.798	<0.001*
	Session	1.312	0.083	15.875	<0.001*
	Treatment	0.532	0.238	2.234	0.025*
	Session $\times$ treatment	-0.116	0.096	-1.202	0.229
Mandarin recognition RT	(Intercept)	7.306	0.060	121.370	<0.001*
	Session	-0.101	0.018	-5.503	<0.001*
	Treatment	0.013	0.082	0.163	0.871
	Session $\times$ treatment	-0.028	0.026	-1.075	0.282
Arabic recognition RT	(Intercept)	7.355	0.052	141.366	<0.001*
	Session	-0.106	0.013	-8.134	<0.001*
	Treatment	-0.042	0.071	-0.597	0.551
	Session $\times$ treatment	0.014	0.018	0.783	0.434

**Table 1.** Learning Model Results (Sessions 2 to 4). Estimates refers to unstandardized regression weights. Final random effects (all models): (Session | Participant) + (Session + Treatment | Item). n obs = 12300, n items = 100, n participants = 41 for Mandarin Recall model. n obs = 10800, n items = 100, n participants = 36 for Arabic Recall model. n obs = 11021, n items = 100, n participants = 40 for Mandarin Recognition RT model. n obs = 9232, n items = 100, n participants = 36 for Arabic Recognition RT model.

Model	Fixed effects	Estimate	SE	z-value	p-value
Mandarin recall	(Intercept)	2.441	0.491	4.966	<0.001*
	Session	0.061	0.099	0.612	0.541
	Treatment	-0.237	0.644	-0.368	0.713
	Session $\times$ treatment	0.021	0.138	0.150	0.881
Arabic recall	(Intercept)	0.662	0.540	1.224	0.221
	Session	0.115	0.095	1.213	0.225
	Treatment	0.542	0.726	0.747	0.455
	Session $\times$ treatment	0.030	0.133	0.227	0.821
Mandarin recognition RT	(Intercept)	7.117	0.060	119.170	<0.001*
	Session	-0.081	0.031	-2.631	0.009*
	Treatment	-0.043	0.081	-0.531	0.596
	Session $\times$ treatment	0.044	0.043	1.008	0.313
Arabic recognition RT	(Intercept)	7.150	0.058	124.208	<0.001*
	Session	-0.050	0.017	-2.869	0.004*
	Treatment	-0.028	0.078	-0.360	0.719
	Session $\times$ treatment	-0.012	0.024	-0.486	0.627

**Table 2.** Retention Model Results (Sessions 4 to 5). Estimates refers to unstandardized regression weights. Final random effects (Mandarin, Arabic Recall): (1 | Participant) + (Treatment | Item). Final random effects (Mandarin Recognition RT): (Session | Participant) + (Session | Item). Final random effects (Arabic Recognition RT): (Session | Participant) + (Treatment | Item). n obs = 8200, n items = 100, n participants = 41 for Mandarin Recall model. n obs = 7200, n items = 100, n participants = 36 for Arabic Recall model. n obs = 7709, n items = 100, n participants = 40 for Mandarin Recognition RT model. n obs = 6054, n items = 100, n participants = 36 for Arabic Recognition RT model.

**Experiment 2 – Arabic recall accuracy during retention.** The results can be found in Table 2. There was no significant effect of tVNS treatment, or interaction between session and tVNS condition. There was also no significant effect of session, indicating that overall, performance on recall accuracy did not significantly decrease the day after training. The lack of significant session-by-treatment interaction suggests that there is no evidence of a difference in retention rate of the participants in the tVNS condition relative to the participants in the sham condition.

*Retention models: recognition reaction time*

Two separate linear mixed-effects models were built to predict item-level log reaction time to correct trials across sessions 4 and 5. The maximal model structure provided to buildmer was:  $\log(\text{RT}) \sim \text{Session} \times \text{Treatment} + (\text{Session} | \text{Participant}) + (\text{Session} \times \text{Treatment} | \text{Item})$ .

**Experiment 1—Mandarin recall log reaction time during retention.** Table 2 displays the results of the final model. There was a significant reduction in reaction time to correct trials between the last session and the next day, as indicated by the negative effect of session. No significant effect of tVNS treatment or its interaction with session was found.

**Experiment 2—Arabic recall log reaction time during retention.** The final model (Table 2) has a significant negative effect of session, indicating that reaction time to correct trials decreased across participants between the last training session and the post-test. No significant effect of tVNS treatment or its interaction with session was found.

*Retention models: summary*

For the Session 4 to 5 models examining the potential difference in retention rate after the 24-h delay, no effects of tVNS treatment were found. Recall performance did not significantly increase or decrease over time. Importantly, the retention rate for the recall performance for the tcVNS condition and sham condition (Experiment 2) was almost identical (Estimated  $B = 0.030$ ,  $p = 0.821$ ) indicating that the recall performance advantage that emerged during training (Session 2–4) in the tcVNS condition was sustained at Session 5 when there was no stimulation.

**Mood questionnaire data**

Given the previous empirical demonstration of tVNS's efficacy on fatigue mitigation<sup>26</sup> as well as the tVNS' presumed noradrenergic and cholinergic enhancement of arousal and focus, we targeted the fatigued/energized, calm/excited, and distracted/focused scales, as a priori variables of interest. As such, we only present the results on these scales in this section. However, the results on all other scales are provided in Supplemental Material (Fig. S1). Similar to the language learning task performance analysis, we took a mixed-effects modeling approach, with each mood analyzed as a function of (fixed effects) language (Arabic, Mandarin, Farsi or Mandarin and Farsi combined), treatment (tVNS or sham), session (1–5, treated as categorical), and time (*pre*-priming or *post*-consolidation) plus their interactions, and participant included as a random effect. We have decided to include the Farsi as well as the Mandarin and Farsi combined conditions in the mood questionnaire data analysis as the repeated measures design provided enough data to analyze the Farsi learners on their own and the Mandarin and Farsi combined analysis would give the most statistical power to test whether tVNS affected participants' mood. The Mandarin and Farsi combined data are not depicted in the plots (e.g., Fig. 3). In this section, we only present the results on the critical treatment  $\times$  time interaction on Day 2–4 that examines the tVNS effect on all days on which participants received stimulation. However, the full results are reported in Supplemental Material (fatigued/energized: Fig. S2 & Table S3; calm/excited: Fig. S3 & Table S4; distracted/focused: Fig. S4 & Table S5; worried/peaceful: Fig. S5 & Table S6; contented/disappointed: Fig. S6 & Table S7; frustrated/satisfied: Fig. S7 & Table S8; tranquil/angry: Fig. S8 & Table S9; composed/nervous: Fig. S9 & Table S10; optimistic/pessimistic: Fig. S10 & Table S11; at-ease/restless: Fig. S11 & Table S12; happy/sad: Fig. S12 & Table S13; tense/relaxed: Fig. S13 & Table S14; confused/not-confused: Fig. S14 & Table S15; able/unable: Fig. S15 & Table S16; distressed/delighted: Fig. S16 & Table S17).

*Fatigued/energized*

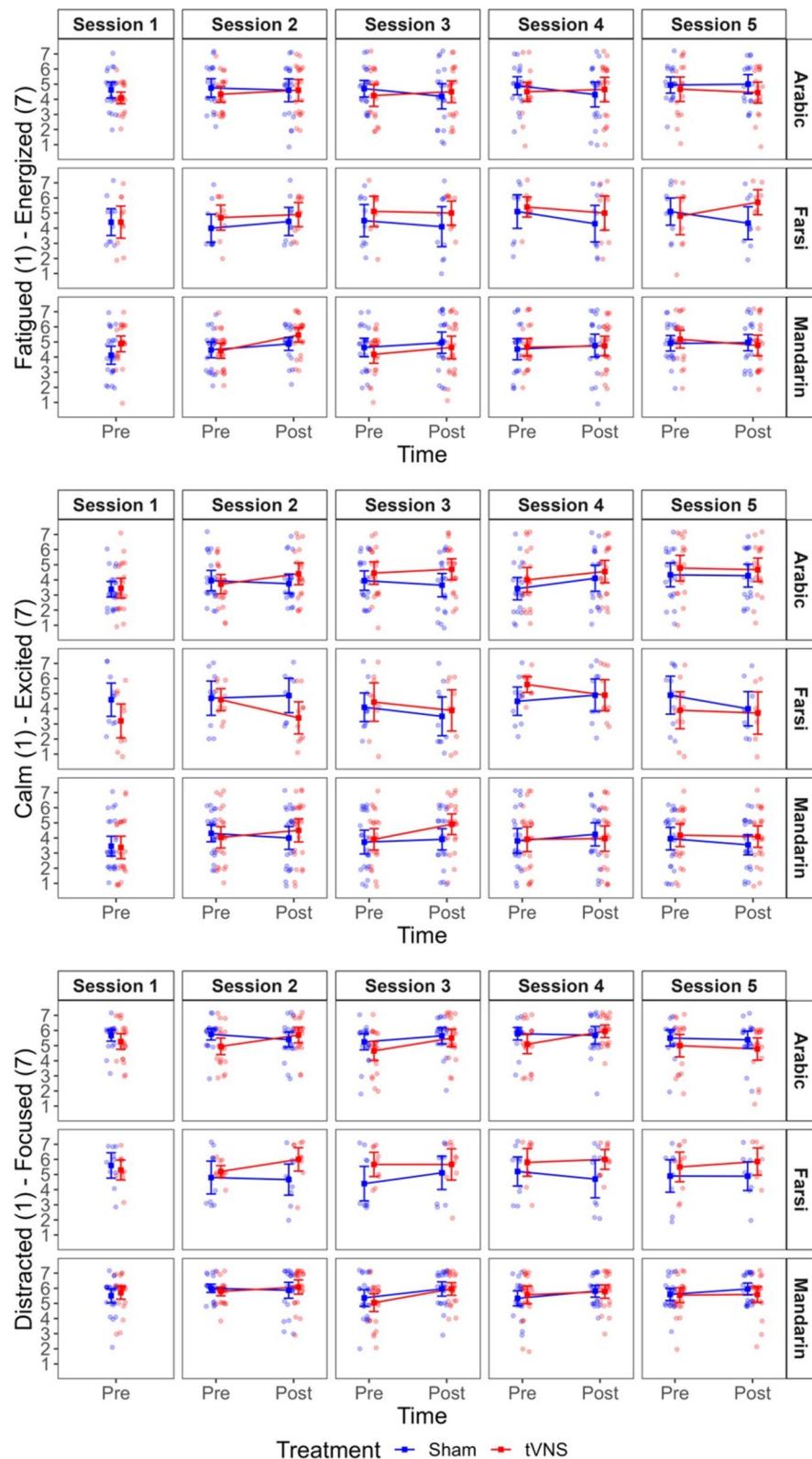
The estimated change in the mean difference between the post-consolidation survey and the pre-priming survey on the fatigued/energized scale across the two treatment groups is shown in the top panel of Fig. 3 and Table 3. Among the critical models examining the tVNS efficacy throughout the training days when stimulations were administered (the top panel of Table 3), Arabic participants receiving tcVNS showed a significant fatigue-mitigating effect,  $p = 0.036$ .

*Calm/excited*

The estimated change in the mean difference between the post-consolidation survey and the pre-priming survey on the calm/excited scale across the two treatment groups is shown in the middle panel of Fig. 3 and Table 3. None of the critical models examining the tVNS efficacy throughout the training days when stimulations were administered (the middle panel of Table 3) showed a significant effect although Arabic and Mandarin participants' rating numerically shifted towards excited as hypothesized.

*Distracted/focused*

The estimated change in the mean difference between the post-consolidation survey and the pre-priming survey on the distracted/focused scale across the two treatment groups is shown in the bottom panel of Fig. 3 and Table 3. Among the critical models examining the tVNS efficacy throughout the training days when stimulations were administered (the bottom panel of Table 3), Arabic participants receiving tcVNS showed a significant fatigue-mitigating effect,  $p = 0.001$ .



**Figure 3.** Participants' ratings on the fatigued/energized, calm/excited, and distracted/focused scales of the Mood Questionnaire in Experiment 1 Mandarin (bottom panel) and Farsi (middle panel) and Experiment 2 Arabic participants (top panel) across 5 sessions (i.e., Day 1 through 5). Note that there was no stimulation on Day 1 and 5, and there was only one measurement on Day 1 as a baseline. The error bars denote a 95% confidence interval about the mean.

Scale	Language	Session	Estimate	SE	Statistic	p-value
Fatigued/energized	Arabic	Sessions 2–4	0.313	0.149	2.103	0.036*
	Farsi	Sessions 2–4	0.077	0.213	0.364	0.716
	Mandarin	Sessions 2–4	0.106	0.140	0.758	0.449
	Mandarin & Farsi	Sessions 2–4	0.092	0.127	0.721	0.471
Calm/Excited	Arabic	Sessions 2–4	0.204	0.169	1.205	0.228
	Farsi	Sessions 2–4	-0.358	0.243	-1.471	0.141
	Mandarin	Sessions 2–4	0.199	0.159	1.252	0.210
	Mandarin & Farsi	Sessions 2–4	-0.079	0.145	-0.546	0.585
Distracted/Focused	Arabic	Sessions 2–4	0.421	0.129	3.271	0.001*
	Farsi	Sessions 2–4	0.157	0.185	0.848	0.397
	Mandarin	Sessions 2–4	0.080	0.121	0.663	0.507
	Mandarin & Farsi	Sessions 2–4	0.119	0.111	1.072	0.283

**Table 3.** Estimated change in the difference between post-consolidation and pre-priming across treatments (tVNS—sham) for the fatigued/energized, calm/excited, and distracted/focused scales of the Mood Questionnaire.

### Mood questionnaire data: summary

From the a priori-selected three scales of the AFRL Mood Questionnaire, Arabic participants receiving tcVNS showed significant increases in energy and focus over the course of each training session compared to sham participants, and their rating on the calm/energized also trended in the hypothesized direction. The results on the rest of the scales of the AFRL Mood Questionnaire are reported in Supplemental Materials.

### Discussion

The current findings on the tcVNS' efficacy on second-language vocabulary acquisition in Experiment 2 is notable for several reasons. First, the tcVNS' learning enhancement was observed among highly selected career linguists under training. The Defense Language Institute (DLI), where the study was conducted, is the Department of Defense's premier language school that consistently trains students to be mission-ready linguists capable of explaining and supporting opinions, hypothesizing, and dealing with unfamiliar topics in 36 to 64 weeks often from little prior experience in the target language<sup>37</sup>. To be admitted into DLI, one has to score high on the Defense Language Aptitude Battery (DLAB), a test designed to measure one's ability to learn new languages<sup>38,39</sup>. The hypothesized mechanism underlying the learning augmentation by tVNS hinges on the noradrenergic and cholinergic enhancement of attention and arousal. Accordingly, it was possible that this highly selected and motivated population could already employ sufficient attention and arousal to effectively learn the foreign vocabulary in our experimental protocol, and thus would not benefit from the tVNS treatment. This was not the case as our results showed that tcVNS enhanced learning even among this highly skilled population, extending our understanding of the efficacy of tcVNS further.

Also significant is that there was no indication that this learning enhancement diminished on Day 5 when there was no stimulation. In a large majority of studies demonstrating the learning enhancement effect of tVNS, the final performance was assessed on the day participants received a stimulation<sup>12,27,28</sup>. By comparison, in the current experiment, participants received no stimulation on the day of the final test (Day 5) and thus took the final test at least 24 h removed from the last stimulation. The findings that there was a significant training effect (Days 2–4) and the retention rate between Day 4 and 5 did not differ between the tcVNS and the sham groups in Experiment 2 suggests that tcVNS can enhance training and the learning that is achieved during the enhanced training can be sustained at a later time when there is no stimulation. Mechanistically, our findings suggest that tVNS' effect on encoding and consolidation is sufficient to produce memory enhancement independent from its effect on the memory retrieval process.

We did not observe statistically significant learning enhancement effects with taVNS. There are a number of potential reasons for the difference between tcVNS and taVNS. First, the tcVNS and taVNS devices and protocol that we used differed in several parameters, such as duration (tcVNS: 2 min × 2 times vs taVNS: 3 min × 1 time), intensity (40.59 mA vs 32.64 mA, as the taVNS was intentionally sub-threshold), and pulse width (1000 µs vs 250 µs; see the Methods section for details). It is possible, for example, that the taVNS did not produce enough stimulation given the lower average intensity and shorter duration. Alternatively, the calibration process employed for the taVNS device to determine the just-noticeable threshold did provide some stimulation to sham participants; this may have minimized a difference between the stimulation groups, especially as the stimulation was already lower for the group receiving stimulation. The whole calibration process took about 3 min. As the effect of taVNS, as measured through pupil dilation and EEG response, can be observed as quickly as within a second of the stimulation<sup>24,40</sup>, this 3-min calibration process might have provided sufficient stimulation to enhance learning in the control group and consequently masked the learning enhancement that emerged in the experimental condition. Additionally, for some participants in the active tcVNS condition, observable physical symptoms (e.g., facial and neck muscle contraction) were present, and thus the double-blinding was compromised.

Separately, there were a variety of linguistic factors that differed between the Mandarin and Arabic word lists and learning tasks. First, Arabic vocabularies were harder to learn (Fig. 2 panel a). As the results depicted in Fig. 2, Panel c show, more difficult items seem to benefit more from tVNS. Thus, it is possible that the difference between taVNS and tcVNS observed across Experiments 1 and 2 was a function of the stimuli difficulty. Stimulus lists were tightly controlled within-language, but word/character length is inherently different between the languages, and the list of English words differed between the languages. Unlike previous studies, this learning paradigm involved both audio and visual presentation of the words. Because Mandarin is tonal, and the words were not explicitly chosen to highlight minimal triplets or quadruplets (e.g., the same consonant–vowel information but differing in tone only) it is possible that the addition of the tonal audio cues enhanced the learning of the sham participants. Additionally, in order to create four sub-lists per study (varying in low and high frequency and concreteness), frequency and concreteness parameters and cutoffs were different per language. It is important to note that due differences in the stimulation parameters, calibration methods, and the different language materials used, the current results are not intended to compare tcVNS and taVNS efficacy directly. Additionally, these results do not rule out the possibility that taVNS enhances second-language vocabulary acquisition. Indeed, in a closely related paradigm, taVNS has been shown to enhance the learning of tone-word pairing that is observed in tonal language, such as Mandarin<sup>24,25</sup>. Rather, the current results provide important stimulation and experimental parameters to consider when implementing taVNS for cognitive enhancement purposes.

Compared with the sham stimulation, there was a significant shift from fatigued to energized mood in the tcVNS condition (Experiment 2) as measured by the fatigued/energized scale of the Mood Questionnaire. While the previous demonstration of this effect was among the participants who went through a 24 h protocol with repeated cognitive testing that mimicked the demand on common shift workers<sup>26</sup>, we have replicated this fatigue-mitigation effect on a sample of participants on their regular circadian cycle, demonstrating the generality of this effect. As fatigue is a prevalent public health and safety issue, and common countermeasures, such as caffeine and other stimulants, have undesirable side effects, tcVNS represents an intriguing and promising alternative that warrants future investigation. tcVNS also had a significant positive effect on the focused/distracted scale of the AFRL Mood Questionnaire. It is interesting that on its face value, this scale clearly describes the attentional capacity, one of the cognitive constructs that tVNS is hypothesized to influence, and thus it can be taken as indirect support for the underlying noradrenergic and cholinergic mechanism of tVNS.

In summary, we conducted two experiments investigating the efficacy of auricular transcutaneous vagus nerve stimulation (taVNS: Experiment 1) and cervical transcutaneous vagus nerve stimulation (tcVNS: Experiment 2) on second-language vocabulary acquisition among career linguists under training at the US Department of Defense's premier language school. The current findings showing the tcVNS efficacy among a highly selected population in a realistic multi-day protocol (i.e., 5-day) with the final test after a 24 h delay with no stimulation represents the first empirical demonstration of its kind, and it expands our understanding of tVNS' potential for enhancing cognitive skill acquisition. Furthermore, this learning enhancement was accompanied by a simultaneous fatigue-mitigation effect and a focus promoting effect. Based on the current results, combined with previous relevant findings showing tVNS's efficacy on enhancing second-language tone acquisition<sup>24,25</sup>, associative learning<sup>23</sup>, and reasoning<sup>28</sup>, as well as the extensive demonstration of its safety<sup>3</sup>, we believe the implementation of this technique could benefit language-teaching institutions, especially those that have intensive and challenging curricula like the Defense Language Institute.

## Methods

### Design

We employed an active-versus-sham between-subjects double-blind design for both experiments.

### Participants

Fifty-one volunteers from DLI's Mandarin schoolhouse and 20 volunteers from the Farsi schoolhouse were recruited to participate in Experiment 1. The volunteers were compensated for \$20 per hour for their participation. After excluding five participants for not completing the whole 5-day protocol and five pilot participants, 41 participants from the Mandarin schoolhouse (14 female,  $N_{\text{Active}} = 21$  &  $N_{\text{Sham}} = 20$ ) and 20 participants from the Farsi schoolhouse (6 female,  $N_{\text{Active}} = 10$  &  $N_{\text{Sham}} = 10$ ) were included in the final analysis on language learning. Forty-eight volunteers from DLI's Arabic school house were recruited to participate in Experiment 2. After excluding seven participants for not completing the whole 5 day protocol and five pilot participants, 36 participants from the Arabic school house (18 female,  $N_{\text{Active}} = 18$  &  $N_{\text{Sham}} = 18$ ) were included in the final analysis on language learning. For both experiments, the data from the participants who dropped out before completing all 5 days were included in the AFRL Mood Questionnaire analysis examining tVNS' effects on pre- and post-training change in subjective mood in each training day (see Results). The eligibility criteria included between the age of 18–42 (representative of the most common age groups in military population), have completed at least 3 weeks of classes at DLI and learned the letters of the target language, did not have any conditions that could compromise the safety of the electrical stimulation (e.g., narrowing of the arteries, heart disease, recent severe concussions or brain injuries, not being or planning to get pregnant), and had no neurological and psychological abnormalities (e.g., no recent hospitalization, diagnosis, psychoactive medication, or drug and alcohol treatment). Participants were quasi-randomly assigned to the active and the sham conditions, so that the two conditions roughly had equal distribution of gender and the number of weeks completed at DLI. Indeed, there was no significant difference between the participants who were assigned to be in the active versus the sham conditions on gender, The Proportion of Curriculum Completed at DLI (i.e., the number of weeks completed at DLI/the number of weeks required to complete the curriculum in the given language), or the Day 1 baseline recall score ( $p > 0.05$ : see Table S1). The study was approved by the Institutional Review Board of the Air Force

Research Laboratory, and all study procedures were performed in accordance with the relevant guidelines and regulations. The participants' consent was obtained prior to their participation.

## Material

One-hundred Mandarin-English and 100 Farsi-English word pairs were used in Experiment 1. One-hundred Arabic-English word pairs were used in Experiment 2. Participants in both experiments studied and were tested on the same 100 word pairs on all 5 days. The word-pair lists were assembled from word lists of DLI material, as available on NetProf<sup>41</sup>, and consisted mostly of nouns, but also included some adjectives, verbs, and adverbs. The words within a language were balanced in terms of English word frequency (as measured by log SUBTLEX values from the English Lexicon Project<sup>42</sup>) and concreteness<sup>43</sup>. Specifically, half of the words in the 100 words in each language were low frequency (Log Frequency<sub>low</sub>:  $M = 1.73$ ) while the other half had high frequency (Log Frequency<sub>high</sub>:  $M = 2.52$ ), and also half of the 100 words had low concreteness (Concreteness<sub>low</sub>:  $M = 2.80$ ) while the other half had high concreteness (Concreteness<sub>high</sub>:  $M = 4.57$ ). The crossing between these two variables created 25 words each of high frequency-high concreteness, high frequency-low concreteness, low frequency-high concreteness, and low frequency-low concreteness categories that did not significantly differ in English or Mandarin/Arabic word length or number of English syllables. For each task (learning, recall, recognition), stimulus lists were created with different pseudorandomized orders that were counterbalanced among participants on a given administration day. All stimulus lists contained no more than 4 consecutive items with the same frequency/concreteness category, and started and ended with unique items to prevent recency/primacy effects across days. Stimulus lists for recognition also contained no more than 4 consecutive items with the same correct response (left or right), and no two consecutive trials involved the same word as a target or a lure (i.e., the wrong answer in a forced two-choice recognition test trial). Words within the recognition task were paired such that the lure for a given target was of the same frequency/concreteness category and was consistent across days.

## Stimulation device, stimulation parameters, calibration—Experiment 1

The auricular tVNS device used for Experiment 1 was Xen by Neuvana (See Fig. 1a). Xen is a commercially available device for relaxation and wellness promotion, and it delivers stimulation through the left ear canal. Specifically, the left earbud, containing an embedded electrode with two contact points, touched the superior and inferior walls of the outer ear canal, ensuring contact with the most lateral 1 cm of these walls upon insertion. Participants in both the active and sham stimulation conditions went through a calibration procedure to determine the just-noticeable threshold before the first stimulation each day. The active stimulation was delivered for 3 min at the maximum intensity of 84 mA with a 100-level adjustment (0–100;  $M_{Intensity} = 32.64$  mA,  $SD_{Intensity} = 11.15$  mA), pulse width of 250  $\mu$ s, frequency of 25 Hz, duty cycle of on only, and pulse shape of symmetrical, square shape.

Prior to the first stimulation of the day, all participants (both active and sham) went through a calibration procedure to determine the just-noticeable threshold of the stimulation intensity through a custom iOS application. During the calibration procedure, the experimenter first increased the intensity one level at a time with approximately 2 s intervals and asked participants to raise their hands when they felt the stimulation. Next, the experimenter decreased the intensity one level at a time until the participants no longer felt the stimulation. The up-down calibration process was repeated 5 times (i.e., up-down-up-down-up), and the intensity at which a given participant felt the stimulation sensation at the 5th calibration point was taken as the just-noticeable threshold for the day. The calibration procedure was based on a previous study that demonstrated taVNS efficacy on Mandarin tone learning<sup>24</sup>.

The participants in the active stimulation condition then received the stimulation for 3 min at an intensity that was slightly below the just-noticeable threshold (i.e., just-noticeable threshold\*0.95–1 intensity level). The stimulation duration was chosen to be similar to the tcVNS duration. The participants in the sham condition sat quietly with the Xen device on their ears for 3 min, but no stimulation was delivered during this time. Following the calibration procedure, the iOS application automatically delivered an active or a sham stimulation that had been previously assigned by a remote member of the research team who were not involved in the data collection. Thus, both the participants and the experimenters were blind to the experimental condition.

## Stimulation device, stimulation parameters, calibration—Experiment 2

The cervical tVNS device used for Experiment 2 was the gammaCore<sup>®</sup> by electroCore<sup>®</sup>, Inc. (Rockaway, NJ, USA, see Fig. 1b). The gammaCore is an FDA-approved device for the acute episodes and prevention of cluster headaches and migraines. Active stimulations were delivered for 2 min each for both sides of the neck for a total of 4 min at the maximum intensity of 60 mA with a 40-level adjustment, pulse width of 1000  $\mu$ s, frequency of 25 Hz, duty cycle of 1 ms on/39 ms off, and pulse shape of full sinusoidal, symmetrical biphasic shape.

Participants in the sham condition were given a sham device that looked identical to the active device and provided a similar sound and tactile sensations but did not deliver any electrical stimulations. Participants in both the active and the sham conditions adjusted the intensity on their own. They were told to increase the intensity to where they could feel the sensation but still tolerable, thus the participants in the active condition self-calibrated ( $M_{Intensity} = 40.59$  mA,  $SD_{Intensity} = 10.42$  mA). The active and sham devices were simply labeled by a number (1 or 2), and the conditions were assigned by a remote member of the research team who were not involved in the data collection. Thus, both the participants and the experimenters were blind to the experimental condition. The self-calibration procedure was based on a previous study that demonstrated tcVNS efficacy on cognitive performance<sup>26</sup>.

## Experiment overview and procedure

The current experiments employed a 5-day protocol with baseline test in Day 1 (Monday), training with or without tVNS in Day 2 to 4 (Tuesday, Wednesday, & Thursday), and final test in Day 5 (Friday). The list and the order of tasks completed on each day are shown in Fig. 1c. The detailed descriptions of each task are below.

*Day 1*-Upon arriving at the testing room, participants completed eligibility screening, provided consent, and filled out the Mood Questionnaire before completing the baseline test (identical to the Language Learning Test-Recall above) to determine the prior knowledge on the to-be-learned list of second-language vocabulary. The Day 1 protocol lasted approximately 40 min. *Day 2 through 4*-In each of the training days, participants first completed the Mood Questionnaire, completed the calibration and received the priming stimulation for 3 min, and filled out the Sensation Questionnaire before completing the Language Learning Task. Then participants received the consolidation stimulation for 3 min, filled out the Sensation Questionnaire, completed the recall and recognition test, and filled out the Mood questionnaire again. The Day 2 through 4 protocol lasted approximately 60 min. *Day 5*-On the final testing day, participants filled out the Mood questionnaire, completed the final recall and recognition test, and were given debriefing before being dismissed. The Day 5 protocol took approximately 20 min.

## Language learning task

Performance was measured using a language learning task similar to Pavlik and Anderson<sup>44</sup>. In this task, participants memorized 100 vocabulary words in the target language that were paired with the English translation. When the word pairs were first presented, they were displayed together so that the target language word and English translation appeared at the same time. Along with the word pairs, the audio recording of a native speaker (female for Mandarin and Arabic; male for Farsi) pronouncing the target language word was played. Each of these initial presentations lasted for 3 s. Following this presentation phase, participants moved on to the feedback learning phase. During the feedback learning phase, the target language word and the audio recording appeared without the English translation. Participants were then required to recall and type out the English translation from memory within 6 s. After each feedback learning trial, participants received feedback on whether their answers were correct or incorrect along with the correct answer (Fig. 1d).

## Language learning test—recall

The recall test was identical to the feedback learning phase wherein each foreign word was presented along with its pronunciation and the participants had 6 s to type up the English translation, except there was no feedback after each trial. The baseline and recall tests were scored as correct if they were synonymous with the exact translation, or if they differed in part of speech or singular/plural.

## Language learning test—recognition

The recognition test was aimed at assessing how quickly the participants were able to access the learned information (i.e., the reaction time). The test was a two-alternative forced choice recognition test. Participants were presented with a foreign word that they learned and their pronunciation along with two English words, a correct English translation and a lure that was matched with the correct translation in terms of word frequency, concreteness, and number of letters. Participants were instructed to rest their index fingers on the “x” key (left option) and the “m” key (right option) and press the key corresponding to the correct translation as quickly as possible (Fig. 1e).

## Mood questionnaire

The 15-item mood questionnaire assessed participant's mood at the beginning and end of each testing day. Questions were answered using a 7-point semantic differential scale by which participants rated how accurately the mood descriptor fit their current mood. Examples of mood pairs include focused/distracted, excited/calm, and distressed/delighted. We were mainly interested in the fatigued/energized scale, because this scale has been sensitive to tVNS intervention under sleep deprivation<sup>26</sup>, as well as the calm/excited and distracted/focused scales because they align conceptually with tVNS' presumed noradrenergic and cholinergic enhancement of arousal and focus.

## Sensations questionnaire

The 6-item sensation questionnaire was used to ensure the safety and comfort of the participants after each exposure to the stimulation. The questionnaire asked participants to rate whether they experienced pain, metallic taste, muscle twitching, or headaches on a scale of 1–10 (1 = none, 10 = unbearable). The questionnaire also included one fill-in-the-blank question to report any other noticeable side effects, such as shortness of breath and dizziness. As no significant safety issues or side effects were recorded, we do not report the results of this questionnaire.

## Statistical approach—language learning task performance

We took a mixed-effects modeling approach to leverage the information unique to the participants and the studied items. R version 4.1.2<sup>45</sup> was used, along with *buildmer* version 2.3<sup>46</sup>, to build mixed-effects models. The maximum possible model was provided to *buildmer*, and the *bobyqa* optimizer was used. *buildmer* was used for its “order” step only, to order the terms in the model based on the amount of variance explained (as measured by a likelihood-ratio tests), and not for any backward stepwise elimination of non-significant effects, as we were interested in all of the fixed effects specified. To obtain the most maximal form of the model that converges, *buildmer* builds the model in a stepwise fashion, first adding fixed effects, and then adding random effects, until either the maximal structure is reached, or the model no longer converges.

We ran two models for each of the three task analyses (recall accuracy, recognition accuracy, and recognition reaction time): 1) learning models, which examined performance across training (sessions 2 through 4); and 2) retention models, examining performance across the last training session and the one day post-test session (sessions 4 to 5). As the recognition test was used primarily to measure reaction time, the results of the recognition accuracy are reported in Supplemental Material (Fig. S17 & Table S18).

For all model structures within the language tasks, an interaction (e.g., Session  $\times$  Treatment) implies the presence of both the interaction (Session  $\times$  Treatment) and all lower-order terms (Session, Treatment). When any baseline recall score was used as a covariate, it was mean-centered (i.e., the mean of the scores across participants was subtracted from each datapoint). Plots displaying model estimates employed the R package effects, version 4.2–2<sup>47</sup>.

### Statistical approach—mood questionnaire

Each question was analyzed separately using a mixed-effects model with session (1–5), treatment (tVNS or Sham), language (Arabic, Mandarin, or Farsi), and time (Pre or Post) and their interactions modeled as fixed categorical variables and each participant given a random intercept using the lme4 package<sup>48</sup>. Estimates and contrasts were calculated using the multcomp package<sup>49</sup>.

### Data availability

Anonymized data can be shared upon request and completing appropriate paperwork to the funding agency. Contact the corresponding author Toshiya Miyatsu (tmiyatsu@ihmc.org).

Received: 12 March 2024; Accepted: 18 July 2024

Published online: 26 July 2024

### References

1. Dubljević, V., Venero, C. & Knafo, S. What is cognitive enhancement? In *Cognitive Enhancement* (eds Dubljević, V. *et al.*) (Elsevier, 2015).
2. Brunyé, T. T. *et al.* A review of US army research contributing to cognitive enhancement in military contexts. *J. Cogn. Enhanc.* **4**, 453–468 (2020).
3. Redgrave, J. *et al.* Safety and tolerability of Transcutaneous Vagus Nerve stimulation in humans; A systematic review. *Brain Stimul.* **11**, 1225–1238 (2018).
4. Tyler, W. J. *et al.* Transdermal neuromodulation of noradrenergic activity suppresses psychophysiological and biochemical stress responses in humans. *Sci. Rep.* **5**, 13865 (2015).
5. Couto, L. B. *et al.* Descriptive and functional neuroanatomy of locus coeruleus-noradrenaline-containing neurons involvement in bradykinin-induced antinociception on principal sensory trigeminal nucleus. *J. Chem. Neuroanat.* **32**, 28–45 (2006).
6. Nichols, J. A. *et al.* Vagus nerve stimulation modulates cortical synchrony and excitability through the activation of muscarinic receptors. *Neuroscience* **189**, 207–214 (2011).
7. Dorr, A. E. & Debonnel, G. Effect of vagus nerve stimulation on serotonergic and noradrenergic transmission. *J. Pharmacol. Exp. Ther.* **318**, 890–898 (2006).
8. Yuan, H. & Silberstein, S. D. Vagus nerve and vagus nerve stimulation, a comprehensive review: Part I. *Headache* **56**, 71–78 (2016).
9. Rufener, K. S., Geyer, U., Janitzky, K., Heinze, H.-J. & Zaehle, T. Modulating auditory selective attention by non-invasive brain stimulation: Differential effects of transcutaneous vagal nerve stimulation and transcranial random noise stimulation. *Eur. J. Neurosci.* **48**, 2301–2309 (2018).
10. McIntyre, C. K., McGaugh, J. L. & Williams, C. L. Interacting brain systems modulate memory consolidation. *Neurosci. Biobehav. Rev.* **36**, 1750–1762 (2012).
11. Mather, M., Clewett, D., Sakaki, M. & Harley, C. W. Norepinephrine ignites local hotspots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *Behav. Brain Sci.* **39**, e200 (2016).
12. Jongkees, B. J., Immink, M. A., Finisguerra, A. & Colzato, L. S. Transcutaneous vagus nerve stimulation (tVNS) enhances response selection during sequential action. *Front. Psychol.* **9**, 1159 (2018).
13. Beste, C., Steenbergen, L., Sellaro, R. & Grigoriadou, S. Effects of concomitant stimulation of the GABAergic and norepinephrine system on inhibitory control—a study using transcutaneous vagus nerve stimulation. *Brain Stimul.* <https://doi.org/10.1016/j.brs.2016.07.004> (2016).
14. Keute, M., Ruhnau, P., Heinze, H.-J. & Zaehle, T. Behavioral and electrophysiological evidence for GABAergic modulation through transcutaneous vagus nerve stimulation. *Clin. Neurophysiol.* **129**, 1789–1795 (2018).
15. Sellaro, R. *et al.* Transcutaneous vagus nerve stimulation enhances post-error slowing. *J. Cogn. Neurosci.* **27**, 2126–2132 (2015).
16. Fischer, R., Ventura-Bort, C., Hamm, A. & Weymar, M. Transcutaneous vagus nerve stimulation (tVNS) enhances conflict-triggered adjustment of cognitive control. *Cogn. Affect. Behav. Neurosci.* **18**, 680–693 (2018).
17. Colzato, L. S., Ritter, S. M. & Steenbergen, L. Transcutaneous vagus nerve stimulation (tVNS) enhances divergent thinking. *Neuropsychologia* **111**, 72–76 (2018).
18. Burger, A. M. *et al.* The effects of transcutaneous vagus nerve stimulation on conditioned fear extinction in humans. *Neurobiol. Learn. Memory* **132**, 49–56 (2016).
19. Verkuil, B. *et al.* Transcutaneous vagal nerve stimulation to promote the extinction of fear. *Brain Stimul. Basic Transl. Clin. Res. Neuromodulation* **10**, 395 (2017).
20. Colzato, L. S., Sellaro, R. & Beste, C. Darwin revisited: The vagus nerve is a causal element in controlling recognition of other's emotions. *Cortex* **92**, 95–102 (2017).
21. Sellaro, R., de Gelder, B., Finisguerra, A. & Colzato, L. S. Transcutaneous vagus nerve stimulation (tVNS) enhances recognition of emotions in faces but not bodies. *Cortex* **99**, 213–223 (2018).
22. Maraver, M. J. *et al.* Transcutaneous vagus nerve stimulation modulates attentional resource deployment towards social cues. *Neuropsychologia* **143**, 107465 (2020).
23. Jacobs, H. I. L., Riphagen, J. M., Razat, C. M., Wiese, S. & Sack, A. T. Transcutaneous vagus nerve stimulation boosts associative memory in older individuals. *Neurobiol. Aging* **36**, 1860–1867 (2015).
24. Pandža, N. B., Phillips, I., Karuzis, V. P., O'Rourke, P. & Kuchinsky, S. E. Neurostimulation and pupillometry: New directions for learning and research in applied linguistics. *Annu. Rev. Appl. Linguist.* **40**, 56–77 (2020).
25. Llanos, F. *et al.* Non-invasive peripheral nerve stimulation selectively enhances speech category learning in adults. *NPJ Sci. Learn.* **5**, 12 (2020).

26. McIntire, L. K., McKinley, R. A., Goodyear, C., McIntire, J. P. & Brown, R. D. Cervical transcutaneous vagal nerve stimulation (ctVNS) improves human cognitive performance under sleep deprivation stress. *Commun. Biol.* **4**, 634 (2021).
27. McIntire, L., Goodyear, C. & McKinley, A. Peripheral Nerve Stimulation to Augment Human Analyst Performance. in *2019 IEEE Research and Applications of Photonics in Defense Conference (RAPID)* 1–3 (ieeexplore.ieee.org, 2019).
28. Klaming, R., Simmons, A. N., Spadoni, A. D. & Lerman, I. Effects of noninvasive cervical vagal nerve stimulation on cognitive performance but not brain activation in healthy adults. *Neuromodulation* **25**, 424–432 (2022).
29. Lerman, I., Klaming, R., Spadoni, A., Baker, D. G. & Simmons, A. N. Non-invasive cervical vagus nerve stimulation effects on reaction time and valence image anticipation response. *Brain Stimul.* <https://doi.org/10.1016/j.brs.2022.06.006> (2022).
30. Giraudier, M., Ventura-Bort, C. & Weymar, M. Transcutaneous vagus nerve stimulation (tVNS) improves high-confidence recognition memory but not emotional word processing. *Front. Psychol.* **11**, 1276 (2020).
31. Mertens, A. *et al.* Transcutaneous vagus nerve stimulation does not affect verbal memory performance in healthy volunteers. *Front. Psychol.* **11**, 551 (2020).
32. Mertens, A. *et al.* The potential of invasive and non-invasive vagus nerve stimulation to improve verbal memory performance in epilepsy patients. *Sci. Rep.* **12**, 1984 (2022).
33. Charles Alderson, J. *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment* (A&C Black, 2005).
34. Laufer, B. How much lexis is necessary for reading comprehension? In *Vocabulary and Applied Linguistics* (eds Arnaud, P. J. L. & Béjoint, H.) (Palgrave Macmillan UK, 1992).
35. Stæhr, L. S. Vocabulary size and the skills of listening, reading and writing. *Lang. Learn. J.* **36**, 139–152 (2008).
36. Lange, K. 64 Weeks to Fluency: How Military Linguists Learn Their Craft. *defense.gov* <https://www.defense.gov/News/Inside-DOD/Blog/Article/2061759/64-weeks-to-fluency-how-military-linguists-learn-their-craft/> (2018).
37. Wang, C. H. *An analysis of factors predicting graduation of students at Defense Language Institute Foreign Language Center* (Monterey, California. Naval Postgraduate School, 2004).
38. Petersen, C. R. & Al-Haik, A. R. The development of the defense language aptitude battery (Dlab. *Educ. Psychol. Meas.* **36**, 369–380 (1976).
39. Silva, J. M. & White, L. A. Relation of cognitive aptitudes to success in foreign language training. *Mil. Psychol.* **5**, 79–93 (1993).
40. Sharon, O., Fahoum, F. & Nir, Y. Transcutaneous vagus nerve stimulation in humans induces pupil dilation and attenuates alpha oscillations. *J. Neurosci.* **41**, 320–330 (2021).
41. Marius, T., Melot, J. & Vidaver, G. NetProf IOS pronunciation feedback demonstration. *MIT Lincoln Laboratory*. Available at: <https://www.ll.mit.edu/r-d/publications/netprof-ios-pronunciation-feedback-demonstration> (2015) (Accessed 22 July 2024).
42. Balota, D. A. *et al.* The English lexicon project. *Behav. Res. Methods* **39**, 445–459 (2007).
43. Brysbaert, M., Warriner, A. B. & Kuperman, V. Concreteness ratings for 40 thousand generally known English word lemmas. *Behav. Res. Methods* **46**, 904–911 (2014).
44. Pavlik, P. I. Jr & Anderson, J. R. Practice and forgetting effects on vocabulary memory: an activation-based model of the spacing effect. *Cogn. Sci.* **29**, 559–586 (2005).
45. R Core Team R: A language and environment for statistical computing. *R Foundation for Statistical Computing* Vienna, Austria. Version 4.1.0. <https://www.R-project.org/> (2021).
46. Voeten, C. C. buildmer: Stepwise elimination and term reordering for mixed-effects regression. *R package version 1.0-0* <https://CRAN.R-project.org/package=buildmer> (2020).
47. Fox, J. & Weisberg, S. *An R Companion to Applied Regression* (SAGE Publications, 2018).
48. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* <https://doi.org/10.18637/jss.v067.i01> (2014).
49. Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biom. J.* **50**, 346–363 (2008).

## Acknowledgements

The authors thank Lt. Ellison Deschamps, Dr. Julie Cantwell, TSgt Zachery Mcveay, MSgt Kevin Beavers, MSgt Abraham Ducatte, MSgt Christopher Baldwin, MSgt Shaun Henderson, TSgt Philip Rowe, TSgt Kelly Hudspeth, TSgt Ronald Ware, TSgt Jason Franklin, TSgt Marilyn Williams, MSgt Jediah Hanes, and Dr. Leah Graham at USAF 517th Training Group for their assistance in coordinating the data collection, Carmen Asman, Jessica Stukins-Wayman, and Joel Schooler for their assistance in preparing the language learning and testing materials as well as for data collection, Nick Pandža for advice on graphing model effects, and DARPA Program Managers and contractors Dr. Tristan McClure-Begley, Dr. Gretchen Knaack, Ms. Elizabeth Kilpatrick, Dr. Mathew Pava, Dr. Robbin Miranda, Mr. Zachary Kuhn, Ms. Kelly Waud, and Dr. Joanna Arthur for their guidance throughout the project. This document was approved by DARPA for Public Release, Distribution Unlimited. A part of this manuscript has been presented at the 2023 Annual Meeting of American Academy of Neurology, Boston, MA.

## Author contributions

TM, VO, JR, VPK, DM, PO, LM, WA, RM, PP, & TB led the design of the study. TM, VO, and JR collected the data in coordination with VPK who took care of the pseudo-random assignment of the participants remotely, and VPK and MK analyzed and visualized the data. TM, VO, JR, VPK, and MK drafted the manuscript. All authors contributed to critical revisions and approved the final manuscript draft on its accuracy and integrity of the manuscript, and approved the submitted version.

## Funding

This work was supported by DARPA contract to AFRL/IHMC (FA8650-14-D-6500), Parallax Advanced Research (N66001-17-2-4019), and University of Maryland (N66001-17-2-4009) within the DARPA Targeted Neuroplasticity Training (TNT) program. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of DARPA, AFRL or the U.S. Government.

## Competing interests

Dr. Polly O'Rourke has received compensation as a member of the scientific advisory board of Neuvana, LLC, the supplier of the taVNS device that was used in the current study. All other authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-68015-4>.

**Correspondence** and requests for materials should be addressed to T.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024