Neural Underpinnings of Phonotactic Rule Learning

Enes Avcu and Ryan Rhodes
University of Delaware

Aim and The Finding. Artificial grammar learning (AGL) studies have clearly shown that learners can extract adjacent and non-adjacent dependencies with relatively short training at the behavioral level. However, less is known about how these patterns are encoded at the neurophysiological level. The aim of the current study is to observe the neurophysiological correlates of implicit learning of a non-adjacent phonotactic pattern using EEG. We find that without a priori explicit learning the brain can distinguish words which follow the pattern from words that violate it. Interestingly, only trials with correct behavioral responses show a significant difference in brain response to grammatical and ungrammatical words, indicating a correlation between neural response and behavioral response within subject.

Methods. We ran an artificial grammar learning experiment testing a simple phonotactic pattern, namely a sibilant harmony rule, a non-adjacent harmony pattern attested in Navajo.
Participants. A total of 24 monolingual English speakers participated in an AGL experiment in which ERPs were recorded.
Stimuli. All training and test stimuli consisted of two syllables of the form of CV.CV, with sibilants ([s, ʃ]) as the first and second consonants. All words were either "harmonic" (both sibilants identical) or "disharmonic" (mixed [s] and [ʃ]). The vowels in the alphabet of the language were [a, ɛ, ɔ, i, u]. The duration of each phoneme was strictly controlled at 100ms, making each word 400ms long.
Procedure. The procedure for the experiment consisted of two phases: a training phase and a testing phase. During the training phase, participants listened to words that conform to sibilant harmony and were instructed to repeat each word orally after they heard it. The training lasted approximately 15 minutes. In the testing phase, the task for the participant was to press a button in response to each stimulus to categorize the stimulus as grammatical or ungrammatical. Participants were tested in an auditory oddball paradigm (a deviant stimulus infrequently appearing among repeated occurrences of a standard stimulus). Participants discriminated between harmonic and disharmonic words, with harmonic words appearing in 80% of trials and disharmonic words appearing in 20% of trials[1], with 300 total trials. Each word was presented for 400ms followed by a blank inter-trial interval of 1100 to 1500ms. The duration of the test session was approximately 10 minutes.
Prediction. If the two EEG waveforms (the brain response to harmonic and disharmonic words, respectively) are significantly different from each other, we can unambiguously conclude that the brain can distinguish harmonic words from disharmonic words, and that this distinction must have been generated as a result of categorization. We can then say that without explicit learning the brain is making a distinction between the harmonic and disharmonic words.
Data Recording and Analysis. In the test phase, button presses made by participants to deviant stimuli (disharmonic words) were recorded. *Hits* (when an ungrammatical word was present and the participant detected it and reported hearing it) and *false alarms* (when a grammatical word was present but the participant thought s/he heard an ungrammatical word and reported it) were counted. Sensitivity Index (*d'*), (Signal Detection Theory (SDT); Green and Swets, 1966; Heeger, 2007), the detection difficulty of a target stimulus, was then derived from the hit and false alarm rates. Finally, *d'* was compared to zero – which would indicate no learning – by using a one sample t-test. The EEG was recorded from 128 carbon fiber core/silver-coated electrodes in an elastic electrode net (Geodesic Hydrocel 128). The

[1] There should NOT be any frequency effect here because this experiment tests learnability; if participants do not know the rule, they will not know whether harmonic words are frequent or not.

signals were re-referenced offline to the average of the left and right mastoids (Luck 2005). We coded each trial as correct or incorrect following the button presses. ERPs were calculated from the fronto central region with a time window of 400 to 800ms (amplitudes were measured as the mean voltage in the given measurement window).

Results. Behavioral results showed that deviants were detected with a mean sensitivity of 0.557 (*SD*=0.815), and significantly different from zero, $t(23)=3.34$, $p=0.003$, $d=0.684$, $1-\beta=0.894$. Mean accuracy (percent correct) was 0.664 (SD=0.129) and significantly greater than chance (which would be 0.5), $t(23)=6.21$, $p<0.001$, $d=1.269$, $1-\beta=0.999$.

Electrophysiological results showed that trials with correct behavioral responses showed a significant difference in brain response between grammatical and ungrammatical words, $t(23)=3.047$, $p=0.006$, $d=0.622$, $1-\beta= 0.830$. Whereas trials with incorrect behavioral responses do not show a significant difference in the brain response (p value >.05) All stimuli elicited a clear auditory evoked potential (AEP). See Figure 1 below for a comparison of these waveforms.

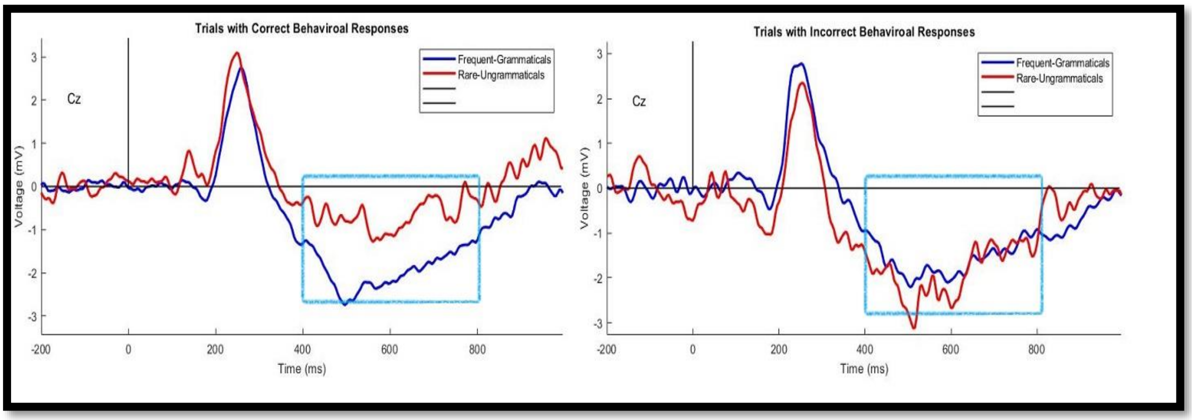

Fig. 1. Trials with correct behavioral responses (left panel) and incorrect responses (right panel); Waveforms show mean voltage in the given time window and Cz is chosen for illustration purposes. Two waveforms started to separate from each other at 400 ms which is the offset of the word (left panel).

Discussion and Conclusion. The behavioral results demonstrated that participants learned the non-adjacent phonotactic pattern (d' was above zero, and accuracy was greater than chance). We replicated the earlier findings of whether an attested and computationally learnable pattern (Heinz, 2010) is inside the hypothesis space of humans' phonological pattern detectors, by testing whether violations of such patterns trigger behavioral responses to novel stimuli after a training session. The novel finding of this study is the correlation of neurophysiological response and the behavioral response. While all subjects learned the pattern, they only displayed a difference in brain response on the trials when they correctly categorized the word. This within-subject correlation suggests that the observed brain response is a necessary precursor to accurate discrimination and is not simply an effect of frequency (responding to rare vs frequent stimuli).

References.

Green, D. M., and Swets, J. A. (1966). Signal Detection Theory and Psychophysics. New York: Wiley.

Heeger, D. (2007). Signal detection theory. Retrieved November 30, 2015, from http://www.cns.nyu.edu/~david/handouts/sdt/sdt.html.

Heinz, J. (2010). Learning long-distance phonotactics. *Linguistic Inquiry* 41:623–661.

Luck, S. J. (2005). An introduction to the event-related potential technique. Cambridge, MA: MIT Press.