Version of March 2017. Manuscript under review.
Comments are welcome; please also contact us before citing.
(clairemoorecantwell@gmail.com, pater@umass.edu,
lsanders@psych.umass.edu)

Event-related potential evidence of abstract phonological learning in the laboratory·

Claire Moore-Cantwell, Joe Pater, Robert Staubs, Benjamin Zobel and Lisa Sanders

University of Massachusetts Amherst

Abstract. The experimental study of artificial language learning has become a widely used means of investigating the predictions of theories of phonology and of learning. Although much is now known about the generalizations that learners make from various kinds of data, relatively little is known about how those generalizations are cognitively encoded. This paper presents an ERP study of brain responses to violations of lab-learned phonotactics. Novel words that violated a learned phonotactic constraint elicited a larger Late Positive Component (LPC) than novel words that satisfied it. Because an LPC has also been found in the study of naturalistically learned phonotactics, this new result provides support for the ecological validity of lab learning of phonology. Furthermore, because the LPC is associated with violations of abstract "rules", such as syntactic violations and violations of musical expectations, this result provides evidence that generalizations acquired in the lab can be encoded at an abstract level.

1. Introduction

Artificial language learning, the experimental study of the learning of constructed linguistic patterns, has become a widely employed technique amongst phonologists (see Moreton and Pater (2012a; 2012b) for a review). It has been used to address a number of research questions in theoretical phonology, including the extent to which speakers have knowledge of phonological universals that are not instantiated in their native language (e.g. Pycha et al. 2003; Wilson 2003; Carpenter 2010), the relationship between phonotactics and alternations (Pater and Tessier 2003; 2006), and the nature of biases for structural simplicity (Moreton et al. To appear; Moreton 2008; Lai 2012). It has also provided evidence that humans generalize from the data they are provided with in training to novel test words (see all of the above), to novel segments (Cristiá and Seidl 2008; Cristiá et al. 2013) and to novel contexts (Myers and Padgett 2014). In this paper,

· Thank you to Elliott Moreton for discussion. This material is based on work supported by NSF Graduate Research Fellowships to R.D.S. and C.M.C under NSF DGE-0907995 and to B.H.Z. under NSF 1451512. The design of the experiment and writing of the paper were supported by NSF grant BCS-424077 to the University of Massachusetts Amherst.

we show that neurophysiological measures can be used to shed light on how generalizations acquired in the course of an experiment are cognitively encoded. In particular, we address the question of how similar lab-learned phonological patterns are to natively acquired phonological patterns. The process of learning an artificial pattern in a lab differs considerably from the process of acquiring a language's phonology naturalistically, but conclusions drawn from artificial phonology experiments typically depend on the assumption that the two processes share an underlying cognitive mechanism. In this paper, we use neurological measures to argue that our participants abstractly represent the pattern they have learned, in a way that is comparable to the representation of naturalistically learned phonological and syntactic rules.

In phonotactic learning studies, participants are trained on a set of words that obey a phonological restriction, and are tested on novel words that either obey or violate that restriction. Frequently, this test demonstrates that participants have acquired either implicit or explicit knowledge of the trained restriction. In this case, we ask whether this knowledge is represented similarly to natively acquired phonological generalizations. There are many relevant differences between the process of learning a pattern in the lab and native acquisition of phonology. The amount of input data provided in the lab, even in longer artificial language experiments, is radically different from the years of exposure that contribute to the acquisition of natural language phonology. Additionally, participants in an artificial language study are typically taught a single 'rule': all of the items in an experiment constitute evidence about that rule, and only one rule is taught at once. In the process of natural language acquisition, not only are many phonological rules being acquired in parallel, but so is a great deal of higher-level grammatical structure. Another potential difference between the two situations lies with the learner: the participants in lab-learning experiments are typically adults, and the acquisition of one's native phonology happens in childhood. It is possible that quite different learning processes are made use of by children and adults, and that very different cognitive representations of generalizations result (though see Birdsong 2006 for a critical overview of the 'critical period' literature).

Although adults demonstrably *can* learn phonological generalizations in the lab, the fact that the process is so different from natural first language acquisition raises the

question of what cognitive mechanism they are using to do so. Since the process of learning phonology in the lab bears a certain resemblance to the process of learning a second language as an adult, we turn to the L2 learning literature in order to understand what alternative strategies adults might use in the very early stages of acquiring a novel grammatical system. McLaughlin et. al. (2010) review several studies comparing early stages of L2 acquisition to later stages, arguing that learners go through two distinct stages, the latter of which closely resembles native grammatical knowledge. During early L2 acquisition, they argue, learners memorize specific strings, or probabilistic dependencies between specific strings. They are thus able to replicate the effects of grammatical knowledge in production and comprehension. Later, learners acquire an abstract grammatical rule, and begin to show native-like processing. The difference between these two stages of acquisition may be related to the difference between declarative memory (memorization of strings or probabilistic dependencies) and procedural memory (grammatical rules); see Ullman (2001, 2005) for more on the neurological grounding of this distinction.

McLaughlin et. al. (2010) specifically examine L2 learners' processing of grammatical violations, both syntactic and phonological, using event-related potentials (ERPs). They find that for both native speakers and more proficient L2 learners, grammatical violations elicit a late positivity, a positive-going potential beginning about 600ms after the onset of the grammatical violation. This component is typically found in response to grammatical violations (Osterhout and Holcomb 1992, et seq.), where it is called a P600, but a similar late positivity has been observed for phonological violations, both in the McLaughlin et al. paper, and in Domahs et al. (2009). Early L2 learners, in contrast, exhibit a larger N400 in response to grammatically illicit novel words and structures than to grammatically licit ones. The N400 is a component related to the process of lexical access and semantic integration. Syntactically acceptable but semantically surprising words elicit a large N400 ("I take my coffee with cream and socks", Kutas and Hillyard 1980), and nonwords elicit a larger N400 than do words (Rugg and Nagy 1987). We examine participants' knowledge of lab-learned phonotactics in light of these results, asking specifically whether lab-learned phonotactics are represented more like the early stages of L2 learning, or more like native grammatical

knowledge and the later stages of L2 learning. If lab-learned patterns are represented more like the early stages of L2 acquisition, we would expect novel items which violate a lab-learned pattern to exhibit a larger N400 than novel items which observe that pattern. If they are represented more like the later stages of L2 learning, and also like native grammatical knowledge, we would expect to see a late positivity in response to novel items that violate the lab-learned pattern.

We trained participants on a pattern of stop voicing agreement or disagreement in CVCV words. The patterns were learned in the course of the experiment by our English-speaking participants, whose native language allows them to be violated. Novel words that violated the pattern did not differ in N400 from novel words that observed the pattern, but they did elicit a larger late positivity. From this we conclude that the learned patterns of voicing agreement/disagreement are represented by our participants in a way that is similar to naturalistically learned grammatical generalizations. We believe this result rules out one possible interpretation of participants' performance on artificial language learning tasks, namely that participants are using a mechanism of analogy, judging the similarity of the novel test words to lexically stored representations of the training data. Such a mechanism would rely on participants memorizing the individual items they were trained on, but would not require them to have formed an abstract generalization of the pattern. Daelemans et al. (2000: 76) state that these sorts of exemplar-based / analogical models "do not derive explicit abstractions such as rules, but instead basically memorize the learning material and generalize by analogical reasoning on the basis of examples stored in memory". Other explicit models of analogy of this type include Skousen (1989), Eddington (2000), Nakisa and Hahn (1996), and many others.

For behavioral data, even in natural language, the predictions of these models often do not diverge very much from the predictions of models which appeal to abstract rules or constraints, and in fact, Daelemans et al. (2000) and others argue for the superiority of an analogical model based on its better match to data produced by human learners (see Albright and Hayes 2003 and Daland et al. 2011 for arguments going the other way). Like McLaughlin et. al. (2010) and Ullman (2001, 2005), we assume that many generalizations can be produced by either an lexical / analogical mechanism or by

an abstract / grammatical mechanism. Because the late positivity we observe is so closely linked in previous work to violations of abstract structure, including syntactic, phonological, and musical structure, we believe that its presence in our experiment implies that here, too, participants have learned an abstract generalization.

In the next section, we provide background on ERP methodology and some of the relevant prior results. We then present the methodology for our experiment (section 3), its results (section 4), and a discussion of the significance of the results for our understanding of the cognitive underpinnings of lab-learned phonology, and how they fit with other ERP results on laboratory learning, and on naturalistically learned phonology (section 5).

2. ERP background

Electroencelogram (EEG) recordings measure electrical activity of the brain using recording electrodes placed on the scalp. Event-related potentials (ERPs) are obtained by averaging many segments of EEG that are time-locked to a particular stimulus or event. The resulting waveforms show average voltage over time, which can be negative or positive relative to a pre-stimulus baseline. A comparison of ERPs elicited in different experimental conditions yields information about differences in brain activity across those conditions. In what are called violation paradigms, ERPs are compared in trials in which an expectation or constraint is violated, versus when it is satisfied. Linguistic research using the violation paradigm has led to the identification of several 'components', characteristic waveform differences that are relatively consistently found in similar experimental manipulations. Of these, we will be primarily concerned with the already mentioned LPC/P600, as well as the N400. These components receive their names from the direction of the voltage deflection – either positive (P) or negative (N) – and from the latency of the deflection (e.g. a relatively late peak at about 600 ms or 400 ms after stimulus onset).

The N400 was first identified as a response to semantically incongruous, but syntactically acceptable words, such as "He took a sip from the transmitter" (Kutas and Hillyard 1980; Kutas and Federmeier 2011). The amplitude of the N400 has been interpreted as an index of the effort involved in lexical access. For example, nonwords

elicit larger N400s than actual words (Rugg 1984), and novel words (e.g., toose) preceded by rhyming strings (e.g., buice) elicit a smaller N400 than the same items preceded by unrelated strings (e.g., gock) (Coch et al. 2015). Many other factors affect the amplitude of the N400; high-cloze words elicit a smaller N400 than equally plausible low-cloze words (Kutas and Hillyard 1984), more recently accessed words elicit a smaller N400 (Petten et al. 1991), and both words and pseudowords with fewer lexical neighbors elicit a smaller N400 (Holcomb et al. 2002).

An LPC has been observed for a range of syntactic violations, including agreement, phrase structure, subcategorization, and constraints on long-distance dependencies (see Gouvea et al. 2010; Morgan-Short et al. 2012 for overviews). Because in this context it usually has a peak at around 600 ms post-stimulus, it is usually referred to as a P600, though a variety of factors affect its latency (i.e., the peak is not always at 600 ms), as well as its distribution, the scalp regions in which it is found (Gouvea et al. 2010). There are a variety of proposals about the functional interpretation of the LPC/P600, though there is general agreement that it reflects the evaluation of an abstract structural relation. The LPC was first identified for syntactic violations, but has subsequently been found for phonotactic violations: Finnish vowel harmony in McLaughlin et. al (2010), and the German SCVC constraint in Domahs et. al. (2009). LPC's have also been found for violations of musical structure (Patel et al. 1998; see Carrión and Bly 2008 for an overview), and for rule violations in arithmetic tasks (Núñez-Peña and Honrubia-Serrano 2004). While the LPC is not language-specific, it is also an indicator of abstract structural relations in these other cognitive domains: "an index of detection for any anomaly in rule-governed sequences" (Núñez-Peña and Honrubia-Serrano 2004, 130); [it] "reflects processes of knowledge-based structural integration" (Patel et al. 1998, 51).

The LPC has been argued to be an instance of the P300 - a component which is elicited when certain stimuli are more infrequent or more salient than others (Sutton et al. 1965, Duncan-Johnson and Donchin, 1977). Coulson et al. 1998 examine various factors that influence the amplitude and latency of the P300, such as salience and task relevance, and show that they also influence the P600. More recently, Sassenhagen et al. 2014 argue that the P600 also exhibits one important property of the P300, namely that its timing is

dependent on when a response is made, rather than when a stimulus is experienced. Arguments against this P600-as-P300 hypothesis come from studies such as Osterhout et al 1996, who find both a P600 and a P300 in the same experiment, occurring in different conditions, and with slightly different timing and distribution. We will ultimately argue in this paper for distinct P600 and P300 components, since we find something quite similar to Osterhout et al: we find two distinct positivities occurring in different conditions and with different timing.

Previous research using ERPs to examine phonotactic knowledge has focused on cases of 'perceptual repair', in which listeners have difficulty accurately perceiving sequences which are disallowed in their language (e.g. Dehaene-Lambertz et al, 2000, Breen et al, 2013). We do not expect perceptual repair to occur with lab-learned phonotactics, since both observers and violators of the pattern are legal in the participants' native language, English. Instead, we compare the processing of lab-learned phonotactics to the correctly-perceived phonotactic violations examined in McLaughlin et al (2010) and Domahs et al (2009), which both elicit an LPC. Based on previous research on the N400 and the LPC, we expect to find a larger N400 for words presented in the training phase of an artificial language learning experiment than for words that were newly presented in the test phase. Such an N400 difference would mean that participants had successfully learned the words presented in the training, and could distinguish between them and novel words in the same way that participants in Rugg (1984) and many subsequent studies can distinguish between words of their language and novel words. We might also expect to find a larger LPC for novel words that violate a generalization over the training words than for novel words that satisfy it, if the knowledge of that generalization is abstractly represented, in a way that is similar to naturalistically learned phonology (McLaughlin et al. 2010). Finally, if participants learn a generalization over the training items in a way that is similar to the early stages of L2 learning, we would expect to find a greater N400 for novel violators than for novel observers.

3. Methods

We taught 24 adult, native English speakers 16 word-object pairings by asking them to match an auditorily presented word to one of four pictures, after which they were given the correct pairing. The words each participant learned were all consistent with a phonotactic pattern. In testing, participants were asked to rate on a 4-point scale how likely it is that each word is part of the language they were learning. These words included half of the trained words (Studied), eight novel words that fit the pattern (Novel-Fit), and eight words that violated it (Novel-Violate). Testing and training blocks alternated, with a total of five each. We adopted this alternating training-testing procedure so that we could collect sufficient EEG data in the test blocks for our ERP analysis without "untraining" the participants with a too-long single test block that contains words that violate the restriction.

3.1 Materials

As the phonotactic patterns for our participants to learn, we selected voicing agreement and disagreement between stops in CVCV words. The patterns, and the stimulus space, were as studied by Moreton (2008; 2012). The stops were drawn from the set [d, g, t, k], and the vowels from the set [i, æ, u, ɔ]. We constructed 48 words, with the consonants in half of them agreeing in voice (e.g. [dugi], [tikɔ]), and in the other half disagreeing (e.g. [kædu], [tigæ]). Vowels were chosen to avoid patterns in terms of co-occurrence with one another, or with the consonants. From this set of items, four groups of items were created, shown in Table 1. For the two Voice-Match lists, half of the voice-agreeing items were used as training items, and half were used in testing. The items that were used in training for Group 1 were used in testing for Group 2. Additionally, half of the voice-disagreeing items were used as the novel pattern-violating items in testing for Group 1, and the other half of those items were used for Group 2. Two more lists were created for the Voice-Mismatch condition in a similar way. Twelve participants were assigned to the Voice-Match condition, and twelve to the Voice-Mismatch condition. Within each condition, participants were assigned to one of two sets of items. Each participant saw each item in only one role, but across participants each stimulus appeared during training,

during testing, and as a pattern-violating nonword. For Group 1, eight pattern-observing items were chosen to be used in training, and the other eight were. There were two sets of materials, which were created to achieve a design in which all test words appeared overall equally often as Studied, Novel-Fit, and Novel-Violate items. In this way, physical differences amongst the stimuli were perfectly controlled, and any differences in behavioral or ERP data could be attributed to differences in training. Of the 24 words that either had voicing agreement or disagreement, 8 appeared only in the training session for the appropriate condition (Voice-Match or Voice-Mismatch). The other 16 were used in the test phase: as Studied and Novel-Fit items in that same condition, and as Novel-Violate items in the other. The distribution of the 32 test words across our 4 groups of 6 participants is shown in Table 1.

Table 1. Items, and counterbalancing scheme.

	List	Agreeing		Disagreeing	
		Set 1 tɔtu gigæ dægɔ kuti didɔ kikɔ tɔki gɔdi	Set 2 tɔti gɔgi digɔ kitɔ dædɔ kuki tɔku gidæ	Set 3 tɔdu gikæ dækɔ kudi ditɔ kigɔ tɔgi gɔti	Set 4 tɔdi gɔki dikɔ kidɔ dætɔ kugi tɔgu gitæ
Voice-Match	Group 1	**Studied**	Novel-Fit	Novel-Violate	
	Group 2	Novel-Fit	**Studied**		Novel-Violate
Voice-Mismatch	Group 3	Novel-Violate		**Studied**	Novel-Fit
	Group 4		Novel-Violate	Novel-Fit	**Studied**

As inspection of Table 1 will show, our design was counter-balanced in that the Studied and Novel-Fit forms for participants in the Voicing-Match condition were Novel-Unfit forms for participants in the Voicing-Mismatch condition, and *vice versa*. The 24 words in each "language" were equally likely to start with a voiced or voiceless consonant or with a velar or alveolar place of articulation. Vowel features were also controlled in each position. English words that fit the constraints (e.g., *duty* and *gaudy*) were excluded.

The words were pronounced by a 26-year-old linguistically trained male native speaker of English. The speaker pronounced the words with stress on the initial syllable

and a full vowel in the second syllable in the frame sentence "It was X said Tim." The sentences were recorded to 32 bit / 44.1 khz digital format. The words were segmented from the sentences using the offset/onset of noise for the surrounding sibilants as criterion. The peak amplitude of all items was normalized to their mean, and a 10 ms sinusoidal fadeout was applied to the end of each recording to eliminate the effects of trimmed formants.

Words averaged 367 ms (*SD* = 34) in duration with the 2nd syllable beginning 136-245 ms after the first (*M* = 191, *SD* = 29). All sounds were presented over a pair of M-Audio StudioPro3 loudspeakers with EPrime software running on a PC with a Creative Audigy 2 ZS sound card. Both loudspeakers were located directly above the computer monitor. Sounds were presented at 65 dB SPL (A-weighted) as measured from the location of participants' heads.

The objects used in training were presented as color photographs of a common concrete object (e.g., puppy, ship, shoe) that participants would be expected to describe with a single English word. Images were cropped to leave minimal background behind the objects. Pictures were then resized such that four pictures shown at the same time along with their response labels (1-4) filled the space available on the computer monitor. The image that was "correct" in training for a given aurally presented label was presented at an identical size when shown among the three distractors and when it was shown afterwards with the aural label.

3.2 Procedure

Instructions for both the training and testing tasks were given at the beginning of the session. For training, participants were told they would be learning some of the words in a made up language by matching the spoken words to pictures of the objects the words refer to. They were warned that their initial responses would be guesses, but that seeing the correct answer after every response would help them to learn the words. Participants initiated a training block by pressing any button on a response pad. Each training trial began with the appearance of a fixation cross on the computer monitor. One of the 16 training words was presented from the loudspeakers 700 – 1200 ms (randomly selected, rectangular distribution) later. The fixation cross remained on the screen for 500 ms after

the offset of the word and was then replaced by pictures of 4 objects. Each object was labeled with a number (1-4) that corresponded with a button on the response pad. Participants pressed a button to indicate the meaning of the word they had heard. Immediately following any button press, the correct picture was shown in the middle of the computer monitor for 1000 ms and the word was played again. EEG data collected while pictures were shown were not analyzed and participants were free to move their eyes and blink during these presentations. In each training block, all 16 words were presented 5 times each in random order (80 trials per block).

During testing, participants were asked to rate how likely it is that each word is part of the language they were learning. They were told that some of the test words had been heard during training and were clearly part of the language. They were instructed that even though they had not learned a meaning for most of the test words, some of them were part of the same language and some were not. The experimenter shared the example that people can often tell if a word sounds like it could be Italian or Japanese even if neither of those languages is familiar. They were also encouraged to use all four response buttons rather than only the 1 labeled "unlikely a word" and the 4 labeled "very likely a word." Participants began each test block by pressing any button. At the beginning of each trial, the fixation cross was shown on the computer monitor. One of the test words was presented over the loudspeakers 700 – 1200 ms later. The fixation cross remained on the screen for 1000 – 1500 ms after the word onset and was then replaced by the response prompt "Likely a word?" with the labeled scale. A test trial ended after any response was given. In each test block, all 24 words (eight each of Studied, Novel-Fit, and Novel-Violate) were presented once in random order.

The training block – test block sequence was presented 5 times for a total of 400 training trials and 120 test trials (40 of each type). For all trials, participants were asked to refrain from blinking, moving their eyes, or moving any other part of their body, including moving a finger to press a button, whenever the fixation cross was shown on the screen. They were encouraged to make these and any other movements while the pictures (training) or response prompt (testing) were shown. Participants were asked to continue from each training block to the following test block without pause; they were encouraged to take breaks after each test block. At the end of the experiment, participants

were asked if they had noticed anything about the language they learned and if they had developed any strategy to distinguish between words that were and were not in the language.

EEG was recorded continuously throughout the training and test trials (250 Hz sampling rate, 0.01-100 Hz bandpass) from 128 electrodes (EGI, Eugene OR). Scalp impedances at all electrode sites were maintained under 50 kΩs. Segments of EEG from 100 ms before to 500 ms after the onset of training words and from 100 ms before to 1000 ms after the onset of test words were examined. Trials with artifacts from muscle tension, blinks or eye-movements, or motion were excluded from analysis. EEG from remaining training trials was averaged together by each block; EEG from test trials was averaged by condition (Studied, Novel-Fit, Novel-Violate) across all blocks. The 100 ms before word onset were used as a baseline and ERPs were rereferenced to the averaged mastoid recording.

For training trials, average amplitude measurements were taken 40 – 70 ms (P1), 90 – 130 ms (N1), and 230 – 500 ms (N400) after word onset. For test trials, mean amplitude measures were made in the same P1 and N1 windows as well as 400 – 700 ms (N400 and P300) and 600 – 1000 ms (LPC) after word onset. Measurements were made at 100 of the 128 electrode sites across the scalp such that electrode position could be included as multiple factors in statistical analyses. Specifically, measurements from 4 electrodes were averaged together in a 5 (Anterior, Anterior-Central, Central, Posterior-Central, Posterior or ACP) x 5 (Left, Left-Medial, Medial, Right-Medial, Right, or LMR) grid. Data from each of the three measurements taken from training trials were entered in 5 (Block) x 5 (ACP electrode position) x 5 (LMR electrode position) repeated-measures ANOVAs. Data from test trials were entered in 3 (Word Type) x 5 (ACP electrode position) x 5 (LMR electrode position) repeated-measures ANOVAs. Greenhouse-Geisser corrected p-values (and uncorrected degrees of freedom) are reported. Significant ($p < .05$) effects of Block in the training data were followed up by comparisons of each training block with block 1. Significant ($p < .05$) interactions of Word Type and electrode position factors were followed by ANOVAs conducted on data collected at subsets of electrodes and to compare the ERPs elicited by Studied words to both types of novel words as well as Novel-Fit and Novel-Violate words.

3.3 Participants

Twenty-four native English speakers (9 females, ages 19 to 33 years) provided the data included in analysis. An equal number of participants (N = 6) were trained on each configuration of each language (see Table 1 above). Data from one participant were excluded because she expressed explicit knowledge of the phonological generalization[1] and from one other because of excessive high-frequency noise in the EEG, likely caused by muscle tension. All participants reported being right-handed, having no neurological problems, and not taking psychoactive medication within a year of the study. Participants provided informed consent and were compensated for their time at a rate of $10/hour.

4. Results

We first discuss the results on the words presented in training. During the first training block, participants were already well above chance performance of 25% on the word-picture matching task (M = 53.9%, SD = 2.4); in the four subsequent training blocks, performance was even better (M = 89.2%, SD = 1.8). The feedback provided after each trial was clearly sufficient for adults to learn the meanings of 16 words in an artificial language over a short period of time. During testing, participants rated the Studied words as more likely to be in the language that they were learning (M = 3.72, SD = .18) than Novel-Fit words (M = 2.71, SD = .29) ($t(23) = 14.77$, $p < .001$). ERPs recorded during training revealed an N400 that decreased in amplitude over the course of the experiment. There was a main effect of block across all electrodes on amplitude in the 230-500 ms range ($F(4,92) = 2.952$, $p = .037$). The negativity in this time window was larger for block 1 (M = -1.38μV, SE = .30) than in the other four blocks (M = -.81μV, SE = .29).

[1] Of the 24 participants that contributed data to analysis, 19 reported adopting a strategy to learn the word/picture relationships. These approaches involved various types of mnemonics (e.g., [kudi] is the "current deep" where the ship sails, and [tɔki] sounds like something you would say to a little puppy). However, only five participants reported trying to come up with a system for determining which words were in the language at the time of testing; all five also stated that their attempts failed. These participants considered patterns such as "if a word starts with k, it has to end with i and if it starts with g it has to end with a." All participants, including these five, agreed with the statement "In the end, I just guessed about the test words." The only participant who suggested that a pattern existed between the consonants in the words identified the correct pattern (i.e., consonants match in voicing); data from this person were excluded from analysis.

These results are consistent with the behavioral data showing that learning of the word meanings occurred rapidly and was maintained. During testing, Studied words elicited a smaller N400 than Novel-Fit words over central and central-posterior regions. At 400-700 ms after stimulus onset, across all electrodes, there was an interaction between word category and anterior-posterior electrode position ($F(4,92) = 3.398$, $p = .047$). Motivated by this interaction, we ran a subanalysis over just the central and central-posterior regions, where the effect was significant for Studied and Novel-Fit words ($F(1,23) = 6.67$, $p = .017$). In contrast, there was no evidence that amplitude 400-700 ms after onset differed between Novel-Fit and Novel-Violate items (p's > .375). Further, there was no evidence that the differences for Studied and Novel words differed by group (p's > .500). This result is consistent with the ERP results from training. Even during the testing trials, the easing of lexical access afforded by familiarity was evident as a difference in N400 amplitude.

In the same time window, and motivated by the same word category x anterior/posterior electrode position, we conducted a subanalysis over central-posterior and posterior regions. There, Studied words also elicited a larger positivity than Novel-Fit words ($F(1,23) = 11.19$, $p = .003$). We believe that this positivity is an example of a P300. P300's can be elicited by less probable items relative to more probable items (Donchin, 1981) and by items which require a response relative to items which do not (Duncan-Johnson and Donchin, 1977). In the testing session, trained items were the minority, making up one third of the stimuli. While all items in the testing session required a response, trained items required a qualitatively different type of response than untrained items (whether pattern-fitting or not), since participants only had to remember the item from training, and did not have to judge it based on its sound pattern. The N400 and P300 are illustrated in Figures 1 and 2.

The critical comparison for generalization of the phonotactic patterns is between the Novel-Fit words and the Novel-Violate words. The participants rated the Novel-Fit words as more likely to be in the language than Novel-Violate ($M = 2.21$, $SD = .26$) ($t(23) = 7.98$, $p < .001$). We found no evidence that the pattern of ratings differed for the Voicing-Match and Voicing-Mismatch languages (across all types of test words, $p > .15$).

We now turn to the comparison of ERPs elicited by Novel-Fit and Novel-Violate items during testing, which provides information on how the learned phonotactic generalizations are represented. Mean amplitude in a later time-window, 600-1000 ms after word onset, was investigated. Across all electrodes, a word category x anterior/posterior interaction was found ($F(4,92) = 3.423$, $p = .042$). This interaction motivated a subanalysis, at posterior and posterior-central regions. Here, novel words that did not fit the pattern of the trained language elicited a larger positivity ($F(1,23) = 5.55$, $p = .027$). There was no indication that this effect interacted with language group ($p = .987$). This LPC in response to Novel-Violate items is shown in Figures 1 and 3. The P300 effects evident for Studied compared to Novel-Fit words does show some temporal and spatial overlap with the Fit/Unfit LPC, but the P300 started earlier than the LPC. The timing and distribution of all three components can be compared visually in Figure 1. Although it has been argued that late positivities like our LPC (and the well-known P600) are instances of the P300 (e.g. Sassenhagen et al, 2014), in the present data the two components are distinct in timing as well as in the conditions they appear in. While the LPC is elicited by Novel-Unfit items relative to novel-fit items, the P300 is elicited by Studied items relative to Novel-Fit items. While a sensible explanation exists for each of these differences independently, we can think of no account that unifies the novel-unfit items and the trained items as opposed to the novel-fit items.

5. Conclusions and Discussion

The participants in our experiment learned a dependency between the voicing of the two stop consonants of CVCV words.[2] They were exposed to a set of words obeying the restriction in the context of learning the meanings of the words, and then in testing they rated novel words that fit the restriction as more likely to belong to the language than novel words that violated it. From the study of EEG data collected during the experiment, we conclude that the phonotactic generalization is abstractly or grammatically

[2] For half the participants the consonants always agreed, and for half they always disagreed. The fact that we saw no differences between the groups fits with a general lack of evidence for a difference between long-distance assimilation and dissimilation in the artificial phonology learning literature (Moreton and Pater 2012a; Moreton and Pater 2012b). Some recent work indicates, however, that vowel harmony may have a learning advantage over vowel dissimilation (Guevara Rukoz 2015).

represented, rather than the product of lexical search or analogy. The ERP response to the Novel-Violate items included a Late Positive Component (LPC), similar to that found in response to syntactic and musical harmonic structure violations. We did not find an N400 between Novel-Fit and Novel-Violate items, which might have been expected if lexical search was involved in determining the acceptability of these new words.

The only previous neurophysiological study of the outcome of laboratory phonological learning of which we are aware is that of Wong et al. (2013), who focus on a distinction between the learning of what they call analogical and concatenative grammars, though for them analogy is a way of characterizing the knowledge of an opaque alternation (concatenation is simple addition of a suffix). Their focus is also different from ours in that they are concerned with individual differences in the learning of these two types of paradigmatic relation. The intersection of our study's concerns and theirs is a good topic for future research: are there individual differences in the learning of phonotactics in terms of a reliance on different neural subsystems?

Our results add to the broader literature on laboratory learning of language, in which there has been some previous ERP research on the outcome of morpho-syntactic acquisition in the lab. As in our study, Morgan-Short et al. (2012) show a relatively quick acquisition of an LPC. They also find a difference between implicit and explicit learning conditions, in that only implicit learning yielded an early anterior negativity (see Morgan-Short et al. To appear on early negativities in naturalistically learned syntax). We did not find this component in our study, and it is an open question whether it will be observed in phonological violations, which are less well-studied than syntax (Loui et al. 2009 find both an LPC and an early anterior negativity in responses to unexpected chords in a newly learned harmonic system). Again, the intersection between our study and this earlier work seems like a fruitful area for further work: do implicit and explicit training differentially affect phonological learning (see Moreton and Pertsova 2015)?

Finally, a general implication of our result is that it lends support to the view that laboratory learning of phonology, while different in many ways from naturalistic acquisition, has some ecological validity, given the Domahs et al. (2009) finding discussed above that a long-distance restriction on the place of consonants in sCVC words in German also yields an LPC.

References

Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90: 119–161.

Birdsong, David. 2006. Age and second language acquisition and processing: a selective overview. *Language Learning* 56:s1 9-49.

Carpenter, Angela C. 2010. A naturalness bias in learning stress. *Phonology* 27: 345–392.

Carrión, Ricardo E., and Benjamin Martin Bly. 2008. The effects of learning on event-related potential correlates of musical expectancy. *Psychophysiology* 45: 759–775.

Coch, Donna, Giordana Grossi, Wendy Skendzel, and Helen Neville. 2015. ERP Nonword Rhyming Effects in Children and Adults. *Journal of Cognitive Neuroscience* 17: 168–182. doi:10.1162/0898929052880020.

Cristiá, Alejandrina, Jeff Mielke, Robert Daland, and Sharon Peperkamp. 2013. Similarity in the generalization of implicitly learned sound patterns. *Laboratory Phonology* 4: 259–285.

Cristiá, Alejandrina, and Amanda Seidl. 2008. Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development* 4: 203–227.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis, and Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 29: 197–234.

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. doi:10.1177/0023830909336581.

Eddington, David. 2000. Spanish stress assignment within the analogical modeling of language. *Language 76:* 92-109

Frisch, Stefan, Sonja A. Kotz, D. Yves von Cramon, and Angela D. Friederici. 2003. Why the {P600} is not just a P300: the role of the basal ganglia. *Clinical Neurophysiology* 114: 336 – 340. doi:http://dx.doi.org/10.1016/S1388-2457(02)00366-8.

Gillis, Steven, Walter Daelemans, and Gert Durieux. 2000. 'Lazy Learning' : A Comparison of Natural and Machine Learning of Stress. In P. Broeder & J. Murre (eds.), *Models of Language Acquisition*. Oxford University Press, 76-99.

Gouvea, Ana C., Colin Phillips, Nina Kazanina, and David Poeppel. 2010. The linguistic processes underlying the P600. *Language and Cognitive Processes* 25: 149–188.

Guevara Rukoz, Adriana. 2015. The role of phonetic naturalness in shaping sound patterns: Evidence from artificial language learning and computational modeling. M.A. Thesis, EHESS – ENS –Université Paris Descartes.

Holcomb, Phillip J., Jonathan Grainger, and Tim O'Rourke. 2002. An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience* 14: 938–950.

Kutas, Marta, and Kara D. Federmeier. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology* 62: 621–47.

Kutas, Marta, and Steven A. Hillyard. 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207: 203–205. doi:10.1126/science.7350657.
Kutas, Marta, and Steven A. Hillyard. 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307: 161–163.
Lai, Yeeking Regine. 2012. Domain specificity in learning phonology. University of Delaware.
Lenneberg, Eric H. 1967. *Biological foundations of language.* New York: Wiley
Loui, Psyche, Elaine H. Wu, David L. Wessel, and Robert T. Knight. 2009. A generalized mechanism for perception of pitch patterns. *J. Neurosci.* 29: 454–459.
McLaughlin, Judith, Darren Tanner, Ilona Pitkänen, Cheryl Frenck-Mestre, Kayo Inoue, Geoffrey Valentine, and Lee Osterhout. 2010. *Language Learning* 60: 123-150.
Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25: 83–127.
Moreton, Elliott. 2012. Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language* 67: 165–183.
Moreton, Elliott, and Joe Pater. 2012a. Structure and substance in artificial-phonology learning: Part I, Structure. *Language and Linguistics Compass* 6: 686–701.
Moreton, Elliott, and Joe Pater. 2012b. Structure and substance in artificial-phonology learning: Part II, Substance. *Language and Linguistics Compass* 6: 702–718.
Moreton, Elliott, Joe Pater, and Katya Pertsova. To appear. Phonological concept learning. *Cognitive Science*.
Moreton, Elliott, and Katya Pertsova. 2015. Implicit and explicit phonology: What are artificial-language learners really doing? *Handout from the Manchester Phonology Meeting*.
Morgan-Short, Kara, Mandy Faretta-Stutenberg, and Laura Bartlett. To appear. Contributions of event-related potential research to issues in explicit and implicit second language acquisition. In *Implicit and Explicit Learning of Languages.* Amsterdam: John Benjamins.
Morgan-Short, Kara, Karsten Steinhauer, Cristina Sanz, and Michael T. Ullman. 2012. Explicit and Implicit Second Language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience* 24: 933–947.
Myers, Scott, and Jaye Padgett. 2014. Domain generalisation in artificial language learning. *Phonology* 31: 399–433. doi:10.1017/S0952675714000207.
Nakisa, Ramin and Ulrike Hahn. 1996. Where defaults don't help: the case of the German plural system. *Proceedings of the 18th Annual Conference of the Cognitive Science Society*: 177-182
Núñez-Peña, M. Isabel, and M. Luisa Honrubia-Serrano. 2004. P600 related to rule violation in an arithmetic task. *Cognitive Brain Research* 18: 130 – 141. doi:http://dx.doi.org/10.1016/j.cogbrainres.2003.09.010.
Osterhout, Lee, and Phillip J. Holcomb. 1992. Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language* 31: 785-806.
Patel, Aniruddh D., Edward Gibson, Jennifer Ratner, Mireille Besson, and Phillip J. Holcomb. 1998. Processing Syntactic Relations in Language and Music: An

Event-Related Potential Study. *J. Cognitive Neuroscience* 10: 717–733. doi:10.1162/089892998563121.

Pater, Joe, and Anne-Michelle Tessier. 2003. Phonotactic knowledge and the acquisition of alternations. In *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*.

Pater, Joe, and Anne-Michelle Tessier. 2006. L1 phonotactic knowledge and the L2 acquisition of alternations. In *Inquiries in linguistic development: in honor of Lydia White*, ed. Roumyana Slabakova, Silvina Motrul, Philippe Prévost, and Lydia White, 115–131. Benjamins.

Petten, Cyma Van, Marta Kutas, Robert Kluender, Mark Mitchiner, and Heather McIsaac. 1991. Fractionating the word repetition effect with event-related potentials. *Cognitive Neuroscience, Journal of* 3: 131–150.

Pycha, Anne, Pawel Nowak, Eurie Shin, and Ryan Shosted. 2003. Phonological rule-learing and its implications for a theory of vowel harmony. In *Proceedings of the 22nd West Coast Conference on Formal Linguistics (WCCFL 22)*, ed. M. Tsujimura and G. Garding, 101–114.

Rugg, Michael, and Margaret Nagy. Lexical contribution to nonword-repetition effects: evidence from event related potentials. *Memory and Cognition* 15: 473-481

Skousen, Royal. 1989. *Analogical modeling of language*. Springer.

Ullman, Michael T. 2001. The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition,* 4: 105-122.

Ullman, Michael T. 2005. A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In *Mind and context in adult second language acquisition*, ed C. Sanz, 141-178. Washington, DC: Georgetown University Press.

Wilson, Colin. 2003. Experimental investigation of phonological naturalness. In *Proceedings of the 22nd West Coast Conference on Formal Linguistics (WCCFL 22)*, ed. G. Garding and M. Tsujimura, 533–546. Somerville: Cascadilla Press.

Wong, Patrick C. M., Marc Ettlinger, and Jing Zheng. 2013. Linguistic grammar learning and DRD2-TAQ-IA polymorphism. *PloS one* 8: e64983.

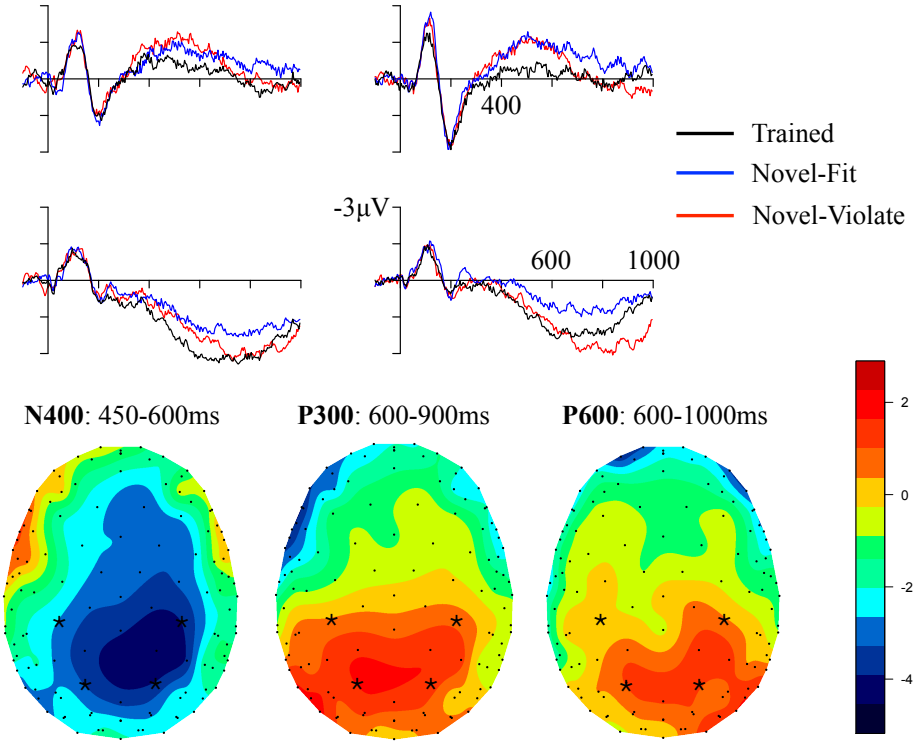

Figure 1. Studied, Novel-Fit, and Novel-Violate. Waveforms are time locked to the onset (vertical lines) of test items that had been studied (black), that were presented only during testing that fit the phonological pattern (blue), and that were presented only during testing but did not fit the phonological pattern (red). These data were measured at the four electrodes indicated (stars). Studied words elicited a smaller N400 and a larger P300. Novel-Violate items elicited a larger Late Positive Component (LPC) over posterior regions.

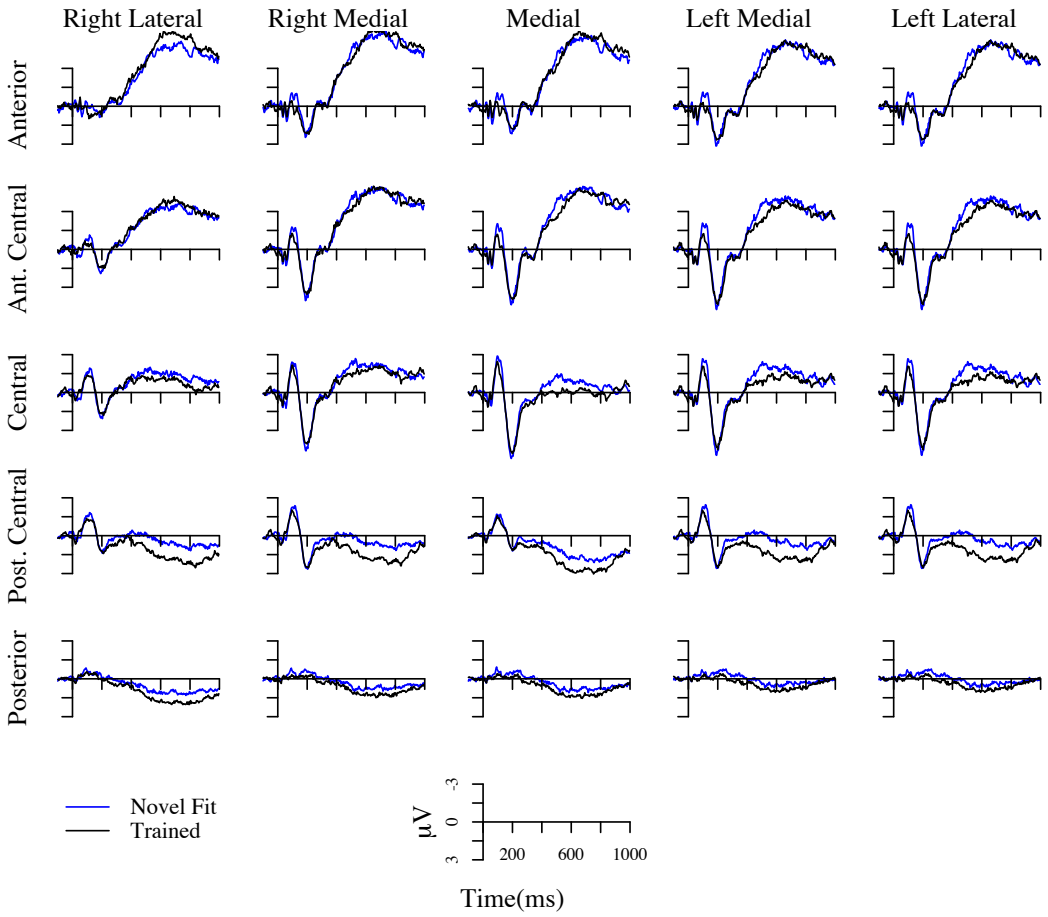

Figure 2. Novel Fit and Trained. Waveforms timelocked to the onset of the stimulus, as in Figure 1. Each one is an average over 4 electrodes, located as indicated by the row and column labels.

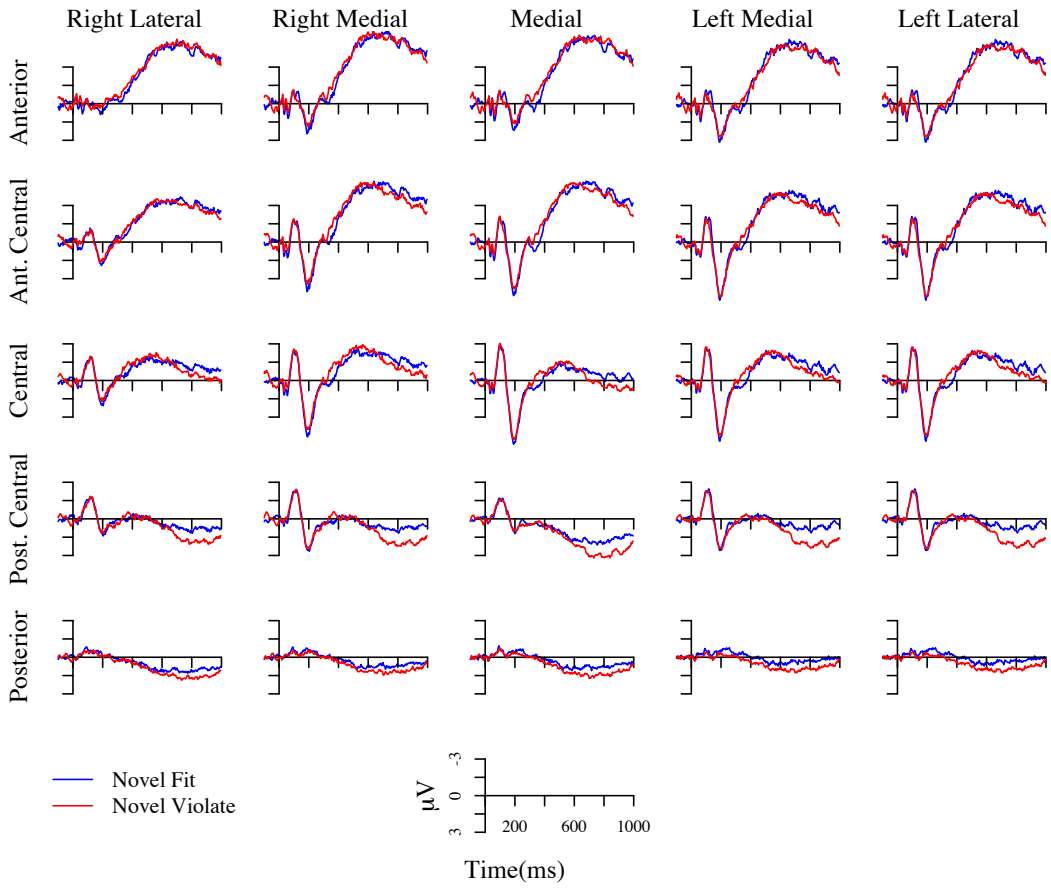

Figure 3. Novel Fit and Novel Violate. ERP waveforms as in Figure 2, each one an average over 4 electrodes, located as indicated by the row and column labels.

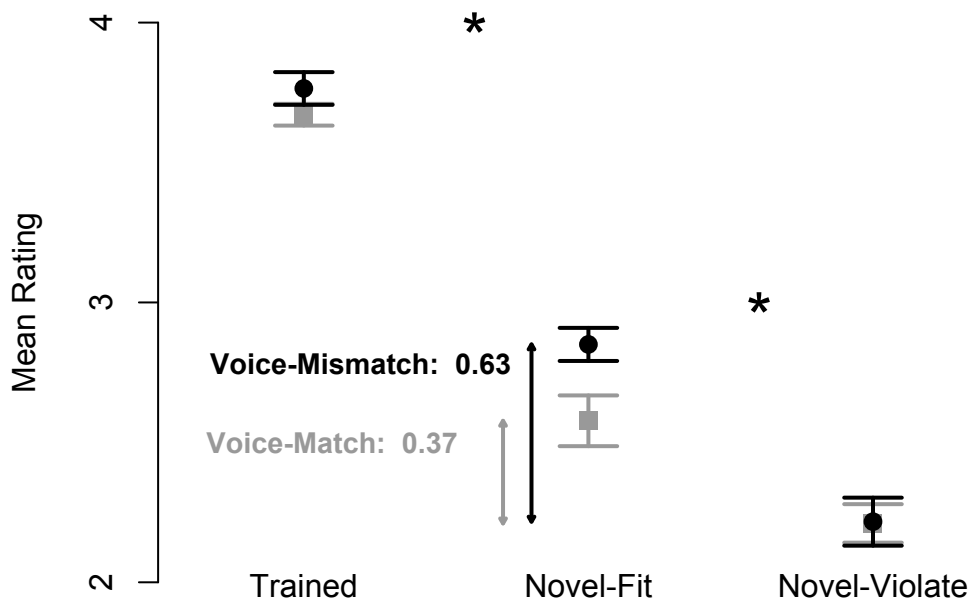

Figure 4: Rating responses during the training session. Voice-Match condition values are grey squares and Voice-Mismatch black circles. In both Voice-Match (assimilation) and Voice-Mismatch (dissimilation groups), Novel-Violate items were rated as less likely to be in the language than Novel-Fit items.

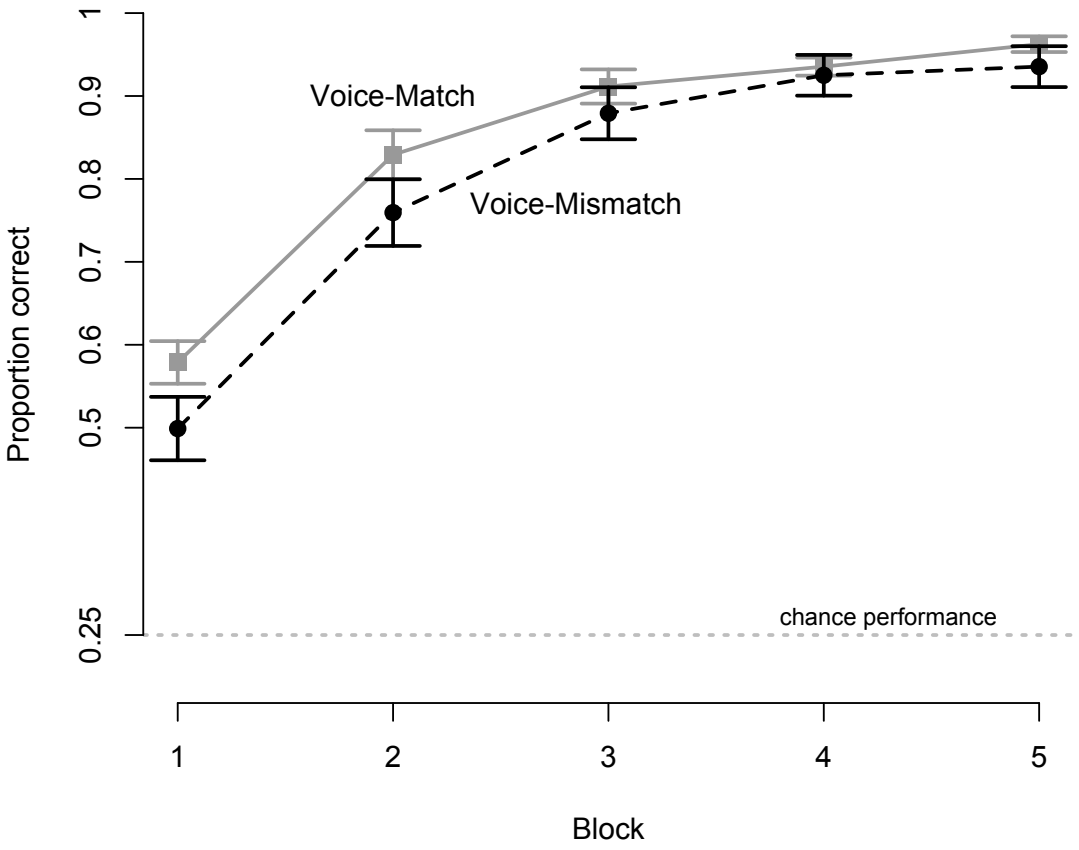

Figure 5: Proportion correct responses across blocks of training. Chance performance is at 25% because participants are choosing between four pictures while they are learning the words.