JOURNAL OF UNDERGRADUATE RESEARCH

Statistical Language Learning and Test-Retest Reliability: A Pilot Study

Ivette De Aguiar and Edith Kaan
University of Florida

Faculty mentor: Edith Kaan, Department of Linguistics

Abstract

Applying transcutaneous stimulation to the auricular branch of the vagus nerve (tVNS) has been shown to enhance associative learning in humans. The goal of a new project in our lab is to investigate the effect of tVNS on procedural learning, specifically implicit statistical language learning. The aim of the experiment reported in this paper was to determine which statistical language learning paradigms would be appropriate to use with tVNS. Since we would be looking at within-subject changes between two sessions (one session with, one session without stimulation), we tested the test-retest reliability of two statistical learning paradigms: one testing word segmentation and adjacent dependencies, the other testing non-adjacent dependencies. We also tested the correlation between an explicit phonological memory task and the implicit statistical learning tasks to determine whether phonological memory was involved in the statistical learning tasks. Our results showed a high test-retest reliability for the word segmentation task. However, the task dealing with non-adjacent dependencies had low test-retest reliability, meaning it would not be appropriate for future studies incorporating tVNS. There was a high correlation between the phonological memory task and both statistical learning tasks, indicating implicit statistical learning may recruit phonological memory.

Keywords: statistical learning, implicit learning, phonological memory, vagus nerve stimulation

Introduction

Transcutaneous stimulation of the auricular branch of the vagus nerve (tVNS) has been shown to enhance associative learning in humans (Jacobs, Riphagen, Razat, Wiese, & Sack, 2015). The overall goal of the current project is to investigate whether tVNS can improve procedural learning, in particular, implicit statistical learning of language-like materials. Implicit learning requires that a subject has learned information without awareness of what has been learned. This is a subconscious form of learning. Using tVNS for our purposes will require two sessions per subject to properly monitor within-subject change between a session with and a session without stimulation. As a first step towards this goal, we implemented a pilot study where we tested the test-retest reliability of two statistical language learning paradigms without tVNS in order to discover which paradigms have the most consistent results between sessions and, therefore, which will most effectively show within-subject changes due to tVNS. We did this by presenting the same tasks in two different sessions separated by 3 weeks. Additionally, we tested the correlation between a phonological memory task based on Nittrouer, Lowenstein, Wucinich, & Moberly (2016) and statistical language learning. In both statistical learning paradigms, participants listened to an artificial language. The first paradigm was based on Gómez (2002) and tested the learning of non-adjacent dependencies between two morphemes; the second paradigm was based on Saffran, Newport, & Aslin (1996) and Isbilen, Stewart, Kidd, & Christiansen (2017), and investigated the sensitivity to transitional frequencies between syllables.

The Saffran et al. (1996) task investigated word segmentation and adjacent dependencies. Specifically, it covered the transitional properties found within and between words in a language, which helps humans decipher when one word ends, and another word begins in fluent speech. When learning a language, there are numerous statistical regularities that occur within fluent speech, with the highest transitional regularities occurring when two sounds follow each other in the same word. Transitional probabilities crossing a word boundary will be low and unpredictable. An example in the English language is *funny#story*. The transitional probability from *fun* to *ny* is greater than from *ny* to *sto*. In this task, participants listened to a pseudo language during the exposure phase. After exposure, a forced choice task was used which required the participants to choose between a pseudo word which did not follow the transitional probability rules of the pseudo language and a pseudo word found in the pseudo language. Participants were asked to determine which word was previously heard in the pseudo language used.

The Gómez (2002) task tested the implicit learning of non-adjacent dependencies. Such dependencies occur in natural language and are essential for expressing relations between elements in a sentence. An example in English is: *John walks*, or *John sleeps*. In this case, there are three components of the sentence which we can call *X, Y,* and *Z,* where *X* is *John*, *Y* is *walk* or *sleep* and *Z* is *s* indicating the proper conjugation of the verb “to walk” for third person singular. In this example, *X* and *Z* are non-adjacent dependencies: the intervening verb stem Y can change, but Z (-*s*) is dependent on the subject X. In this task, participants listened to a pseudo language during the exposure phase. The testing phase required participants to engage in a familiarity yes/no task. A pseudo phrase was presented to participants one at a time, and they had to determine whether the phrase presented sounded like it belonged in the pseudo language they previously heard.

Finally, we used a phonological memory task based on Nittrouer et al. (2016). We tested whether phonological working memory is correlated with implicit statistical learning. In the statistical language learning task, participants were told to listen carefully to a pseudo-language. Short-term memory may therefore be important for remembering and accessing contingencies in the language. We hypothesized that statistical learning and phonological working memory would be related, where the better the phonological working memory, the better the statistical learning ability will be.

If both statistical learning paradigms have a high test-retest reliability, we will apply these tasks in a tVNS study. Learning the effects of vagus nerve stimulation paired with implicit statistical language learning tasks may improve our understanding of how individuals learn languages. This may have applications, for example, in military settings where individuals must become proficient with a language in a short amount of time.

Methods

Participants

A total of 33 native English speakers from the University of Florida (10 men, 23 women, 18-33 years, mean age 18.75 years) were recruited. They were paid or given course credit to participate. All reported having been raised monolingually, having no history of reading or hearing problems, and having no history of neurological problems. Four additional participants failed to return for the second session, hence we did not include their data in the analyses involving session 2. The study protocol was approved by the University of Florida Institutional Review Board.

Materials and Procedure

The experiment consisted of two sessions, which occurred 3 weeks apart. Each session consisted of a phonological memory task based on Nittrouer et al. (2016), the statistical learning tasks based on Saffran et al. (1996), and finally the statistical learning task based on Gómez (2002). The order of the three tasks was the same for each participant and session.

phonological memory. The phonological memory task had 16 experimental items: 8 non-rhyming (*ball, rake, ham, teen, seed, dog, pack, coat*) and 8 rhyming (*vat, cat, hat, pat, rat, bat, mat, gnat*) 1-syllable English words (Figure 1). Participants listened to the 8 non-rhyming and the 8 rhyming words spoken in a random order and were then required to choose the corresponding pictures representing the items in the order that they had heard the words. Participants did this for 10 different sequences of non-rhyming words, and 10 sequences of rhyming words. The number of errors in the order of recall was recorded.

word segmentation and adjacent dependencies. The first statistical learning task given to participants was based on the task used in Saffron et al. (1996) and Isbilen et al. (2017), and deals with word segmentation and adjacent dependencies (Figure 1). The input language consisted of 18 syllables (*ta, ti, to, ma, mo, pa, po, ri, lo, la, lu, ki, ka, ga, du, di, bu, bi*), combined into 6 trisyllabic words: *latibi, lomari, modipa, tagalu, topoka, kibudu.* In the exposure phase, participants listened to an 11-minute continuous auditory stream, in which these words were randomly concatenated without pauses between words (*modipalatibilomarikibudutagalu…).* Participants were asked to monitor the stream for occasional repetitions of syllables by pushing the space bar when a repetition was heard. For example, in the following stream, (*modi**papa**latibiloma**riri**ki…*), the participant would be expected to push the space bar when "papa", "riri", or any other syllable repetition was heard. A two-alternative forced choice task (2AFC) was administered in the test phase following the auditory exposure phase. For the 2AFC task, six pseudo words were randomly generated *dikabi, kigala, tatori, polubu, mamoti, lopadu.* These pseudo words did not conform to the transitional probability rules used in the words in the exposure phase. Instead, they were intended to act as foil words, words not found in the exposure phase, and to be paired with one of the 6 trisyllabic words found in the exposure phase. A foil word and a word found in the artificial language presented in the exposure phase were presented auditorily. The participants were asked to choose the word they thought was in the exposure phase by pushing a corresponding key on the keyboard for 36 trials. For instance, if a participant heard "*kibudu*" and "*kigala*", they would be expected to recognize "*kibudu*" as correct since it was found in the exposure phase.

non-adjacent dependencies. The second statistical learning task was based on Gómez (2002), and deals with non-adjacent dependencies on the syntactic level (Figure 1). Two artificial languages were created, each one built by combining pseudo words to create 3-word strings, where the first and third word belonged together (*e.g. pel X rud, vat X jic,* where X was varied). The first language followed the pattern *aXb* and *cXd*, while the second language took the pattern of *aXd* and c*Xb*. The X spot could potentially be filled randomly by 24 items (e.g. *wadim, kicey,* *puser*, ...). The exposure phase consisted of a two-minute stream with 250 ms pauses between words and 750 ms pauses between strings to help distinguish between strings and the words within them. A familiarity yes/no task was administered in the test phase consisting of 12 items, half following the transitional probability rules of the language they were exposed to and half violating those rules. The participant was asked to indicate if a string was a correct sequence in the previously exposed language (*pel wadim rud* -correct string, *pel kicey jic*-incorrect string, etc.). Participants heard one of the two languages in the first session, and the other in the second session. Half of the participants heard the first language in the first session, and the second language in the second session; in half of the participants the order of presentation was reversed.

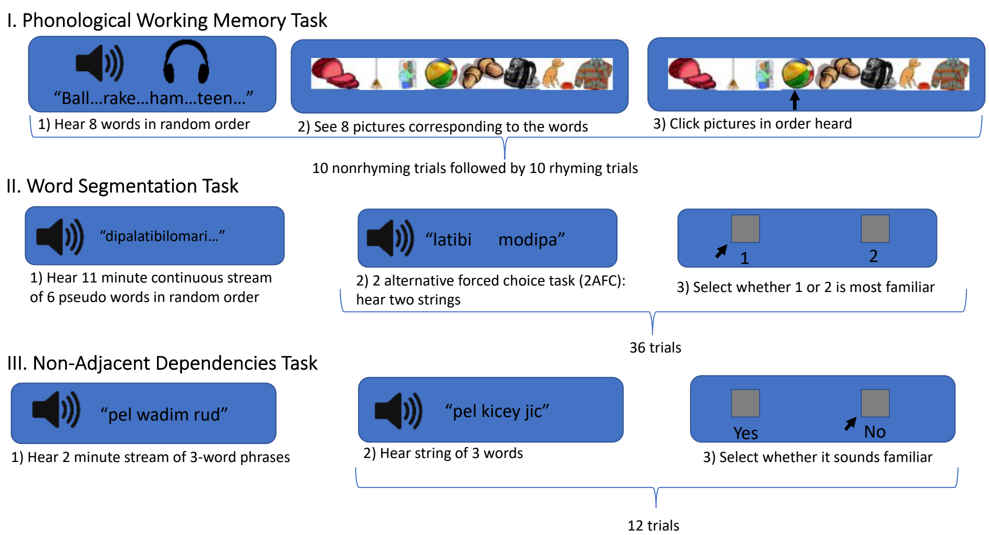

Figure 1. Overview of the three tasks used in the study.

statistical methods. Statistical analyses were conducted in R (R Core Team, 2018). We first obtained the mean performance accuracy of each participant per task per session (number of errors in the rhyming and non-rhyming portion of the phonological memory task, proportion of correct responses in the 2AFC task in the word segmentation task, and the proportion of correct responses in the test phase of the non-adjacent dependencies task). Next, to find to what extent performance was in the statistical learning tasks was above chance in session 1, a one-sample t-test (one-tailed) was used for each measure. The same was done for session 2. The correlation between both sessions was calculated using Pearson's product-moment correlation coefficient. If there was a high and significant correlation between sessions 1 and 2, then there would be high test-retest reliability. However, if there was low correlation between the two sessions, then there would be low test-retest reliability. Finally we computed correlations between the number of errors in the phonological memory task and the proportion of accurate responses on each of the statistical learning tasks.

Results

Word Segmentation with Adjacent Dependencies

In the first statistical learning task, the word segmentation task, participants performed above chance in both session 1 and 2 in the 2AFC task (session 1 $t_{(32)}$=8.78, p<0.0001, mean accuracy=0.67±0.11; session 2: $t_{(28)}$=12.20, p<0.0001, mean accuracy=0.76±0.12). Performance on the 2AFC in session 1 correlated with that in session 2, meaning that the test has a good test-retest reliability (r=0.67, p<0.0001) (Figure 2).

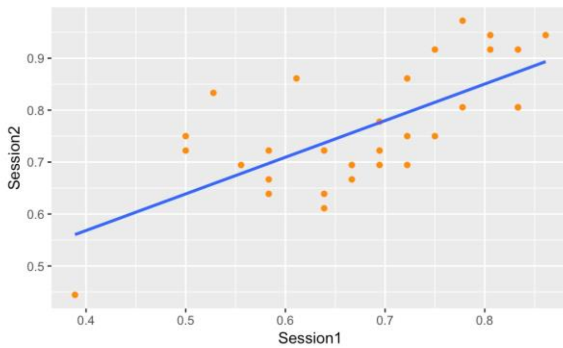

Figure 2. Correlation between session 1 and 2 of the 2AFC task. The orange dots represent proportion accuracy for session 1 and 2 for each participant. The blue line represents the statistical correlation trend and demonstrates how the participants' performances follow this trend.

Non-Adjacent Dependencies

In the statistical learning task testing non-adjacent dependencies, performance was above chance in both session 1 and 2 (session 1: $t_{(32)}$=2.80, p<0.01, 0.61±0.23, session 2: $t_{(27)}$=5.55, p<0.0001, 0.76±0.26). However, there was no correlation between sessions 1 and 2 (r=0.15, p=0.44). Furthermore, the distribution of the data was bimodal: participants either performed at chance or at ceiling (Figure 3). The weak test-retest reliability may be largely due to the bimodal distribution.

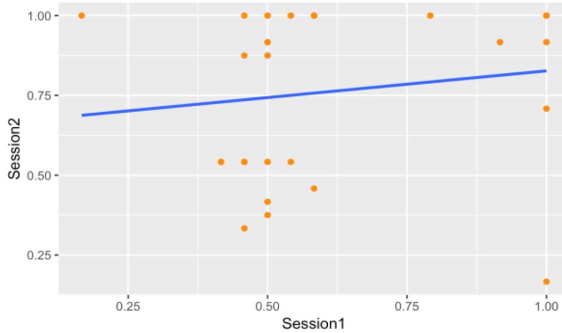

Figure 3. Correlation between session 1 and 2 of the familiarity yes/no task. The orange dots represent the proportion accuracy in session 1 and 2 for each participant. Note the bimodal distribution of the data. Test-retest reliability was low.

Phonological Working Memory Task

We next tested the correlations between the Nittrouer et al. phonological working memory task and the statistical language learning tasks. A significant correlation was found between session 1 of the rhyming version and the word segmentation statistical learning task accuracy ($r = -0.44$, $p < 0.01$) (Figure 4). A correlation was also found between session 1 of the rhyming task and the non-adjacent statistical learning task ($r = -0.37$, $p < 0.05$) (Figure 5). The pattern seen here shows that the better the participant's phonological working memory (the fewer errors they made), the better they performed on the statistical learning tasks.

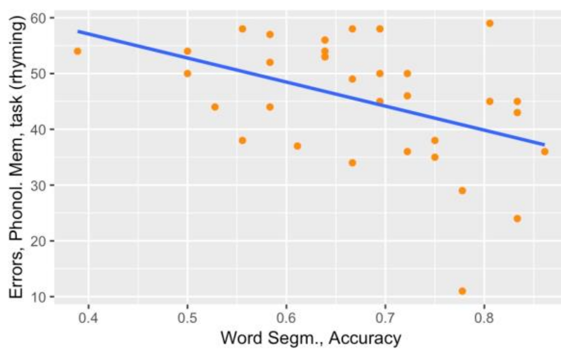

Figure 4. Correlation between session 1 of the memory task and session 1 of the word segmentation task. The x-axis shows the proportion accuracy in the word segmentation task. The y-axis represents the number of errors in the working memory task. The orange dots represent the performance in session 1 on both tasks for each participant. The blue line represents the statistical correlation.

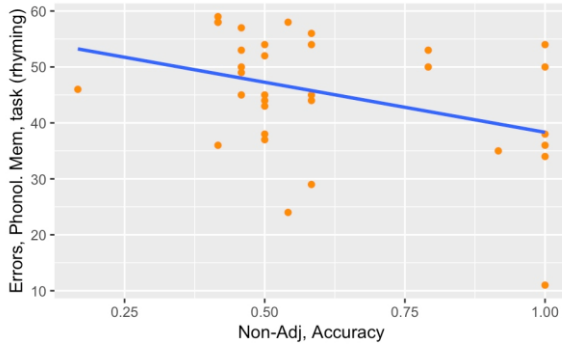

Figure 5. Correlation between session 1 of the memory task and session 1 of the non-adjacent dependencies task. The x-axis shows the proportion accuracy in the non-adjacent dependencies task. The y-axis represents the errors in the working memory task. The orange dots represent the performance in session 1 on both tasks for each participant. The blue line represents the correlation.

Discussion

The purpose of the study was to measure the test-retest reliability between two sessions of two statistical learning paradigms. Furthermore, we tested the effects of explicit phonological working memory on implicit statistical learning. We hypothesized that statistical learning and phonological working memory would be related. That is, the better the phonological working memory the better the statistical learning.

We found that the word segmentation task had a good test-retest reliability, making it an appropriate task to use in an upcoming tVNS study. However, the non-adjacent dependencies task had no correlation between session 1 and 2; its test-retest reliability was low; hence, this task would not be appropriate for an upcoming tVNS study. Additionally, as we predicted, good performance on either statistical learning task in session 1 was associated with good performance on the rhyming version of the phonological memory task, suggesting that auditory statistical learning recruits phonological working memory. We believe tasks implemented in the testing phases of both statistical learning tasks used explicit working memory. Working memory is an explicit learning function, meaning that it is done consciously and is considered part of short-term memory (Nittrouer et al., 2016). When participants heard the pseudo-language stimuli in the exposure phase, it makes sense that participants tried to remember as much of what they heard as possible since they knew there would be a testing phase following exposure. By doing this, although they were implicitly learning, they consciously tried to remember language information meaning implicit and explicit learning was occurring simultaneously.

Future research in this area could explore why performance in the non-adjacent dependency task had no correlation between the two sessions, specifically looking to see if changing the task employed during exposure to a more involved one would change the results. A better task may be one that tests learning during exposure (Misyak, Christiansen, & Tomblin, 2010). The task employed in the Misyak et al. study used the same pseudo language utilized in the Gomez et al. task and presented participants visually with syllables in a grid like fashion on a computer screen. The participants were then asked to click on the pseudo words they heard while simultaneously listening to the pseudo language. The response time was measured; faster response times would mean that the participants were beginning to anticipate certain patterns within the language structure.

The current pilot study suggested that the word segmentation task has good test-retest reliability, and therefore is suitable to use in a future two-session study testing whether tVNS improves implicit language learning.

References

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436. doi:10.1111/1467-9280.00476

Isbilen, E.S., McCauley, S.M., Kidd, E. & Christiansen, M.H. (2017). Testing statistical learning implicitly: A novel chunk-based measure of statistical learning. In G. Gunzelmann, A. Howes, T. Tenbrink, & E.J. Davelaar (Eds.), Proceedings of the 39th Annual Conference of the Cognitive Science Society (pp. 564-569). Austin, TX: Cognitive Science Society.

Jacobs, H. I. L., 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. doi:10.1016/j.neurobiolaging.2015.02.023

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, *2*(1), 138-153. doi:10.1006/jmla.1996.0032

Nittrouer, S., Lowenstein, J. H., Wucinich, T., & Moberly, A. C. (2016). Verbal working memory in older adults: The roles of phonological capacities and processing speed. *Journal of Speech, Language, and Hearing Research, 59*(6), 1520-1532. doi:10.1044/2016_jslhr-h-15-0404

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Languag*e, *35*(4), 606-621.