What is learned during statistical learning?

Ava Kiai, Lucia Melloni

Abstract

Contents

1 Introduction

Statistical learning (SL) is one mechanism by which the brain is able to parse continuous streams of sensory information, based on simple statistical features alone. Dehaene et al. 2015 SL paradigms have shown the brain to be surprisingly effective at detecting statistical regularities in highly impoverished or artificial inputs, leading to a discrimination of meaningful boundaries where previously none were discernible. Pelucchi, Hay, and Saffran 2009; Richard N. Aslin, Saffran, and Newport 1998; Turk-Browne, Jungé, and Brian J. Scholl 2005; Turk-Browne, Brian J Scholl, et al. 2008

Statistical learning is canonically measured using an explicit word discrimination task after exposure to the continuous speech stream.

Several online statistical learning measures have been previously shown to be effectively modulated by implicitly-acquired knowledge of the statistical regularity of reoccurring syllables.

We investigated whether dynamic online measures of statistical learning are able to predict performance in the canonical offline task. In a first experiment, we implemented a target detection task to measure SL online and an offline explicit word recognition task. In a second study, we asked participants to perform the online target detection task to further explore the informative capacity of this task, in both a structured and random condition.

2 Experiment 1

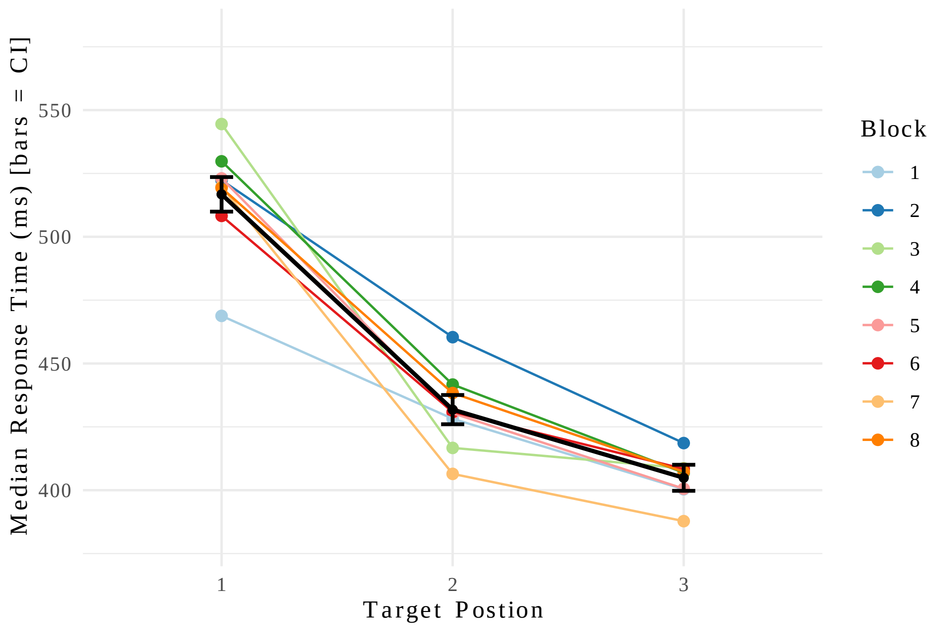
2.1 Method

41 individuals participated in the study (x female, mean age, +- sd). Two participants were removed from the data pool due to technical failure. Of the 39 remaining datasets, 33 were used in analyzing the target detection task (one participant failed to follow instructions, and technical issues caused partial data loss for the other five). However, technical failures in our program were limited to the current task being run. Thus, we used data from 38 participants in analyzing the word recognition task (data from one participant was overwritten).

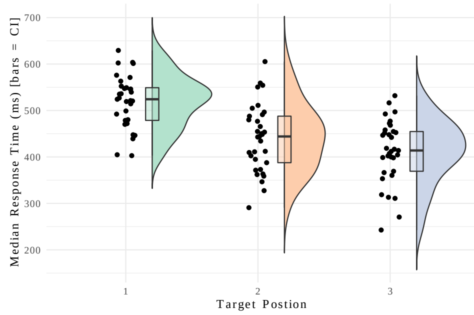
2.2 Results

To replicate findings that showed graded reaction times in response to syllables in different ordinal positions, we ran a generalized linear model with reaction time (in seconds) as outcome variable, fitted with a gamma function and log link function. Our full model included both ordinal position and block as fixed effects factors, and subject as random intercept random effects factor. This model was compared with a lesser model in which only ordinal position was input as a fixed effect. The lesser model provided a better fit of the data, with a lower AIC (-6389.2) value and significantly lower deviance (-6399.2, $\chi^2(21, \mathrm{N} = 33) = 49.066$, $p = 0.00049$). (See 2 for regression results.) We also compared both the fuller and the lesser models with random slopes in the random effects term, but the lesser model with only varying random intercepts in the random effects term still proved a better fit for observed data (see Table 1. Model output the random slopes fuller model not shown; lesser vs. fuller random slopes models deviance was -6624.3 and -6678.3, respectively; $\chi^2(21,\ \mathrm{N} = 33) = 54.025$, p ¡ 0.0001). Thus, we conducted further analysis on results of the lesser model.

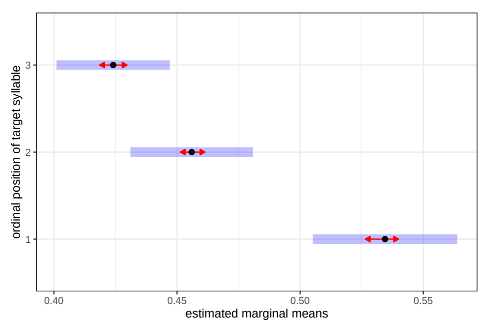
We found that reaction times are modulated by ordinal position such that RTs to word-initial syllables are notably slower than those to word-medial and word-final syllables. (**??**) We conducted pairwise comparisons on estimated marginal means for levels of the factor position with Tukey adjustment, to explore the drop in reaction times between each positions (See 1). The predicted coefficients from the winning model accurately reflected the pattern of graded reaction times in the raw data (**??**).



(a) Ordinal Position of target syllable predicts reaction time.



(b) Averaged over blocks.



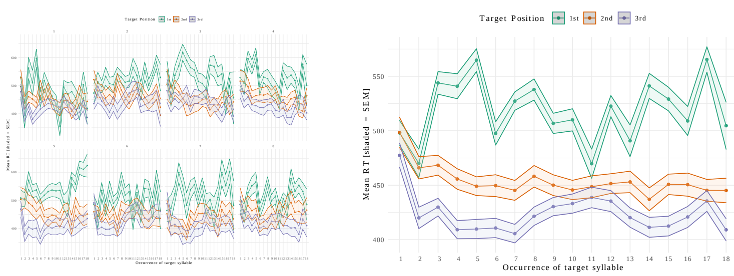
(c) Contrast of estimated marginal means.

To improve interpretability of the results, we also computed s-values for each p-value. The s-values is a metric of self-information, or surprisal information measure. It can assist in the interpretation of p-values by providing an intuitive transformation of the p-value, taking the negative log of the p-value: -$\log_2(p)$. A value of 0 ($p = 1$) is perfectly unsurprisingly and surprisal increases exponentially as s approaches zero. Thus, S becomes a measure of information in bits against the null hypothesis. The difference in means between positions 1 and 3 can be quantified in roughly 139.3 bits. If the null hypothesis of no difference in means is true, this result is as surprising as getting all heads in 140 fair coin tosses (s rounded to the nearest integer).

Table 1: Contrasts for ordinal positions of target syllables.

contrast	estimated marginal diff.	SE	df	z ratio	p value	s value
1 - 2	0.078	0.006	Inf	13.770	<.001	139.307
1 - 3	0.110	0.006	Inf	18.964	<.001	262.422
2 - 3	0.032	0.004	Inf	7.285	<.001	39.918

Furthermore, this pattern of reaction times emerged within the first few presentations of the target syllable, and remained stable throughout the rest of the experimental blocks (**??**)). This observation accounts for why factor block did not contribute much to model fit; reaction times differentiate early on and exhibit little change thereafter.



(a) Rolling means of reaction times in each block.

(b) Grand averaged rolling means.

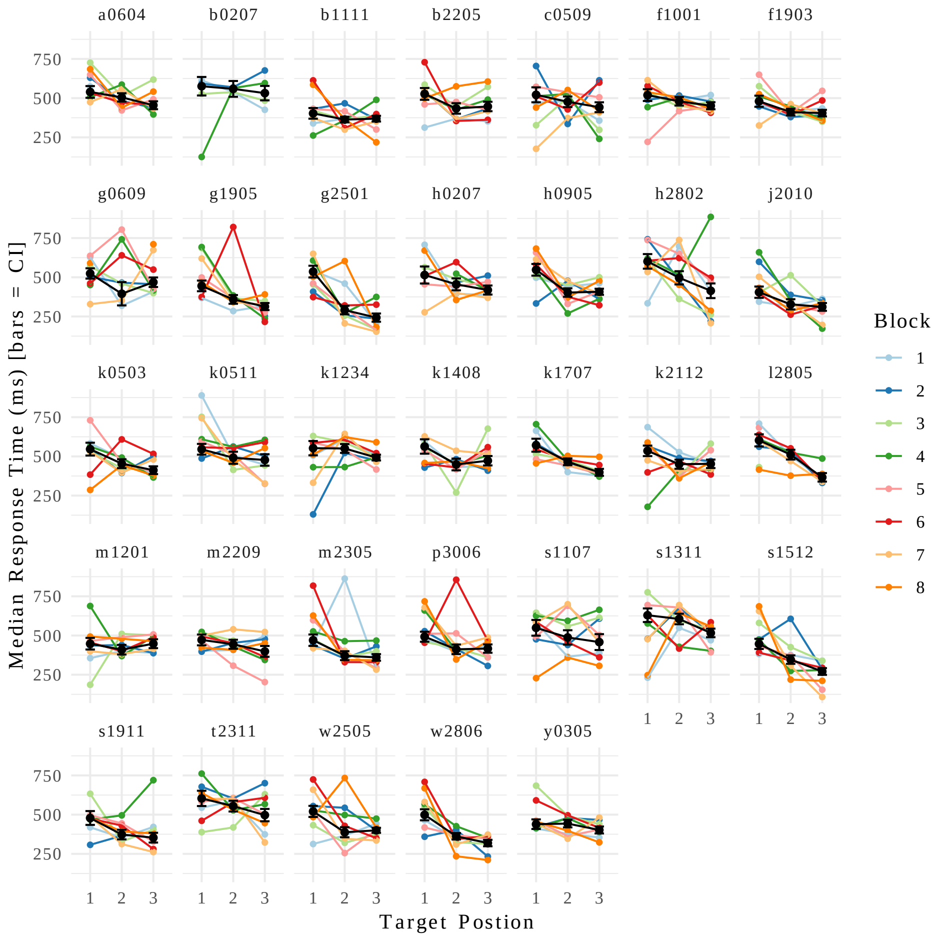


Figure 3: Reaction times for each position in each participant.

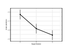
3 Experiment 2

3.1 Method

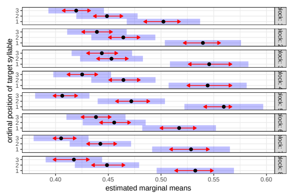
3.2 Results

4 Supplementary Materials

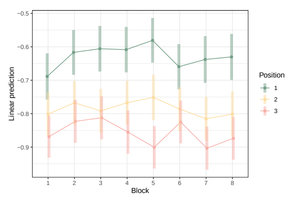
We quantified accuracy in the target detection task to ensure participants complied with task instructions. The hit rate in Experiment 1 was 0.70



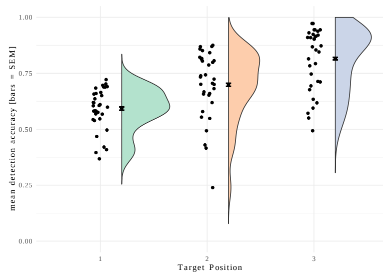
(a) Predictions from linear model, with only position as predictor.



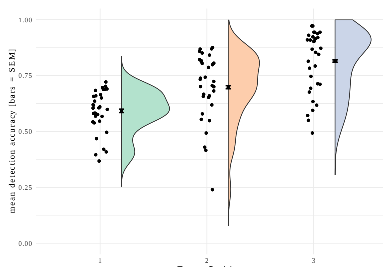
(b) Estimated means and contrasts from linear model with position and block as predictors.



(c) Predictions from linear model, with position and block as predictors.



(a) Accuracy for Experiment 1



(b) Accuracy for Experiment 2

Table 2: GLM Results

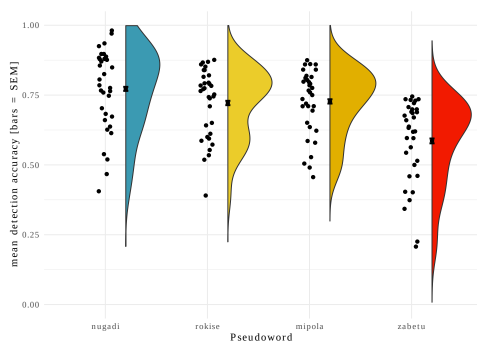
	Dependent variable:		
	reaction time (s)		
	lesser	lesser (random slopes)	fuller
	(1)	(2)	(3)
Intercept(Pos 1)	-0.63^{***} $(-0.68, -0.57)$	-0.64^{***} $(-0.68, -0.61)$	-0.69^{***} $(-0.76, -0.62)$
Pos 2	-0.16^{***} $(-0.18, -0.14)$	-0.14^{***} $(-0.19, -0.10)$	-0.11^{***} $(-0.17, -0.05)$
Pos 3	-0.23^{***} $(-0.25, -0.21)$	-0.22^{***} $(-0.28, -0.15)$	-0.18^{***} $(-0.24, -0.12)$
Block 2			0.07^{**} $(0.01, 0.13)$
Block 3			0.08^{***} $(0.02, 0.15)$
Block 4			0.08^{**} $(0.02, 0.14)$
Block 5			0.11^{***} $(0.05, 0.17)$
Block 6			0.03 $(-0.03, 0.09)$
Block 7			0.05 $(-0.01, 0.11)$
Block 8			0.06^{*} $(-0.005, 0.12)$
Pos 2:Block 2			-0.04 $(-0.12, 0.04)$
Pos 3:Block 2			-0.03 $(-0.11, 0.05)$
Pos 2:Block 3			-0.07^{*} $(-0.16, 0.01)$
Pos 3:Block 3			-0.03 $(-0.11, 0.05)$
Pos 2:Block 4			-0.05 $(-0.13, 0.04)$
Pos 3:Block 4			-0.07 $(-0.15, 0.01)$
Pos 2:Block 5			-0.06 $(-0.14, 0.03)$
Pos 3:Block 5			-0.14^{***} $(-0.22, -0.06)$
Pos 2:Block 6			-0.01 $(-0.10, 0.07)$
Pos 3:Block 6			0.01 $(-0.06, 0.09)$
Pos 2:Block 7			-0.06 $(-0.15, 0.02)$
Pos 3:Block 7			-0.09^{**} $(-0.17, -0.004)$
Pos 2:Block 8			-0.06 $(-0.14, 0.03)$
Pos 3:Block 8			-0.06 $(-0.14, 0.02)$
Fixed Effects	*Subject*	*Position|Subject*	*Subject*
Fixed Effects Struct.	*Rand.Int.*	*Rand.Int., Slope*	*RandInt.*
Observations	9,531	9,531	9,531
Log Likelihood	3,199.61	3,312.13	3,224.15
Akaike Inf. Crit.	-6,389.23	-6,604.26	-6,396.29
Bayesian Inf. Crit.	-6,353.42	-6,532.64	-6,210.07

Note: *p<0.1; **p<0.05; ***p<0.01
Fitted using Gamma distribution and log link function.

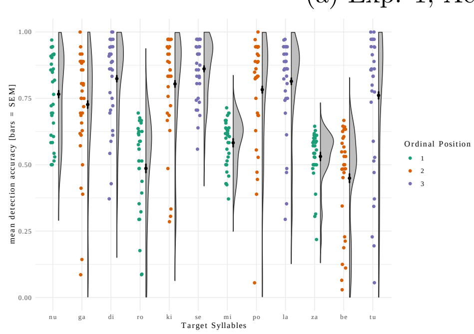
(sd = 0.45). The hit rate was also modulated by ordinal position, with each successive position having a higher mean accuracy ($M\_Pos\ 1 = 0.59$, $sd\_Pos\ 1 = 0.49$; $M\_Pos\ 2 = 0.70$, $sd\_Pos\ 2 = 0.46$; $M\_Pos\ 3 = 0.82$, $sd\_Pos\ 3 = 0.39$) (**??**(a)). When averaging across all syllables in a pseudoword, accuracy varied between the four words. This effect appears to be driven by differences in recognizibility of individual syllables; certain CV syllable pairs may have been easier to detect than others, due to minor variations in stimuli acoustics (**??**).

References

Dehaene, Stanislas et al. (Oct. 2015). "The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees". In: *Neuron* 88.1, pp. 2–19. ISSN: 0896-6273. DOI: `10.1016/`



(a) Exp. 1, Accuracy by Word



(b) Exp. 1, Accuracy by Syllable

j.neuron.2015.09.019. URL: http://dx.doi.org/10.1016/j.neuron.2015.09.019%20https://linkinghub.elsevier.com/retrieve/pii/S089662731500776X.

Pelucchi, Bruna, Jessica F. Hay, and Jenny R. Saffran (Nov. 2009). "Learning in reverse: Eight-month-old infants track backward transitional probabilities". In: *Cognition* 113.2, pp. 244–247. ISSN: 0010-0277. DOI: 10.1016/J.COGNITION.2009.07.011. URL: https://www.sciencedirect.com/science/article/pii/S0010027709001826.

Richard N. Aslin, Jenny R. Saffran, and Elissa L. Newport (1998). "Computation of Conditional Probabilities by Infants". In: *Psychol. Sci.* 9.4, pp. 321–324. URL: https://journals.sagepub.com/doi/pdf/10.1111/1467-9280.00063.

Turk-Browne, Nicholas B., Justin A. Jungé, and Brian J. Scholl (2005). "The Automaticity of Visual Statistical Learning". In: *J. Exp. Psychol. Gen.* 134.4, pp. 552–564. ISSN: 1939-2222. DOI: 10.1037/0096-3445.134.4.552. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0096-3445.134.4.552.

Turk-Browne, Nicholas B., Brian J Scholl, et al. (2008). "Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness". In: *J. Cogn. Neurosci.* 21.10, pp. 1934–1945. URL: www.omniglot.com..