

RESEARCH ARTICLE 

Neurostimulation and Pupillometry: New Directions for Learning and Research in Applied Linguistics

Nick B. Pandza^{1,2*}, Ian Phillips¹, Valerie P. Karuzis^{1,3}, Polly O'Rourke¹
and Stefanie E. Kuchinsky^{4,1}

¹University of Maryland Applied Research Lab for Intelligence & Security, ²University of Maryland Program in Second Language Acquisition, ³University of Maryland Program in Measurement, Statistics & Evaluation and ⁴Walter Reed National Military Medical Center, Audiology and Speech Pathology Center


*Corresponding author; E-mail: npandza@umd.edu

Abstract

This paper begins by discussing new trends in the use of neurostimulation techniques in cognitive science and learning research, as well as the nascent research on their application in second language learning. To illustrate this, an experiment designed to investigate the impact of transcutaneous vagus nerve stimulation (tVNS), which is delivered via earbuds, on how learners process and learn Mandarin tones is reported. Pupillometry, which is an index of cognitive effort, is explained and illustrated as one way to assess the impact of tVNS. Participants in the study were native English speakers, naïve to tone languages, pseudorandomly assigned to active or control conditions, while balancing for nonlinguistic pitch ability and musical experience. Their performance after tVNS was assessed using a range of more traditional language outcome measures, including accuracy and reaction times from lexical recognition and recall tasks and was triangulated with pupillometry during word-learning to help understand the mechanism through which tVNS operates. Findings are discussed in light of the literatures on lexical tone learning, cognitive effort, and neurostimulation, including specific benefits for learners of tone languages. Recommendations are made for future work on the increasingly popular area of neurostimulation for the field of applied linguistics in the 40th anniversary issue of ARAL.

Introduction

Learning a second language (L2) is extremely difficult for adults, in part because it places great demands on diverse memory and attentional mechanisms (Doughty & Long, 2003) and perceptual abilities (Sebastián-Gallés & Díaz, 2012). To enhance language learning, these mechanisms and abilities have commonly been targeted with

 The experiment in this article earned an Open Materials badge for transparent practices. The data and materials are available at <https://www.iris-database.org/iris/app/home/detail?id=york%3a938003&ref=search> (Lexical Recall Task & Materials); <https://www.iris-database.org/iris/app/home/detail?id=york%3a938002&ref=search> (Lexical Recognition Matching & Materials); and <https://www.iris-database.org/iris/app/home/detail?id=york%3a938001&ref=search> (Passive Word Learning & Materials).

© The Author(s), 2020. Published by Cambridge University Press

behavioral training paradigms (e.g., Colflesh et al., 2016; Ingvalson et al., 2014), and to a lesser extent with neurostimulation techniques, including transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS; Meinzer et al., 2014; Mottaghy et al., 1999). Another neurostimulation technique, vagus nerve stimulation (VNS), has long been studied among patient populations for its therapeutic benefits and has recently gained attention for its ability to improve memory, attention, and auditory processing (Borland et al., 2018; Vonck et al., 2014), but, until now, it has not been systematically tested as an L2 intervention technique. Here we present findings from a recent study testing the pairing of noninvasive transcutaneous VNS (tVNS) with a behavioral training paradigm in which English speakers learned lexical tone contrasts in L2 Mandarin.

Phonology is widely regarded as the linguistic domain that presents the greatest difficulty for adult L2 learners (Moyer, 2014). The notorious difficulty of acquiring novel phonological features, or sound patterns, is well documented, and accurately perceiving novel suprasegmental contrasts such as lexical tone presents a persistent challenge for even advanced learners (Pelzl et al., 2019) and is the subject of a substantial literature (see Pelzl, 2019, for a recent summary). A consistent finding among the many studies that have sought to improve L2 lexical tone acquisition via behavioral training is that certain domain-general abilities and aptitudes largely determine the success of training. The recent findings that VNS can enhance several of these mechanisms make L2 lexical tone learning an ideal case for testing the efficacy of pairing non-invasive tVNS with a behavioral training paradigm.

We begin by providing a broad introduction of neurostimulation techniques for the applied linguist followed by a more detailed review of tVNS and the use of pupillometry as a means of measuring the mechanisms thought to underpin successful L2 learning. We then describe our study, indices of behavioral performance (accuracy and reaction time) that suggest tVNS can accelerate language learning, and an index of physiology (pupillometry) that provides insights into the associated underlying neural mechanisms.

Background

Neurostimulation and Language Science: A Broad Overview

Neurostimulation involves the application of stimulation (e.g., electrical, magnetic, tactile) that modulates the activity of the nervous system. There are a variety of neurostimulation techniques that target different neurocognitive mechanisms, many of which support language learning. In particular, TMS and tDCS involve placing neurostimulators (e.g., electrodes) on or above the surface of the scalp in order to affect cortical activity and have been evaluated for improving language and cognitive performance (Miniussi et al., 2008; Reis et al., 2008).

TMS uses a strong magnetic field to induce electrical current in the brain under its position on the head. A pulse of current can temporarily disrupt neural activity, and TMS has been used to simulate lesions to localize brain regions necessary for a given task, including several regions necessary for language processing (Pascual-Leone et al., 2000; Walsh & Pascual-Leone, 2003). TMS provides a high degree of accuracy of identifying where task-critical regions are in the brain (i.e., spatial localization), especially when combined with structural magnetic resonance imaging (MRI). TMS pulses can also be repeated over an extended period of time (repetitive TMS; rTMS) to facilitate

or inhibit neural activity (Miniussi *et al.*, 2008). It has been used in people with aphasia to promote better language recovery (e.g., Finocchiaro *et al.*, 2006), and healthy individuals to facilitate picture naming and other language tasks (e.g., Mottaghy *et al.*, 1999; Sakai *et al.*, 2002). Downsides of this method include a loud clicking noise associated with pulses (particularly problematic for auditory stimuli) as well as the potential for sensation across the scalp, which can be distracting.

tDCS uses electrical currents applied at low intensities (1–2 milliamps [mA]) to facilitate or inhibit cortical excitability (DaSilva *et al.*, 2015; Miniussi *et al.*, 2008). tDCS has inferior spatial localization to TMS but is able to penetrate deeper brain structures (DaSilva *et al.*, 2015) and is a silent intervention (Miniussi *et al.*, 2008). Stimulation has been found to facilitate working memory (e.g., Ohn *et al.*, 2008), long-term memory for word pairs (Marshall *et al.*, 2004), and vocabulary learning (e.g., Meinzer *et al.*, 2014). tDCS has also been found to promote language comprehension when, for example, used to inhibit cortical activity in the right Wernicke's area in subacute stroke patients (You *et al.*, 2011).

Rather than targeting specific cortical areas directly, peripheral nerve stimulation (PNS) involves stimulating the peripheral branches of a cranial nerve to modulate cortical function more broadly. Cranial nerves are sensory and motor neurons that project from the brainstem and supply nerves to (i.e., innervate) the body, especially the head and neck. Stimulation of their peripheral branches leads to changes in the activity of neuromodulatory systems, which regulate nervous system activity via neurotransmitters, such as changes in attention with the release of norepinephrine (NE) throughout many areas of the cortex. Vagus nerve stimulation (VNS) is a type of PNS that has, until recently, required surgical implantation, limiting its use to clinical populations. However, recent innovations have led to user-friendly, noninvasive transcutaneous VNS (tVNS) technologies that stimulate the vagus by passing electrical current on the skin. These technologies allow for a wider range of VNS applications with neurotypical populations, including the use of tVNS to support L2 learning.

Stimulation of the Vagus Nerve and Language Science

VNS has been investigated invasively in clinical populations since the mid-1980s for its efficacy as an antiepileptic and antidepressant (George & Aston-Jones, 2010; Vonck *et al.*, 2014). Recently, its effects on auditory processing, memory, and cognition have also been studied. VNS involves electrical stimulation applied at low levels to branches of the vagus nerve located in the ear canal or the neck that carry nerve impulses back to the brain. The vagus nerve is the tenth cranial nerve and originates from the medulla in the brainstem. Stimulation to the vagus nerve projects along nerve fibers to a brainstem nucleus called the nucleus tractus solitarii (NTS). The most well-studied mechanism underlying VNS benefits for memory and cognition (e.g., George & Aston-Jones, 2010; Vonck *et al.*, 2014) involves the NTS's innervation of the locus coeruleus (LC) brainstem nucleus, though other mechanisms have also been investigated (e.g., Manta *et al.*, 2009). The LC produces all of the neocortex's supply of the neurotransmitter norepinephrine (Samuels & Szabadi, 2008), which plays a critical role in attention modulation (Aston-Jones & Cohen, 2005).

tVNS-related benefits may be due in part to the LC-NE system's role in optimizing behavior by controlling the trade-off between scanning and focused states of attention. Peak task performance is associated with moderate tonic patterns of LC neuron firing (slow, baseline activity indicative of one's general arousal level), and high levels of phasic patterns of LC neuron firing (fast, task-evoked activity, indicative of one's attention

to a stimulus; Aston-Jones & Cohen, 2005). In addition, NE facilitates cortical long-term potentiation, a form of synaptic plasticity that may be the major cellular mechanism behind memory formation (Vonck et al., 2014).

The few studies that have investigated the effects of noninvasive tVNS on cognitive function in humans have shown improvements in learning and memory. In Jacobs et al. (2015), 30 older adults participated in a single-blind within-subjects study comparing active and sham tVNS conditions and performed a face-names association memory task. In an encoding phase, participants saw 60 neutral faces with names for five seconds each and then rested for ten minutes (consolidation phase). During a subsequent retrieval phase, participants saw old and new faces, decided if they had seen each before and, if so, selected the correct name. Active tVNS was applied to the auricular branch of the vagus nerve within the outer ear canal while sham tVNS was applied to the earlobe. Conditions were counterbalanced within participants across two sessions, and stimulation was delivered during both encoding and consolidation (17 minutes total). Accuracy significantly improved for the active over the sham condition, although no effects on reaction times (RTs) were observed. Jacobs et al. (2015) also presented data collected from a standard neuropsychological test of episodic memory before and after the faces-names association task, measuring both immediate memorization and delayed recall for 15 monosyllabic words. Performance declined over time in the sham tVNS condition but was maintained for the active tVNS condition.

Also, relevant to the present study, VNS has been observed to cause long-lasting changes in auditory processing, which may have implications for linguistic tone learning. VNS has been associated with plasticity in primary auditory cortex during pure tone learning (e.g., Borland et al., 2018; Kilgard, 2012), which has been shown in animal models to persist at least one day after treatment (Engineer et al., 2011).

There are multiple ways in which tVNS may be implemented, including priming (i.e., conditioning) and peristimulus stimulation. Priming involves applying stimulation for a specified number of seconds or minutes prior to starting a critical learning task, presumably inducing tonic shifts in arousal and thus cortical excitability that prepare the individual to be in an optimal state for learning throughout the task. For minutes to hours after even 30-second VNS pulse trains (or sequence of pulses), studies have observed an increase in the firing rates of neurons in the LC (Groves et al., 2005), activity in LC and related brain structures (Frangos et al., 2015), and concentrations of norepinephrine in the cortex and hippocampus (Follesa et al., 2007). Peristimulus stimulation involves delivering a pulse train of stimulation just prior to the presentation of critical stimuli, presumably inducing phasic changes in task-related attention to, and consolidation of, specific to-be-learned information. Work exploring peristimulus VNS has shown effectiveness at improving low-level auditory processing (Engineer et al., 2011; Kilgard, 2012). Because both tVNS approaches are hypothesized to impact LC function, we predict that either may be beneficial to language learning, and thus both will be included in this study. Similar approaches have also been studied in the context of TMS and tDCS (Klooster et al., 2016).

The topic of language learning is still new to exploration with tVNS interventions, but the cited research provides preliminary evidence that tVNS could impact language learning and tone learning more specifically. The effects of tVNS on attention and memory consolidation could promote more effective language learning, and the results of Jacobs et al. (2015) suggest it could enhance retention rates. Assessing the benefits of any new intervention on language-learning outcomes is nontrivial. Does the

intervention serve to increase phoneme or word recognition accuracy overall? Speed the overall learning rate? Reduce the mental load associated with learning an individual item, freeing up mental resources for other aspects of learning? Particularly challenging for tVNS is that, due to a paucity of established research, expected effect sizes have not been established, and thus it is possible that tVNS-induced changes in neural function might be subtle or very focused, and thus primarily observable for only a subset of possible language learning outcomes. Given the range of possibilities, our assessment of tVNS-driven language-learning benefits includes multiple outcome measures at multiple timescales (i.e., trial-level accuracy, trial-level reaction time, moment-by-moment deployment of mental resources as assessed with pupillometry). We conclude by discussing potential limitations of this design as well as suggesting ways for the field to advance.

Using Pupillometry to Assess the Impact of tVNS on Lexical Tone Training

While accuracy and reaction time measures have traditionally been used to assess language-learning outcomes, individuals who achieve the same level of proficiency in terms of phoneme discrimination accuracy or vocabulary size may have exerted vastly different degrees of effort to achieve that level of performance. Thus, there has been increasing interest in objectively measuring the mental effort that individuals deploy throughout the course of learning, above and beyond measures of accuracy. Effort has been broadly defined in terms of the mental resources that are allocated to meet the demands of a task (Pichora-Fuller *et al.*, 2016). According to models like the Framework for Understanding Effortful Listening (FUEL), the allocation of effort to a task is driven by not only the demands of a task (e.g., the difficulty of parsing an acoustic stimulus), but also individual differences in auditory and cognitive capacities, and motivation and arousal levels. Importantly, differences in cognitive effort have been observed even when performance is high or when variation in performance or reaction time is otherwise matched or controlled for across individuals (e.g., Kuchinsky *et al.*, 2013). This suggests that measuring effort in addition to performance metrics may be important for comprehensively assessing the challenges that learners face.

Pupil dilation, measured with an eyetracker, has been used as a physiological marker of changes in effort in a number of cognitive and sensory domains (e.g., Zekveld *et al.*, 2018), in part because it has been associated with the well-studied LC-NE system that modulates states of attention (Aston-Jones & Cohen, 2005) and, as described above, is the primary mechanism through which tVNS is purported to operate. LC activity is thought to influence pupil size through NE receptors in both the muscle that controls iris sphincter dilation and the Edinger-Westphal brainstem nucleus, which innervates the iris sphincter muscle (Loewenfeld, 1999). Baseline, tonic pupil diameter has been investigated as an index of general arousal while changes in phasic pupil diameter has been linked to stimulus-dependent changes in attention and effort (Gilzenrat *et al.*, 2010). As described, the relationship between task demands and effort is not one-to-one, and indeed pupil dilation has been shown to track this nonlinear relationship: Low demand is typically associated with low effort and a smaller dilation response, moderate load is associated with high effort and a relatively larger response, and high cognitive load may result in fatigue or overexertion associated with less effort and a smaller pupil size (e.g., Ohlenforst *et al.*, 2017). Because of this nonlinear relationship, the predicted impact that an intervention or training program may have on the pupil response depends on participants' performance level (Kuchinsky & Vaden, *in press*). If the task is so difficult that people give up, training may serve to improve performance at the cost of

increased effort. Training for tasks on which performance is moderately good to high may instead result in decreased effort and either better or maintained performance.

Especially relevant to the present study, changes in phasic pupil dilation (the task-evoked pupillary response; TEPR) have been linked to an event-related potential (ERP) component that has been well-studied in the field of SLA: the N400. The N400 has been used to track lexical learning and has been shown in passive word-learning tasks (in which objective performance cannot be measured) to index the formation of semantic representations (e.g., Dittinger et al., 2016). Kuipers and Thierry (2011) found smaller N400 amplitudes to be associated with larger pupil dilation (more phasic LC-NE activity) in a passive picture-word semantic association task, indicating that less effort (less phasic LC-NE activity/smaller pupil diameter) was associated with better integration of the word in the lexicon (larger N400 amplitude/more negative deflection); likewise, more effort (larger pupil dilation) was exerted on unfamiliar words (with weaker lexical representations/smaller N400 amplitudes). This was observed despite there being no behavioral response required of participants. Thus, in a lexical tone-learning study with both tVNS and pupillometry, one could expect smaller pupil dilation to reflect a more robust learning of new words. Importantly, pupillometry may allow us to observe the effect of tVNS as stimulus perception and lexical integration processes unfold, even in the absence of differences in traditional performance metrics of word learning.

Though studied extensively in the domains of auditory (Zekveld et al., 2018) and cognitive processing (van der Wel & van Steenbergen, 2018), pupillometry is a metric newly applied to the field of SLA (see Schmidtke, 2018 for a review). However, it may be especially useful for providing insights into the mechanisms by which tVNS supports second language learning due to its reliable link to the LC-NE system (Eckstein et al., 2017). Linking pupillometry and tVNS during word learning can triangulate outcomes from behavioral training with tVNS, providing both a more detailed picture of learning processes in real time and a validation of the attentional mechanisms purported to be enhanced with tVNS.

The present study takes this approach to determine how tVNS may support native speakers of English naïve to tone languages as they learn novel words distinguished by Mandarin lexical tone contrasts. Lexical tone contrasts are notoriously challenging for speakers of nontonal languages, like English, which does not distinguish between multiple meanings of a word based solely on pitch. Unlike in English, in Mandarin Chinese pitch is contrastive and changing the tone of a word changes its meaning: /ma/ with a high flat tone means ‘mother,’ /ma/ with a rising tone means ‘hemp,’ /ma/ with a dipping tone means ‘horse,’ and /ma/ with a falling tone means ‘scold.’ Mandarin has five tones: a high flat tone (tone 1), a mid-rising tone (tone 2), a low dipping tone (tone 3), a high-to-low falling tone (tone 4), and a fifth neutral tone (Wong, 1953). Many words in Mandarin comprise minimal sets with two or more of the four nonneutral tones, which makes accurate perception and categorization of a word’s tone essential for achieving proficiency.

Short-term interventions, such as sound-perception training, have helped native English speakers perceive lexical tone differences, but naïve trainees regularly fall short of consistently and accurately categorizing lexical tones, even taking as many as 18 training sessions to reach consistent performance (e.g., Bowles et al., 2016; Chandrasekaran et al., 2010; Li & DeKeyser, 2017; Wong & Perrachione, 2007). Native speakers of English who have attained high levels of proficiency in Mandarin often continue to perform below native Mandarin speakers on some measures of

tone discrimination (e.g., Pelzl *et al.*, 2019). Importantly, there is a high degree of variability among individual learning trajectories, which has spurred a number of studies examining individual differences that predict learning, such as music experience, non-linguistic tone aptitude, and executive function (e.g., Bowles *et al.*, 2016), or even differences in attended acoustic cues (e.g., Chandrasekaran *et al.*, 2010).

While Ingvalson *et al.* (2014) noted that understanding learner individual differences will allow for tailored training to provide the greatest benefit to learners, another potential avenue to enhance language learning is neurostimulation, which does not necessarily require modifying the training or input that learners receive. In the long run, a combination of insights gained from neurostimulation and aptitude-by-treatment interaction research could maximize the potential of both. Research in neurostimulation is still nascent, but selective, purposeful neurostimulation could lead to efficient and practical gains in learning.

The Present Study

The present study uses tVNS and investigates its effects on lexical tone learning across multiple outcome measures that are sensitive to changes at varying timescales (across sessions, across trials, and across milliseconds). Priming tVNS, in which tVNS is applied for a continuous period preceding some learning or performance task, and peristimulus (peristim) tVNS, in which short bursts of tVNS are time-locked to individual stimuli in some learning or performance task, were utilized and contrasted with sham tVNS in a double-blind study with tone word-learning tasks and lexical recognition and recall tests. Behavioral outcomes on recognition and recall tests were analyzed as indices of learning. Pupillometry was collected during a passive word learning task and was analyzed as an index of cognitive effort.

The research questions of the present study are:

- Does active (priming and/or peristim) tVNS versus sham tVNS improve performance on lexical recognition and/or recall?
- Is there a differential deployment of cognitive effort across individual trials during the passive word learning task that is explained by active versus sham tVNS?

The results presented here are from a larger endeavor showing positive effects of priming and peristim tVNS interventions compared to a sham stimulation condition in a double-blind study of lexical tone learning. This study represents an initial foray into studying the impact of tVNS on language learning, with a two-day training paradigm for native English speakers naïve to tone languages tasked with learning words with lexical tone. We conclude by discussing how the observed tVNS-related improvements here show promise for further neurostimulation research in the field of second language acquisition.

Methods

Participants

This study was completed by 83 participants. Participants were recruited from the University of Maryland and surrounding community, provided informed consent prior to enrolling in this study, and were paid \$20/hour overall for their time (\$10 for 1.25-hour session 1, \$10 for the three-hour session 2, \$125 for the three-hour session

3). This study was approved by the University of Maryland's Institutional Review Board and the U.S. Army Human Research Protection Program (HRPO).

Participants were native speakers of American English who reported no prior exposure to any tonal language and no significant exposure to another language before age 12.¹ All participants had self-reported normal or corrected-to-normal vision and unimpaired use of their right (dominant) hand, no hearing impairments, learning disabilities, history of neurological or psychiatric disorders, or ocular disorders that would affect eyetracking, and had not taken any psychoactive medications within two months of testing. Participants reported having none of the following conditions: being pregnant or nursing, history of cardiac or vascular disease, diabetes, epilepsy, fainting, head or face injuries, pain, or pain disorders, recent hospitalizations, or implanted electronic or metallic devices including nonremovable facial piercings.

Participants were pseudorandomly assigned to a tVNS group in order to balance groups on two variables known to influence lexical tone training outcomes: musicianship and nonlinguistic pitch discrimination ability (e.g., Bowles et al., 2016; Chandrasekaran et al., 2010; Dittinger, et al., 2016; Wong & Perrachione, 2007). The tasks used to measure nonlinguistic tone aptitude (the Pitch Contour Identification Task [PCID]) and musicianship (from an item in the Ollen Musical Sophistication Index) are described in *Materials*, and descriptives per group on these measures can be found in *Supplementary Materials*.

A total of 69 participants (46 female) ages 18–34 years ($M = 21.56$, $SD = 3.16$) were analyzed, after 14 were excluded for missing data, noncompliance throughout learning tasks, and/or incorrectly interpreting task instructions for at least one of the tasks. There were 17 participants in the priming tVNS group, 17 participants in the peristim tVNS group, and 35 participants in the sham tVNS group.²

Design

This study comprised two lexical tone-training sessions that occurred on consecutive days ($n = 4$ completed two days apart, one priming and three sham participants) and a pretraining session that occurred before the first training session. The pretraining session lasted 1.25 hours, and each training session lasted three hours. During the pretraining session, participants completed computerized tasks measuring several aspects of cognitive and musical ability, demographics, and language history. These tasks were administered to confirm participant eligibility in the study, collect measures used for group balancing, and use as other covariates outside of the scope of the present analysis.

Both training sessions included the same training tasks and tests (see *Supplementary Materials* for a list of tasks and task order for each training session) and included the collection of behavioral data (accuracy and reaction times), pupillometry, and electroencephalography (EEG). The EEG methods and results are beyond the scope of this paper and are omitted here. All tasks and tests were administered via E-Prime 2.0 (Psychology Software Tools, 2012) with a 24" LCD monitor positioned 65 cm from the participant's chinrest and all sounds were presented at 70 dB SPL (decibels, sound pressure level) through a set of Neuvana earbuds (Neuvana, LLC, Boca Raton, FL) with embedded electrodes that deliver tVNS. Pupillary data were collected with an EyeLink 1000 Plus eyetracker (SR Research, Ltd., Ontario, Canada) positioned below the monitor and behavioral responses were collected with a Chronos button box.

At the start of both training days, participants completed a sound check followed by a self-paced introduction to the concept of lexical tone that gave examples of naturally

produced monosyllables featuring Mandarin tones 1, 2, and 4 along with visual depictions of the tone contours (Figure 1). Participants then inserted earbuds modified by placing Hydrogel (Axelgaard Manufacturing Co., Ltd, Fallbrook, CA) over the electrodes on the left earbud to create a stable conductive bridge to the skin of the outer ear canal. After another sound test, participants completed the sequence of tone-word training tasks and tests in the following order: a passive paired-associates word learning task, a match/mismatch lexical recognition test, and a learned-word lexical recall test.³ The ten-minute priming task was administered three times each training day, once before every 20 minutes of task or test time. This involved watching a ten-minute silent animated video, *Inscapes*, which is designed to keep participants awake, engaged, and still during extended resting scans for MRI research (Vanderwal *et al.*, 2015). During the video, participants in the active tVNS group received continuous tVNS at 0.2 mA below an individualized, subperceptual threshold (determined via a calibration procedure described below) for the full ten minutes, while the sham priming tVNS participants received no stimulation excepting a short 7 s ramp-up as part of blinding procedures. Instead of receiving tVNS during priming, the peristim participants in the active tVNS group received 500-ms bursts of tVNS preceding each trial during all training tasks and tests except the learned word recall test due to the nature of the test. Before each task that involved tVNS, participants completed tVNS calibration and ramping, described below, and tVNS was administered during both sessions.

This study implemented a double-blind design—participants and session proctors were unaware of tVNS group assignments. A member of the research team not involved in data collection or analysis determined a participant's group assignment based on their nonlinguistic tone aptitude and self-rated musicianship scores (collected during the pretraining session) and assigned a new number to the participant for use during training and testing. The computerized training tasks and tests were programmed to reference a preloaded tVNS-group list so that entering the participant number for an experimental task triggered the correct tVNS delivery, thus allowing proctors to administer the tasks and tests without knowledge of tVNS group assignment. Before each task or test involving tVNS, all participants calibrated tVNS intensity (described below), providing their perceptual threshold.

Materials and procedures

tVNS calibration. tVNS originated from a Digitimer DS8R Constant Current Stimulator (Digitimer North America, LLC, Fort Lauderdale, FL), which was set to deliver square waves with a 50 μ s pulse and 350 μ s interphase dwell with alternating polarity and a 100% recovery phase ratio. All participants (active and sham) completed the same calibration procedure, which consisted of administering 2000 ms tVNS pulses at random 1000–3000 ms intervals that increased from two to ten mA in 0.5 mA steps until participants indicated they could feel the stimulation by pressing a button. tVNS intensity was then reduced by 1.0 mA (or to 2.0 mA, if the level was below this threshold) and then slowly ramped up in 0.1 mA steps until participants again pressed a button to indicate they felt the stimulation. At the start of the following task or test, all participants received a brief sequence of tVNS pulses that ramped up from 2.0 mA to perceptual threshold, while only the active tVNS groups received tVNS during the task or test at 0.2 mA below their perceptual threshold.

Nonlinguistic tone aptitude. Nonlinguistic pitch discrimination ability was measured with an abbreviated version of the pitch contour identification task (PCID) task used

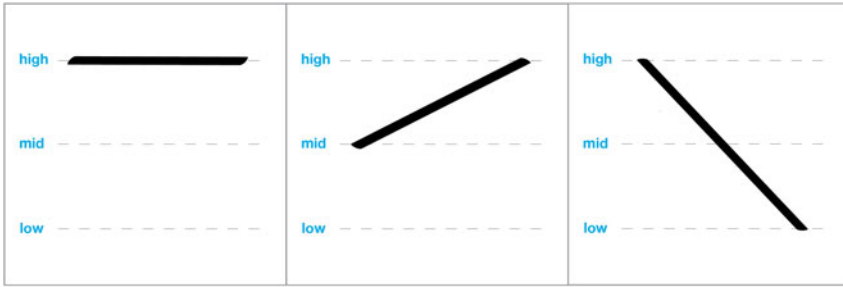


FIGURE 1. The visual aid shown to learners in the tone introduction for the high flat (tone 1), rising (tone 2), and falling (tone 4) tones.

in Bowles et al. (2016; see also Bent et al., 2006). Participants were presented with a pure tone and identified the tone as flat, rising, or falling by pressing a button. Stimuli varied by initial pitch height (200–350 Hz) for the flat tone and pitch contour difference (5–50 Hz) for rising and falling tones. Overall accuracy on this task was used for group balancing purposes.

Musicianship. Self-rated musicianship was assessed from the Ollen Musical Sophistication Index (OMSI) that consists of items pertaining to an individual's experience playing an instrument and listening to music (Ollen, 2006). Self-rated musicianship was measured with the question: Which title best describes you? Possible responses were: 1 = nonmusician, 2 = music-loving nonmusician, 3 = amateur musician, 4 = serious amateur musician, 5 = semiprofessional musician, and 6 = professional musician.

Passive paired-associates word learning task. Recordings of /ba/, /bi/, and /pi/ spoken by a male and female native Mandarin speaker with tones 1 (high flat), 2 (mid rising), and 4 (high-to-low falling) were extracted from Mandarin carrier sentences and taken with permission from a previous tone-learning study (Bowles et al., 2016). These Mandarin syllables were paired with nine English words: TRAY, OVEN, VASE, GOWN, RAFT, SOFA, MENU, LENS, and COIN. The words were all four letters long in order to control for screen luminance, and frequency (logSUBTLEX: 2.29–2.71; Brysbaert & New, 2009) and concreteness (4.61–5.00; Brysbaert et al., 2014) were controlled. Three counterbalances were used to minimize any potential idiosyncrasies of learning a particular English word with a particular Mandarin syllable or tone (“pseudoword”), pairing each English word with each segment (consonant + vowel) and tone.

Participants were instructed to learn the meanings of nine foreign language words, which would vary in sound (consonant, vowel) and tone. The importance of trying to memorize the words was stressed as they would be tested later. Every trial had a 750 ms baseline period in which an English word was presented in the middle of the screen with the visual contour of its tone above (a flat, rising, or falling line). Then, there was a 1750 ms period in which a Mandarin syllable was presented auditorily as the written English word and contour remained on the screen. Participants were not required to respond to the stimuli, but pupillometry was collected during this task. Each English word was presented a total of ten times (five times each per male and female speaker) for a total of 90 trials. Stimulus lists were pseudorandomized to avoid blocking by tone, segment type, or speaker.

Lexical recognition test. In this test, each Mandarin pseudoword was presented 24 times, split by speaker, for a total of 216 trials. These trials were split into two 108-trial blocks with one break in between. Trials were pseudorandomized within block to avoid repetition of the same Mandarin pseudoword in consecutive trials. Half the trials were matches and half the trials were mismatches via tone only, not segment. No feedback was given.

There was a 750 ms baseline period in which a visual English word appeared in the center of the screen. The tonal contour never appeared with the word in this test, unlike in the passive paired-associates word learning task. A subsequent 2,000 ms period began with an aurally presented Mandarin syllable while the English word remained on the screen. During this time, participants indicated whether or not the pairing was a correct translation by pressing a button. Finally, there was a 1,000 ms period in which a four-character visual mask of 'XXXX' replaced the written word on the screen.

Lexical recall test. This test consisted of nine trials. For each trial, participants were presented one of the nine Mandarin pseudowords produced by the female speaker and were given unlimited time to listen to each item as many times as they liked. Participants were instructed to type the correct English translation of the Mandarin syllable on a keyboard. There was no word bank. Responses to this test were reviewed and the only hedge cases in determining accuracy were a limited number of instances where the participant had pluralized the English word (e.g., typed 'COINS' instead of 'COIN'). These responses were accepted as correct. There were no other synonyms or misspellings.

Post-experiment awareness questionnaire. This questionnaire gathered information about participants' awareness of their stimulation condition. Participants were asked to indicate to which condition they believed they were assigned when they received stimulation (answer options described priming, peristimulus, sham, sham without ramping, and other), their confidence in their answer, and whether the stimulation helped them perform better on study tasks (rating from 1–9).

Results

Group balancing and double-blinding procedures

tVNS-group means for the ID measures collected during the pretraining session were compared using two-tailed t-tests. These results indicate that group balancing procedures were successful in balancing active and sham tVNS groups on PCID and self-rated musicianship ($ps > .10$). Participant responses to post-experiment questionnaire items probing their awareness of their assigned tVNS group were analyzed and the results indicate that the tVNS calibration procedures were successful in blinding participants to their tVNS group ($ps > .10$).

tVNS improves tonal language learning performance

Descriptives for behavioral tasks and tests are available in the *Supplementary Materials*. To answer the first research question, whether tVNS improves performance on lexical recognition or recall, the priming, peristim, and sham tVNS groups were compared with binomial logistic mixed-effects models (MEMs) for accuracy (recognition and recall) and a linear MEM for RT (recognition only); participants were not given a

response deadline for the recall task). The MEM for recognition RT analyzed only correct trials, with spurious responses excluded (responses <60 ms, <1% of the data). All MEMs were run with the lme4 (Bates et al., 2015) package in R (R Core Team, 2019), and model testing to arrive at the models of best fit for random and fixed effects (including covariates for musicianship and PCID) was performed with the buildmer package (Voeten, 2019), using the Satterthwaite approximation for degrees of freedom for linear MEM p -values. Final models of best fit are reported below, and all summary tables are reported in *Supplementary Materials*.

Accuracy results for the lexical recognition test are plotted in Figure 2 for the effects of interest. There was a positive effect of peristim tVNS over sham on mismatch trials (est. = 0.581, $SE = 0.192$, $p = .002$), although not for match trials (est. = 0.267, $SE = 0.195$, $p = .171$ when releveling model baseline to match trials). There was no effect of priming tVNS over sham for mismatch trials (est. = 0.184, $SE = 0.194$, $p = .343$), but there was a significant effect of priming tVNS over sham on match trials (est. = 0.403, $SE = 0.198$, $p = .041$, when releveling model baseline to match trials), and, when releveling, a marginal difference from priming to peristim for mismatch trials (est. = 0.397, $SE = 0.229$, $p = .083$) and no difference for match trials. All of the effect sizes were consistent from training day one to two, as everyone improved at the same (logarithmic) rate. Musicianship (est. = 0.190, $SE = 0.088$, $p = .032$) and PCID (est. = 0.349, $SE = 0.090$, $p < .001$) were both significant.

RT results for the lexical recognition test related to our research questions are plotted in Figure 3. No differences in priming or peristim tVNS over sham were observed at day one or day two ($ps > .10$), but there was a significant interaction for priming tVNS and training day indicating that the priming tVNS sped up from day one to day two significantly more than the sham group (est. = -0.107, $SE = 0.036$, $p = .004$). For the peristim group, there was no difference from either priming or sham ($ps > .10$). The observed effect for priming tVNS was consistent across match and mismatch trials. The covariates of musicianship and PCID were not significant ($ps > .10$).

Accuracy results for the lexical recall test are plotted in Figure 4. Priming tVNS was associated with better recall performance over sham (est. = 1.179, $SE = 0.500$, $p = .018$). Accuracy was marginally better with peristim versus sham tVNS (est. = 0.912, $SE = 0.510$, $p = .072$), and no difference was observed between priming and peristim ($p > .10$, when releveling the model baseline). These group effects were consistent from training day one to day two as everyone improved from day one to two at the same (logarithmic) rate. The covariate of musicianship was not significant ($p > .10$) but PCID was significant (est. = 0.676, $SE = .211$, $p = .001$).

Pupillometry reveals differences in effort by tVNS group during learning

After finding differential benefits of stimulation on behavioral performance, the impact of stimulation on pupillometry was examined to try to tease apart the mechanistic differences between groups during learning to answer our second research question. In pupillometry analyses, the entire pupil response over the course of a trial is evaluated, as group differences for the pupil response can be seen in three ways after controlling for variation in both participants and trials: (1) whether one group has an earlier peak in the pupil response than the other (quicker deployment of effort); (2) whether one group has a more peaked (effortful) response than another group; and (3) whether one group's response drops off more quickly over the time course of a word learning trial (less sustained effort over time). Pupillometry data from the passive word learning

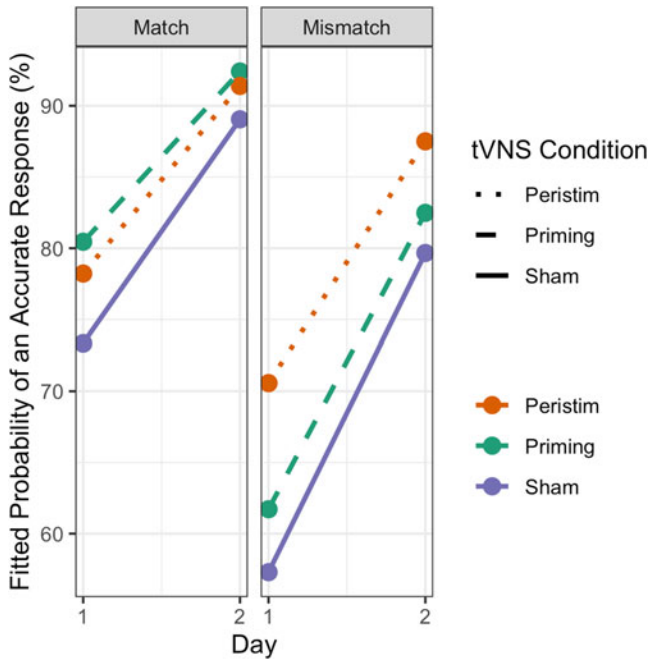


FIGURE 2. Modeled effects of tVNS on lexical recognition test accuracy.

task were analyzed with generalized additive mixed modeling (GAMM), a processor-intensive analysis in which each time point from every trial from every participant can be analyzed.

Data were preprocessed in three steps: (1) data were downsampled to 50 Hz as recommended for GAMMs by van Rij *et al.* (2019), since, above that, the added detail does not significantly change the results but does significantly increase the time it takes a computer to calculate the model; (2) the 750 ms baseline period before each trial was subtracted from the trial for each person so that any observed differences between the groups are due to tVNS impacting the encoding of information in a specific trial rather than conflating it with any effects of tVNS on general arousal; (3) any trials for which more than 33% of the data were missing (due to blinks, saccades, looking off-screen, etc.) were rejected from analysis. This last step resulted in fewer than 15 usable trials for two participants on one training day, who were then excluded from this analysis, resulting in 35 sham, 17 peristim, and 15 priming tVNS participants across two training days for this analysis. Importantly, the number of removed trials was not associated with any one particular condition and thus should not impact the pattern of observed results. GAMMs were implemented with the *mgcv* package (Wood, 2017) following previous recommendations in applying GAMMs to pupillometry data and language science data (Sósokuthy, 2017; van Rij *et al.*, 2019), including an autoregressive model and random smooths for participants and items. GAMMs provide a more appropriate analysis than growth curve analysis for pupillometry data in particular, because they can deal with the issue of autocorrelation (that the position of the pupil at one time point is correlated with its position at the next time point, increasing Type I error if not

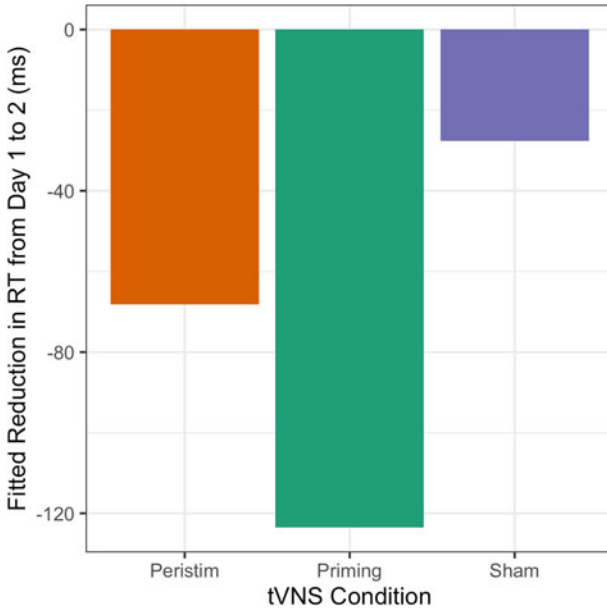


FIGURE 3. Modeled effects of tVNS on reductions in lexical recognition test RT.

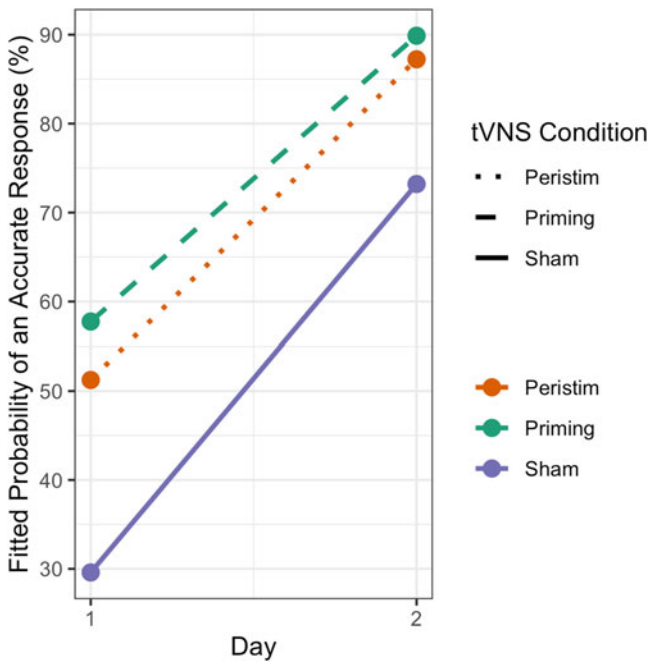


FIGURE 4. Modeled effects of tVNS on lexical recall test accuracy.

controlled for) as well as the fact that pupil size may be influenced by the pupil's position relative to the eyetracker camera (van Rij *et al.*, 2019). Through a computationally intensive algorithm, GAMMs objectively find the number of inflections for the pupil response curve that can support the data. Unlike more traditional analyses, the summary table usually has no utility for inferring statistical significance due to the complexity of the smooth terms for each curve, but significance can be determined by first testing a model with and without the parametric (traditional predictors that one would include in MEMs or regression models) and smooth terms (predictors specific to GAMMs that allow the penalized estimation of a nonlinear relationship) of interest followed by visual inspection of difference curves (subtracting one curve from another) to inspect whether a difference curve and its confidence interval are distinct from zero (Sóskuthy, 2017; van Rij *et al.*, 2019).

Model testing determined that including parametric and smooth terms for tVNS condition, training day, and their interaction significantly improved model fit ($\chi^2(14) = 131.14, p < .001$). The final model's summary table with all terms is presented in *Supplementary Materials* and the estimated TEPRs (task-evoked pupil responses) are depicted in [Figure 5](#), which shows the time course of a trial on the x-axis (0 to 1750 ms), pupil size on the y-axis, and different descriptive TEPR curves for each of the three conditions on each day. The most robust group differences were observed in changes from training day one to two and can be observed in [Figure 6](#), which shows the difference curves from day one to day two for each of the three groups. Statistical significance is supported by inspecting where each of the three curves is different from zero. For the peristim and sham groups, the TEPR increased during the early part of a trial from day one to day two, reflecting an earlier deployment of effort and a less sustained response on the second day of training compared to the first. The timecourse of the TEPR did not change overall from day one to day two, but there was a larger response on day two, reflecting more engagement of cognitive effort from day one to two. Comparing groups, inspecting where each of the three curves separated from each other, a clear effect for the peristim group emerged such that, from day one to day two, there was less sustained effort during a learning trial compared to both priming tVNS and sham tVNS. Sham tVNS also appeared to have a significantly less sustained response than priming tVNS, but this effect was much less robust. Peristim tVNS also showed evidence of more effort being recruited earlier during a learning trial than priming.

Discussion

Two different tVNS interventions were observed to elicit performance improvements over a sham control in a double-blind study of learning novel pseudowords featuring lexical tone. For lexical recognition, peristim tVNS—neurostimulation time-locked to stimulus presentation—showed an advantage over sham in lexical recognition accuracy of mismatch trials by about 5–10% but not RT, while priming tVNS—neurostimulation ten minutes continuously prior to a task or test—showed an RT advantage of about 100 ms and a significant effect on accuracy over sham by about 3–6% on match trials. For lexical recall (which included a small number of test items), priming had a positive effect on accuracy over sham by about 15–30% (depending on day) and peristim showed a marginal effect on accuracy over sham by about 15–20% and was not significantly different from priming.

While these results indicate learning advantages for tVNS recipients, the effect sizes are not easily compared to those found in other tone word learning studies, since these

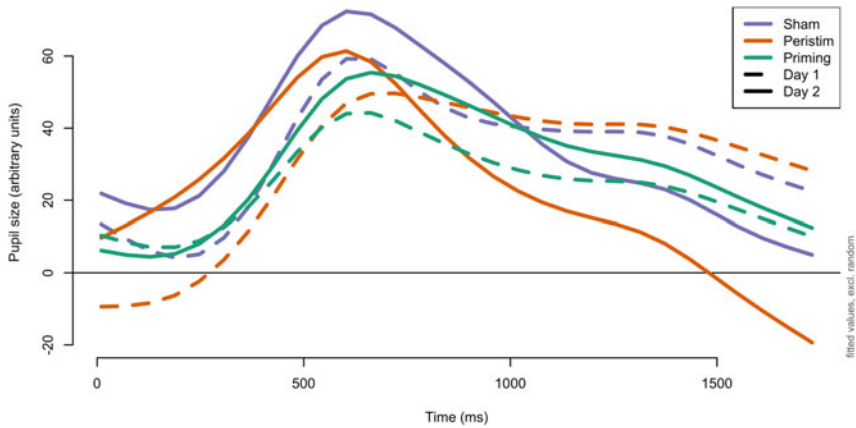


FIGURE 5. Descriptive model curves from the passive word learning pupillometry GAMM.

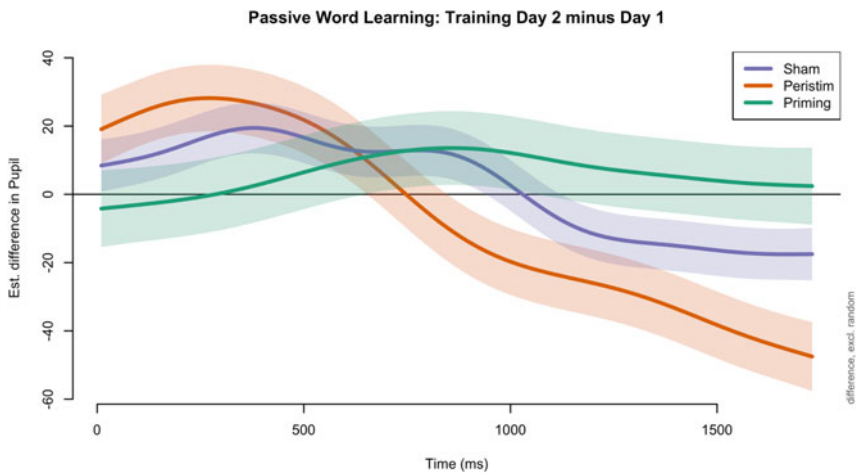


FIGURE 6. GAMM difference curves showing significant differences within and between groups for the change in TEPR from training day one to training day two.

typically involve more training sessions and lexical items and focus on interactions between learner characteristics and training design (e.g., Perrachione et al., 2011) or stimulus manipulations (e.g., Antoniou & Wong, 2016). The passive word learning task and recognition test used here were based on Dittinger et al.'s (2016) study of professional musicians and nonmusicians learning nine Thai words that included tonal contrasts. Although analyses of N400s elicited during these tasks suggested more efficient word learning for musicians, there were no corresponding differences in accuracy or RT scores. However, musicians were more accurate on a semantic-relatedness task involving the same words, which does provide one important benchmark: word learning advantages attributed to a single session of tVNS in the present study emerge as early in training as the learning advantages attributed to years of musical training in Dittinger et al. (2016).

There are a few reasons that priming and peristim tVNS may have had differential impacts on accuracy and RT for recognition and recall tests. One is that the recall test was short; given that there are only nine words to learn and nine items on the recall test, “large” improvements only reflect a relatively small difference in the number of items correctly recalled. Another explanation may be that tVNS may facilitate different aspects of word learning: number of items that can be learned (via peristim) and speed with which learned items can be accessed (via priming). Additionally, the observation that stimulation type differently impacted changes in match vs. mismatch accuracy suggest an avenue for future research regarding the relative benefits of priming and peristim on attention to relevant information versus inhibition of distracting information.

Given that tVNS is hypothesized to affect production of NE and effort allocation, we used pupillometry to investigate the allocation of cognitive effort for each group during the passive word learning task of the experiment. Supporting our expectation that a smaller TEPR reflects a better integration of newly learned words (particularly when the task is not so difficult that people give up), we observed a significantly faster drop-off in the TEPR for peristim than for sham from training day one to day two. This suggests that the peristim group required less sustained effort for a given learning trial than sham while memorizing words and later performed better on accuracy for those words on both days. For priming, the results are less clear, as there were not robust differences from sham. Overall within the priming group from day one to day two, there was a slightly more peaked TEPR. This weaker effect may have arisen, if, as previous work suggests, priming tVNS impacts task-evoked effort indirectly compared to peristim. Priming may alter the tonic firing pattern of the LC, which in turn allows for participants to be in the optimal arousal state to exert mental effort (Aston-Jones & Cohen, 2005). The task in this paper as designed is not appropriate for a tonic analysis, but future planned experiments aligning aspects of this work with invasive animal models of auditory learning with tVNS will be used to explore this possibility.

Even after only one day of training, tVNS had positive effects on lexical tone word learning as measured on lexical recognition and recall tests. These advantages persisted into the second day of training. Moreover, the two types of tVNS interventions, priming and peristim, resulted in different types of benefits for learning. Despite peristim having less total stimulation duration, it resulted in as good or better accuracy than priming. This suggests that the total amount of stimulation may be less relevant than the nature of the stimulation (i.e., time-locked to a given stimulus or primed continuously before a task or test).

Looking Forward

The current study’s results suggest a promising future for tVNS as fast, effective language learning support. Improvements were observed almost immediately and coincided largely with pupillary changes that reflected a predicted influence of tVNS on the LC-NE system. The promise of tVNS lies not only in these results but in the fact that tVNS can be induced safely with consumer-grade equipment straight out of the box.⁴ While implementing a tVNS study with a consumer-grade device is relatively straightforward, the stimulation parameters (e.g., shape of the stimulation waveform, duration of stimulation) that lead to optimal learning benefits are largely unknown. Given the recency of research applying neurostimulation to second language learning and the diversity of neurophysiological mechanisms targeted by different techniques, future research should not only focus on developing effective protocols but also

weigh the relative efficacy of these methods for different aspects of language learning while considering their practical limitations.

While our double-blind study design strengthens causal inferences about tVNS effects on tonal word learning, future work should establish tVNS parameters and protocols that optimize efficacy for different learning tasks and learner characteristics. The present results showed tVNS benefits between groups balanced on pitch aptitude and musician-ship. However, future work would benefit by targeting a larger distribution of musician experience to determine how tVNS efficacy interacts with these variables known to predict success. Likewise, the tVNS protocols used in this study likely improved learning via optimizing effort, yet learning could be further supported by targeting other relevant processes with tVNS. Long-term VNS has been shown to induce changes in tone processing in the primary auditory cortex as a function of stimulation rate (Buell et al., 2018) and intensity (Borland et al., 2016). Establishing whether long-term tVNS leads to similar changes could further support acquiring L2 phonology, improved speech segmentation, and more accurate comprehension in noisy environments. Spending less time on these low-level features of language comprehension in the classroom means more instructional time can be spent on learning higher-level linguistic features, like pragmatics, which are equally vital to developing advanced language proficiency.

As a first attempt to apply tVNS to language learning, this study administered priming and peristimulus tVNS during training and test phases. Future work will systematically compare the effects of different tVNS targets: stimulating only during training to potentially facilitate memory encoding, stimulating only during testing to potentially support recall of learned items, and stimulating following only accurate trials to selectively support encoding. Studies in our laboratory, currently underway, are taking the first step in this direction, but much more work lies ahead to fully understand how tVNS can best support language learning and to maximize its benefits to language learners at different stages of learning and with different learning strengths and backgrounds. At this juncture, however, these results and the previous literature suggest that neurostimulation, when paired with behavioral language learning approaches may provide a much needed boost for adult language learners to overcome the inherent difficulty in learning a second language.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0267190520000069>.

Acknowledgments. We thank Eric Pelzl for his assistance with stimuli selection and design, Matthew Turner and Sara McConnell for their assistance with data collection, Meredith Hughes, Jason Struck, and Alison Tseng for their assistance with blinding the data and assigning groups, Jarrett Lee for assistance programming tVNS, and Henk Haarmann and Greg Colflesh for contributions to earlier portions of the overall project.

This material is based upon work supported by the Naval Information Warfare Center and Defense Advanced Research Projects Agency under Cooperative Agreement No. N66001-17-2-4009. The identification of specific products or scientific instrumentation is considered an integral part of the scientific endeavor and does not constitute endorsement or implied endorsement on the part of the author, DoD, or any component agency. The views expressed in this article are those of the author and do not reflect the official policy of the Department of Army/Navy/Air Force, Department of Defense, or U.S. Government.

Notes

1 Participants who reported experience learning a language before age 12 were admitted into the study if the experience before age 12 was limited to class in a nonimmersion school.

2 There are more participants in the sham group because of the larger study's design. Active and sham priming tVNS participant data were collected simultaneously, and active and sham peristim tVNS participant data were collected simultaneously. Because the different sham group participants differed only in that the sham group for the priming condition was exposed to a ten-minute video and the peristim sham group had an extra 500 ms at the beginning of each trial (during which the peristim group received their stimulation), no differences between sham groups in terms of language learning trajectory were expected and they were combined into one sham group for the present analysis.

3 Participants in all three conditions also completed several tasks and tests that address research questions beyond those considered in this paper, including phonological tone categorization training and testing tasks, and affect and anxiety surveys. The tone categorization training task involved the same Mandarin stimuli used in the word learning tasks.

4 While this study used a research-grade stimulator for precise control of the stimulating waveform and required some training, the earbuds are commercially available as is a handheld stimulation device that is easy to use with more limited waveform options, which was not used here (costing under \$500 total).

References

- Antoniou, M., & Wong, P. C. (2016). Varying irrelevant phonetic features hinders learning of the feature being trained. *The Journal of the Acoustical Society of America*, *139*(1), 271–278.
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.
- Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 97–103.
- Borland, M. S., Engineer, C. T., Vrana, W. A., Moreno, N. A., Engineer, N. D., Vanneste, S., Sharma, P., Pantalia, M. C., Lane, M. C., Rennaker, R. L., & Kilgard, M. P. (2018). The interval between VNS-tone pairings determines the extent of cortical map plasticity. *Neuroscience*, *369*, 76–86.
- Borland, M. S., Vrana, W. A., Moreno, N. A., Fogarty, E. A., Buell, E. P., Sharma, P., & Kilgard, M. P. (2016). Cortical map plasticity as a function of vagus nerve stimulation intensity. *Brain Stimulation*, *9*(1), 117–123.
- Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, *66*(4), 774–808.
- Brysbart, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990.
- Brysbart, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911.
- Buell, E. P., Loerwald, K. W., Engineer, C. T., Borland, M. S., Buell, J. M., Kelly, C. A., Khan, I. I., Hays, S. A., & Kilgard, M. P. (2018). Cortical map plasticity as a function of vagus nerve stimulation rate. *Brain Stimulation*, *11*(6), 1218–1224.
- Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America*, *128*(1), 456–465.
- Colflesh, G., Karuzis, V., & O'Rourke, P. (2016). Effects of working memory training on L2 proficiency and working memory capacity. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 289–294.
- DaSilva, A. F., Truong, D. Q., DosSantos, M. F., Toback, R. L., Datta, A., & Bikson, M. (2015). State-of-art neuroanatomical target analysis of high-definition and conventional tDCS montages used for migraine and pain control. *Frontiers in Neuroanatomy*, *9*, 1–89.
- Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional music training and novel word learning: from faster semantic encoding to longer-lasting word representations. *Journal of Cognitive Neuroscience*, *28*(10), 1584–1602.
- Doughty, C. J., & Long, M. H. (2003). Optimal psycholinguistic environments for distance foreign language learning. *Language Learning & Technology*, *7*(3), 50–80.

- Eckstein, M. K., Guerra-Carrillo, B., Singley, A. T. M., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, 25, 69–91.
- Engineer, N. D., Riley, J. R., Seale, J. D., Vrana, W. A., Shetake, J. A., Sudanagunta, S. P., Borland, M. S., & Kilgard, M. P. (2011). Reversing pathological neural activity using targeted plasticity. *Nature*, 470(7332), 1–15.
- Finocchiaro, C., Maimone, M., Brighina, F., Piccoli, T., Giglia, G., & Fierro, B. (2006). A case study of primary progressive aphasia: improvement on verbs after rTMS treatment. *Neurocase*, 12(6), 317–321.
- Follesa, P., Biggio, F., Gorini, G., Caria, S., Talani, G., Dazzi, L., Puligheddu, M., Marrosu, F., & Biggio, G. (2007). Vagus nerve stimulation increases norepinephrine concentration and the gene expression of BDNF and bFGF in the rat brain. *Brain Research*, 1179, 28–34.
- Frangos, E., Ellrich, J., & Komisaruk, B. R. (2015). Non-invasive access to the vagus nerve central projections via electrical stimulation of the external ear: fMRI evidence in humans. *Brain Stimulation*, 8(3), 624–636.
- George, M. S., & Aston-Jones, G. (2010). Noninvasive techniques for probing neurocircuitry and treating illness: vagus nerve stimulation (VNS), transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS). *Neuropsychopharmacology*, 35(1), 301.
- Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, 10(2), 252–269.
- Groves, D. A., Bowman, E. M., & Brown, V. J. (2005). Recordings from the rat locus coeruleus during acute vagal nerve stimulation in the anaesthetised rat. *Neuroscience Letters*, 379(3), 174–179.
- Ingvalson, E. M., Ettliger, M., & Wong, P. C. (2014). Bilingual speech perception and learning: A review of recent trends. *International Journal of Bilingualism*, 18(1), 35–47.
- Jacobs, H. I., Riphagen, J. M., Razat, C. M., Wiese, S., & Sack, A. T. (2015). Transcutaneous vagus nerve stimulation boosts associative memory in older individuals. *Neurobiology of Aging*, 36(5), 1860–1867.
- Kilgard, M. P. (2012). Harnessing plasticity to understand learning and treat disease. *Trends in Neurosciences*, 35(12), 715–722.
- Klooster, D. C., de Louw, A. J., Aldenkamp, A. P., Besseling, R. M. H., Mestrom, R. M. C., Carrette, S., Zinger, S., Bergmans, J. W. M., Mess, W. H., Vonck, K., Carrette, E., Breuer, L. E. M., Bernas, A., Tijhuis, A. G., & Boon, P. (2016). Technical aspects of neurostimulation: Focus on equipment, electric field modeling, and stimulation protocols. *Neuroscience & Biobehavioral Reviews*, 65, 113–141.
- Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Jr., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, 50(1), 23–34.
- Kuchinsky, S. E. & Vaden, K. I., Jr. (in press). Aging, hearing loss, and effort: Imaging studies of the aging brain. In K. S. Helfer, E. L. Bartlett, A. N. Popper, & R. R. Fay (Eds.), *The Aging Auditory System*. Springer.
- Kuipers, J. R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience*, 5, 61.
- Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition*, 39(4), 593–620.
- Loewenfeld, I. E. (1999). Otto Lowenstein: Neurologic and ophthalmologic testing methods during his lifetime. *Documenta Ophthalmologica*, 98(1), 3–20.
- Manta, S., Dong, J., Debonnel, G., & Blier, P. (2009). Enhancement of the function of rat serotonin and norepinephrine neurons by sustained vagus nerve stimulation. *Journal of Psychiatry & Neuroscience*, 34(4), 272–280.
- Marshall, L., Mölle, M., Hallschmid, M., & Born, J. (2004). Transcranial direct current stimulation during sleep improves declarative memory. *Journal of Neuroscience*, 24(44), 9985–9992.
- Meinzer, M., Jähnigen, S., Copland, D. A., Darkow, R., Grittner, U., Avirame, K., Rodriguez, A. D., Lindenberg, R., & Flöel, A. (2014). Transcranial direct current stimulation over multiple days improves learning and maintenance of a novel vocabulary. *Cortex*, 50, 137–147.
- Miniussi, C., Cappa, S. F., Cohen, L. G., Flöel, A., Fregni, F., Nitsche, M. A., Oliveri, M., Pascual-Leone, A., Paulus, W., Priori, A., & Walsh, V. (2008). Efficacy of repetitive transcranial magnetic stimulation/transcranial direct current stimulation in cognitive neurorehabilitation. *Brain Stimulation*, 1, 326–336.

- Mottaghy, F. M., Hungs, M., Brüggmann, M., Sparing, R., Boroojerdi, B., Foltys, H., Huber, W., & Töpper, R. (1999). Facilitation of picture naming after repetitive transcranial magnetic stimulation. *Neurology*, *53*(8), 1806.
- Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics*, *35*(4), 418–440.
- Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., & Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research*, *351*, 68–79. <https://doi.org/10.1016/j.heares.2017.05.012>
- Ohn, S. H., Park, C. I., Yoo, W. K., Ko, M. H., Choi, K. P., Kim, G. M., Lee, Y. T., & Kim, Y. H. (2008). Time-dependent effect of transcranial direct current stimulation on the enhancement of working memory. *Neuroreport*, *19*(1), 43–47.
- Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings* [Unpublished doctoral dissertation]. Ohio State University, Ohio.
- Pascual-Leone, A., Walsh, V., & Rothwell, J. (2000). Transcranial magnetic stimulation in cognitive neuroscience—virtual lesion, chronometry, and functional connectivity. *Current Opinion in Neurobiology*, *10*, 232–237.
- Pelzl, E. (2019). What makes second language perception of Mandarin tones hard?: A non-technical review of evidence from psycholinguistic research. *Chinese as a Second Language*, *54*(1), 51–78.
- Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2019). Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition*, *41*(1), 59–86.
- Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., & Tremblay, K. L. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37*, 5S–27S.
- Psychology Software Tools (2012). E-Prime (Version 2.0) [Computer software]. Pittsburgh, PA.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>.
- Reis, J., Robertson, E. M., Krakauer, J. W., Rothwell, J., Marshall, L., Gerloff, C., Wassermann, E. M., Pascual-Leone, A., Hummel, F., Celnik, P. A., Classen, J., Flöel, A., Ziemann, U., Paulus, W., Siebner, H. R., Born, J., & Cohen, L. G. (2008). Consensus: Can transcranial direct current stimulation and transcranial magnetic stimulation enhance motor learning and memory formation? *Brain Stimulation*, *1*, 363–369.
- Sakai, K. L., Noguchi, Y., Takeuchi, T., & Watanabe, E. (2002). Selective priming of syntactic processing by event-related transcranial magnetic stimulation of Broca's area. *Neuron*, *35*(6), 1177–1182.
- Samuels, E. R., & Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: Its roles in the regulation of arousal and autonomic function part I: principles of functional organisation. *Current Neuropharmacology*, *6*(3), 235–253.
- Schmidtke, J. (2018). Pupilometry in linguistic research: An introduction and review for second language researchers. *Studies in Second Language Acquisition*, *40*(3), 529–549.
- Sebastián-Gallés, N., & Díaz, B. (2012). First and second language speech perception: Graded learning. *Language Learning*, *62*, 131–147.
- Sóskuthy, M. (2017). Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. *arXiv preprint arXiv:1703.05339*.
- van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015.
- van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing*, *23*, 1–23.
- Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C., & Castellanos, F. X. (2015). *Inscapes*: A movie paradigm to improve compliance in functional magnetic resonance imaging. *NeuroImage*, *122*, 222–232.
- Voeten, C. C. (2019). *buildmer*: Stepwise elimination and term reordering for mixed-effects regression. *R package version 1.3*.

- Vonck, K., Raedt, R., Naulaerts, J., De Vogelaere, F., Thiery, E., Van Roost, D., Aldenkamp, B., Miatton, M., & Boon, P. (2014). Vagus nerve stimulation... 25 years later! What do we know about the effects on cognition? *Neuroscience & Biobehavioral Reviews*, 45, 63–71.
- Walsh, V., & Pascual-Leone, A. (2003). *Transcranial magnetic stimulation: A neurochronometrics of mind*. MIT Press.
- Wong, H. (1953). Outline of the Mandarin phonemic system. *Word*, 9(3), 268–276. DOI: 10.1080/00437956.1953.11659474
- Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, 28(4), 565–585.
- Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd edition). CRC Press.
- You, D. S., Kim, D.-Y., Chun, M. H., Jung, S. E., & Park, S. J. (2011). Cathodal transcranial direct current stimulation of the right Wernicke's area improves comprehension in subacute stroke patients. *Brain & Language*, 199, 1–5.
- Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends in Hearing*, 22, 1–25.

Cite this article: Pandža NB, Phillips I, Karuzis VP, O'Rourke P, Kuchinsky SE (2020). Neurostimulation and Pupillometry: New Directions for Learning and Research in Applied Linguistics. *Annual Review of Applied Linguistics* 40, 56–77. <https://doi.org/10.1017/S0267190520000069>