Neural Underpinnings of Phonotactic Rule Learning

Enes Avcu, Ryan Rhodes, Chao Han and Arild Hestvik

University of Delaware

Key Words: Phonotactics, Learnability, EEG, Oddball Paradigm

Background. Artificial grammar learning (AGL) studies have been widely used for testing the learnability of phonological patterns. It has been clearly shown that learners can extract adjacent and non-adjacent dependencies with relatively short training at the behavioral level. Less is known about how these patterns are encoded at the neurophysiological level. For example, Domahs et al. (2009) and Moore-Cantwell et al. (forthcoming) reported a higher amplitude Late Positive Component (LPC), a response that has been reported in response to syntactic violations, to novel words that violated a learned phonotactic constraint than novel words that satisfied it.

Aim and The Finding. The aim of the current study is to observe the neurophysiological correlates of implicit learning of a non-adjacent phonotactic pattern (a sibilant harmony rule which is an attested long-distance harmonic pattern). We find that without a priori explicit learning, the brain can still distinguish words which follow the pattern from words that violate it. Furthermore, only trials for which participants made the correct grammaticality judgement 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.
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. 24 monolingual English speakers participated in the experiment and the procedure 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, participants were instructed to press a button in response to each stimulus to categorize the stimulus as grammatical (harmonic) or ungrammatical (disharmonic). Participants were tested in an auditory oddball paradigm, and required to discriminate 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 400 ms followed by a blank inter-trial interval of 1100 to 1500 ms. The duration of the test session was approximately 35 minutes[2].
Prediction. If the two EEG waveforms (the brain response to harmonic and disharmonic words, respectively) are significantly different from each other, we can conclude that the brain can distinguish harmonic words from disharmonic words, and that this distinction must have been generated as a result of learning.
Data Recording and Analysis. In the test phase, button presses made by participants to each stimulus (either frequent or rare) 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 they heard an ungrammatical word and reported it) were counted. Learning was measured via the Sensitivity Index (*d'*), a measure of the participants' sensitivity to ungrammatical words. Each participant's *d'* score was derived from the hit and false alarm rates (Signal Detection Theory (SDT); Green and Swets, 1966; Heeger, 2007). The EEG was recorded from 128 carbon fiber core/silver-coated electrodes in an elastic electrode net (Geodesic Hydrocel 128). The signals were re-referenced offline to the average of the left and right mastoids (Luck 2005). We

[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.
[2] The test session originally had four blocks; in the first and third block, harmonic words were frequent and this was reversed in the second and fourth block. However, results of the second, third and fourth block are NOT reported here because those blocks were included as part of a pilot study.

coded each trial as correct or incorrect following the button presses. Given that we did not have a priori hypotheses about the specific ERP, analyses were conducted based on visual inspection of the mean amplitudes for both conditions. This resulted in analyses of the fronto central region with a time window of 400 to 800 ms.

Results. Behavioral results showed that deviants were detected with a mean sensitivity of 0.557 (*SD*=0.815), a score significantly different from zero, $t(23)=3.34$, $p=0.003$, $d=0.684$, $1\text{-}\beta=0.894$. We also analyzed percent correct, and mean accuracy was 0.664 (SD=0.129) and significantly greater than chance, $t(23)=6.21$, $p<0.001$, $d=1.269$, $1\text{-}\beta=0.999$. Electrophysiological results showed that trials with correct behavioral responses showed a significant difference in brain response between grammatical and ungrammatical words, $t(23)=3.047$, $p=0.006$, $d=0.622$, $1\text{-}\beta= 0.830$. Trials with incorrect behavioral responses did not show a significant difference in brain response (p value >.05) All stimuli elicited a clear auditory evoked potential (AEP). See Figure 1 below for a comparison of these waveforms, measured at fronto-central region.

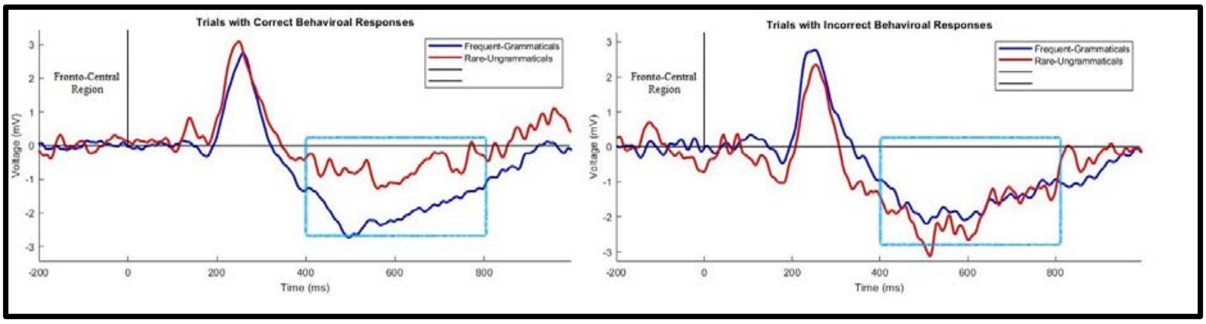

Fig. 1. Trials with correct behavioral responses (left panel) and incorrect responses (right panel); Waveforms show mean voltage in the given time window at the fronto-central region. Two waveforms started to separate from each other at 400 ms which is the offset of the word (left panel).

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 pattern was learned implicitly, with no semantic information (cf. Moore-Cantwell et al. (forthcoming)) or other contextual information to aid learning. The main finding of this study is that after only 15 minutes of implicit learning task, the brain can produce a response for grammaticality. We observed a neurophysiological response to a novel and an artificial rule which indicates how quickly the brain gets used to predicting a linguistic pattern. It is also important to note the overall (without single trial analysis) 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, namely the brain response dictates behavior.

References.

Domahs, Ulrike, W. Kehrein, J. Kraus, R. Wiese, and M. Schlesewsky. (2009). Event- related potentials reflecting the processing of phonological constraint violations. Language and Speech 52: 415–435.

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.

Moore-Cantwell, Claire, J. Pater, R. Staubs, B. Zobel and L. Sanders. (forthcoming). Event-related potential evidence of abstract phonological learning in the laboratory.