Linking the neural basis of hierarchical prediction with statistical learning: The paradox of attention

Preprint · May 2021

DOI: 10.31234/osf.io/8rtsc

4 authors:

Julie M. Schneider
Louisiana State University

Yi-Lun Weng
University of Delaware

Anqi Hu
University of Delaware

Zhenghan Qi
Northeastern University

Linking the neural basis of hierarchical prediction with statistical learning: The paradox of attention

Julie M. Schneider[1,2], Yi-Lun Weng[1], Anqi Hu[1] & Zhenghan Qi[1]

[1] University of Delaware
Department of Linguistics and Cognitive Science
125 E Main St
Newark, DE 19711
United States

[2] Louisiana State University
Department of Communication Sciences and Disorders
217 Thomas Boyd Hall
Baton Rouge, LA 70803

Corresponding Author:
Zhenghan Qi, Ph.D.
Linguistics and Cognitive Sciences
University of Delaware
125 E. Main St.
Newark, DE 19716
Phone: (302) 831-3851
Email: zqi@udel.edu

Abstract

Statistical learning, the process of tracking distributional information and discovering embedded patterns, is traditionally regarded as a form of implicit learning. However, recent studies proposed that both implicit (attention-independent) and explicit (attention-dependent) learning systems are involved in statistical learning. To understand the role of attention in statistical learning, the current study investigates the cortical processing of prediction errors in speech based on either local or global distributional information. We then ask how these cortical responses relate to statistical learning behavior in a word segmentation task. We found ERP evidence of pre-attentive processing of both the local (mismatching negativity) and global distributional information (late discriminative negativity). However, as speech elements became less frequent and more surprising, some participants showed an involuntary attentional shift, reflected in a P3a response. Individuals who displayed attentive neural tracking of distributional information showed faster learning in a speech statistical learning task. These results provide important neural evidence elucidating the facilitatory role of attention in statistical learning.

Highlights:

- The current study elucidates the role of attention in statistical learning
- Monitoring of local probabilities engaged an automatic pre-attentive process (MMN)
- Global distributional information modulated expectation to the local probability
- A larger P3a response was associated with faster online learning in a word segmentation task
- Involuntary attention to statistical information facilitates statistical learning

Introduction

Humans are equipped with remarkable sensitivity to frequencies, regularities, and variabilities of the information encountered in their environments. The process of detecting statistical patterns embedded in the input, known as statistical learning (SL), is posited as a foundational theoretical account for a wide range of complex human behaviors, including language acquisition, category learning, and perception (Aslin & Newport, 2014; Erickson & Thiessen, 2015; Frost, Armstrong, Siegelman, & Christiansen, 2015). In a typical SL paradigm, people learn arbitrary associations between stimuli through passive exposure to streams of information embedded with certain statistical patterns (Fiser & Aslin, 2001; Saffran, Aslin, & Newport, 1996). After a brief period of exposure, participants can correctly identify which patterns were repeatedly presented during the exposure phase without conscious awareness of this knowledge. Successful implicit learning of statistical patterns has been demonstrated with a wide range of stimuli such as visual sequences (e.g., Fiser & Aslin, 2002), auditory tone sequences (e.g., Saffran, Johnson, Aslin, & Newport, 1999), and audiovisual sequences (e.g., Seitz, Kim, Wassenhove, & Shams, 2007) in infants, older children, and adults.

An important feature of SL is that it can occur in the absence of instruction or conscious attention to underlying patterns, such as when stimuli are presented passively without an explicit task (Fiser & Aslin, 2001; Fiser & Aslin, 2002; Saffran et al., 1999) or when participants are engaged in an unrelated cover task (Saffran et al., 1997; Turk-Browne, Jungé, & Scholl, 2005; Turk-Browne, Scholl, Chun, & Johnson, 2009). However, recent studies have found that sufficient attentional resources available for particular stimuli appear to be necessary for successful statistical learning, measured both with online response time and offline recognition accuracy (Toro, Sinnett, & Soto-Faraco, 2005; Turk-Browne et al., 2005), even without explicit awareness of the acquired statistical patterns. For example, Toro and colleagues (2005) manipulated the task demand during a speech SL task. Half of the participants were asked to passively listen to the speech stream, while the other half were asked to perform a concurrent task. When the concurrent task was more demanding or within the same sensory modality as the speech stream, participants were incapable of identifying the words following the embedded pattern, indicating that at least some attentional resources must be available and directed to the speech stream for successful word segmentation. Similarly, Turk-Browne, Junge, and Scholl (2005) found that visual SL outcome is modulated by selective attention using familiarity judgement accuracy and response time measures. These results indicate that both implicit (i.e., attention-independent) and explicit learning (i.e., attention-dependent) systems operate alongside each other during SL (Batterink, Reber, Neville, & Paller, 2015). However, most evidence supporting the importance of attention in SL comes from reflection-based tests of learning, where participants are required to reflect on what they have learned and make explicit decisions about test stimuli. These reflection-based tests are problematic as they introduce additional noise across individuals (Siegelman, Bogaerts, & Frost, 2017) and thus underestimate the effects of learning in SL studies relative to processing-based measures (e.g., Batterink et al., 2015; Vuong, Meyer, Christiansen, & Kau, 2016). These behavioral measures are also problematic as they represent the final result of a complex chain of multiple operations that may include perceptual/cognitive mechanisms, response selection, and motor preparation and execution (Daltrozzo & Conway, 2014; Daltrozzo et al., 2017).

Neurophysiological measurements offer a great opportunity to disentangle the respective correlates of attention and SL without influence of conscious response. Generation of expectations from structured patterns is inherent to SL and is particularly important for learning information that is distributed temporally (Conway, 2020). Prediction has been primarily studied in oddball paradigms, where expectations are built on repetitions of stimuli, or the relative frequency of exemplars in the input. Decades of research has documented the neural responses to violations of auditory expectations, a mismatch negativity response (MMN) between 100 and 250 ms after the stimulus onset followed by a later neural response labeled as the P3 complex. The MMN reflects a pre-attentive, nonconscious response to an auditory stimulus deviance occurring in the concurrent and local context (Chennu et al., 2013; Risto Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001; Paavilainen, 2013), where explicit attention to the auditory stimuli is not required (Alho, Woods, & Algazi, 1994; Van Zuijen, Simoens, Paavilainen, Näätänen, & Tervaniemi, 2006). Importantly, the MMN appears to be only sensitive to violations of expectation based on transient auditory experiences (local information), but not to violations of auditory prediction that accrues over time (global information; (Pegado et al., 2010; Wacongne et al., 2011). The transient nature of the MMN is further supported by findings that MMN responses were drastically reduced when the presentation rate was slowed down (Pegado et al., 2010). From these studies, it is evident that the MMN is a marker of perceptual auditory prediction, or learning from transient sensory inputs; however, the MMN amplitude can be modulated by attention when it is explicitly directed towards the global pattern.

In contrast, the P3 was found to be modulated by mnemonic prediction, a process based on accumulated information perceived over a longer period of time (Bekinschtein et al., 2009a; Chennu et al., 2013; Squires, Squires, & Hillyard, 1975; Sutton, Braren, Zubin, & John, 1965). Crucially, the P3 has been associated with an attentional shift to stimulus changes, which relies on continuous updating of one's memory trace (Bekinschtein et al., 2009a; Donchin & Coles, 1988; Sergent, Baillet, & Dehaene, 2005). It has been found that an explicit search for auditory "oddballs" of global statistical patterns leads to a protracted P3 response, whereas an explicit search for auditory “oddballs” of local statistical patterns leads to shorter and sharper P3 responses (Chennu et al., 2013; Bekinschtein et al., 2009). Importantly, when no attention is required during the task, or when participants are engaged in a cover task, no P3 effect is elicited by the global pattern violations (Wacongne et al., 2011; Bekinschtein et al., 2009). Therefore, the P3 reflects a conscious, top-down attentional shift to the prediction errors derived from the regularities distributed over an extended period of time (Van Zuijen et al., 2006). While the P3 is the most common nomenclature used to refer to this late positive component, the P3 effects reported in these studies are typical P3b responses, distributed across temporal-parietal scalp sites. Alternatively, the P3a is a more frontally distributed component. These effects reflect task-related explicit control of attention (Polich, 2007).

One caveat of these studies is that the global information differed from the local information both in the temporal scale of its distribution (minutes vs. seconds) and in the complexity of the statistical pattern. The global probability is manipulated on the level of conditional statistical information, that is how likely a multi-tone phrase is to occur over an extended period of time, while information related to the local probability is distributional statistical information, that is how likely a particular tone occurred in the local context. Evidence from behavioral work suggests independent learning processes underlie conditional and distributional statistical learning (Endress & Bonatti, 2007; but see Thiessen, Kronstein, & Hufnagle, 2013 for an alternative account). Therefore, the context by which attention is required remains opaque: is it the gradual nature of learning, the complexity of the statistical pattern, or both? Regardless, these findings suggest that the attention-dependent system plays a key role in conditional statistical learning, where learning depends on the gradual accumulation of co-occurring information over a protracted period of time. Yet, the vast literature in implicit statistical learning has presented paramount evidence suggesting learning can be achieved without explicit attention towards the stimuli.

The current study aims to address this paradox by manipulating the temporal scale of distributional information. We ask whether processing local and global distributional information involves a pre-attentive (MMN) or an attentional process (P3), and how neural prediction based on distributional information is related to higher-level conditional statistical learning?

Our research addresses the following gaps in the literature. First, existing evidence for the dissociation between the MMN and P3 during processing of statistical information rely on designs which confound two dimensions of hierarchy: the complexity of the statistical regularity (distributional vs. conditional) and the temporal delay of the distribution (local vs. global). As a result, it is not clear whether the P3 effect represents the necessity of attention for learning conditional statistical patterns, learning patterns distributed over a longer period of time, or both. In our study, we remove the confound by only focusing on the temporal delay of the distribution (local vs. global). We ask how the MMN and P3 are modulated by graded probability of occurrence. Second, previous studies demonstrated the necessity of attention when learning regularities distributed across an extended period of time by contrasting tasks with and without an explicit target-detection component. In this study we sought to investigate the role of involuntary attention in a *passive* oddball paradigm with speech stimuli. Specifically, we ask whether the P3a, a fronto-parietal distributed ERP component which reflects involuntary attention shift to highly salient stimulus (Polich, 2007), can be modulated by local or global distributional statistical information. This approach is more ecologically valid, as a passive paradigm will be more suitable for future studies involving developmental comparison or special populations. Finally, current study is the first to examine the relationship between neural prediction and statistical learning. We ask whether attention-dependent or attention-independent neural prediction relates to statistical learning outcomes. We relate individual differences in a word segmentation task measured by both online and offline learning measures with the neural indices of prediction measured by ERPs.

2. Methods

2.1 Participants

Forty-five adults (mean age = 22.76 years, *SD* = 3.02 years, range = 18.1—34.6 years, 11 males) participated in this study. All participants were right-handed, native English speakers, with no history of neurological or psychiatric disorders, or brain damage. All had average or above-average non-verbal intelligence (age-based standard score > 85) as measured by the Matrices subtest (mean = 107.84, *SD* = 12.89) of the Kaufman Brief Intelligence Test (KBIT-2; Kaufman & Kaufman, 2004), and had average or above average vocabulary as measured by the Picture Vocabulary Test (PVT; mean = 114.2, SD = 13.35, range = 91-149) of the NIH Toolbox (Gershon et al., 2013). All participants gave written consent to participate, in accordance with the Institutional Review Board at the University of Delaware. All participants were compensated for their participation.

2.2 Statistical Learning Behavioral Task

Participants completed an auditory linguistic SL task hosted on a web-based platform (Qi et al., 2019; Schneider, Hu, Legault, & Qi, 2020) programmed using jsPsych (de Leeuw, 2015). The stimuli, modeled after Saffran, Aslin, & Newport (1996), were composed of a set of four triplets, each containing three syllable stimuli which always appeared in the same order. The four triplets were *pa-bi-ku, da-ro-pi, ti-bu-do,* and *go-la-tu*, were generated by a speech synthesizer in a monotone female voice. Each triplet was repeated 48 times and was concatenated in a pseudorandom order so that the same triplet did not repeat more than twice across three consecutive triplets. The stimulus onset asynchrony was 480 ms and each triplet lasted 1440 ms. The exposure phase had a total duration of 4 minutes and 36 seconds.

During the exposure phase, participants were told to track a target syllable (i.e., the alien's favorite word) by pressing a button on the keyboard while listening to a continuous stream of speech syllables. Participants were randomly assigned to track one of the four syllables in the third position of a triplet throughout the exposure phase. For example, the participant may be asked to track *do,* which appeared only in the sequence *ti-bu-do*. They were instructed to press the spacebar on their computer whenever they heard this target syllable. Reaction time (RT) was recorded for the participant's response to the target.

A two-alternative forced-choice task consisting of 32 test trials immediately followed the exposure phase. In each trial, the participant heard a triplet word from the exposure phase and a foil nonword. Foil nonwords were comprised of three stimuli from across triplets, where each syllable maintained its position in the word but never co-occurred during the exposure phase (*pa-ro-do, ti-la-pi, da-bi-tu, go-bu-ku*). Each trial prompted the user to identify which of the two sequences had been heard during the exposure phase.

The linear slope of the RT across trials was calculated, where negative RT slope indicates acceleration of response. Keypresses from trials immediately preceding or following the target were included. More specifically, to be considered a valid response to the target, the keypress must be in the time window of one stimulus before and one stimulus after the target stimulus, that is -480 ms to +960 ms relative to the onset of the target. A keypress prior to the target is considered as anticipation and thus yields a negative reaction time. To control for baseline RT differences across individuals, the raw RTs of an individual were first transformed into z-scores. The z-normed RTs of an individual were then entered as the dependent variable into a linear regression model.

The target trial order (1 to 48) was entered as the independent variable. The reaction time slope was computed as the slope of the linear regression line (beta coefficient). Test accuracy was calculated as the percentage of correct trials during the two-alternative forced-choice test phase. All behavioral analysis of the speech SL task was completed in R (R Core Team, 2017). Seven participants were removed from these behavioral analyses: six participants did not complete the task and one participant did not have enough valid key presses (< 6 trials) during the exposure phase. Therefore, all behavioral and brain-behavior correlational analyses were conducted on the remaining 38 participants (mean age = 20.55 years, *SD* = 2.53 years, range = 18.0—31.0 years, 11 males).

2.3 Auditory EEG Oddball Paradigm

2.3.1 Stimuli. Two female English native speakers produced sounds of "bog" and "dog" 100 times for each word in a picture-naming task. 100 tokens for each word were digitally recorded using a SHURE SM58 microphone and Edirol UA-25EX sound card, sampling at 44.1 kHz. 50 tokens with better recording quality were chosen for the experiment. The */ba/* and */da/* sound were manually cut from the original recordings. The duration of each sound file is 180 ms with ramping at the beginning and the end of the syllable. The intensity of each sound was normalized to 70dB.

Two auditory streams of 1500 stimuli (SOA = 0.7 seconds) were created consisting of three conditions: standard, linguistic deviant and non-linguistic deviant (see Figure 1 for visualization of EEG paradigm). The standard condition included repeated presentations of the /ba/ syllable spoken by one female speaker. The linguistic deviant was a different syllable, /da/, spoken by the same speaker as in the standard condition. The non-linguistic deviant was the same syllable, /ba/, spoken by a different female speaker. In order to investigate the neural responses associated with differences of abstract features of the stimuli as opposed to merely acoustic differences, we included 50 different exemplars for each stimulus type.

To investigate listeners' sensitivity to global probability, we manipulated the frequency of deviant presentation across two experimental blocks. In both blocks, standard stimuli were presented 1200 times, resulting in a global probability of 0.8 (1200 standards /1500 total stimuli). In one block, the linguistic deviant occurred at a high frequency (global probability = 0.13; 200 deviants /1500 total stimuli), while the non-linguistic deviant occurred at a low frequency (global probability = 0.06; 100 deviants /1500 total stimuli). In the other block, the non-linguistic deviant occurred at a high frequency (global probability = 0.13), while the linguistic deviant occurred at a low frequency (probability = 0.06). The block order was counterbalanced across participants.

To investigate listeners' sensitivity to local probability, we manipulated the number of standard stimuli preceding deviant stimuli within each global probability condition. In the low condition, two standard stimuli were presented before a linguistic or non-linguistic deviant (i.e. /ba/ /ba/ /deviant). In the high condition six standard stimuli were presented before a linguistic or non-linguistic deviant (i.e. /ba/ /ba/ /ba/ /ba/ /ba/ /ba/ /deviant/). Low and high conditions were randomly interspersed in each auditory stream. For each deviant type within a block, one third of the deviants were in the local high condition. One third of the deviants were in the local low condition. The rest of the deviants were preceded by four standard stimuli. To maximize the analysis power, we only analyzed the difference between the high and low local probability conditions. The local probability conditions were randomly ordered within each auditory stream.

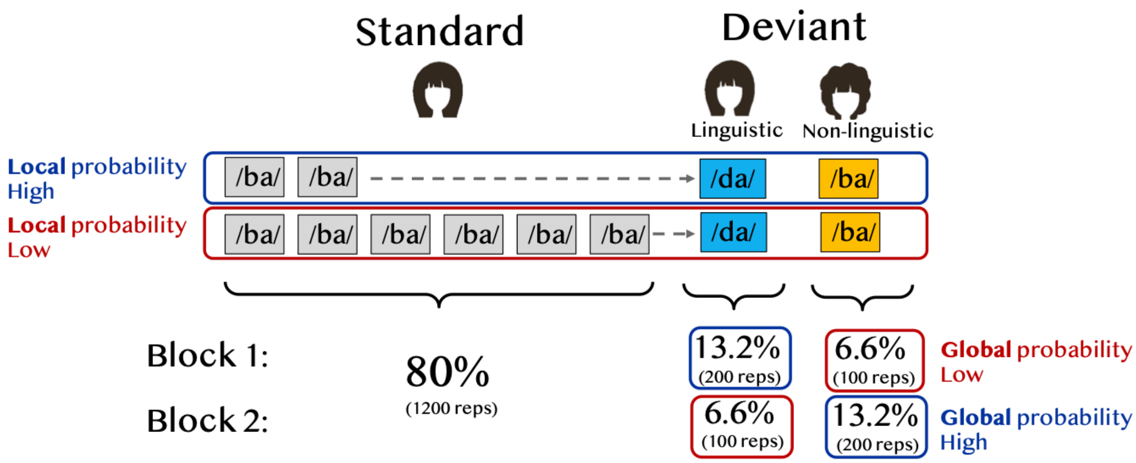

Figure 1. *Schematic illustration of the EEG Paradigm*. The global probability was manipulated across the two blocks. In the first block there were greater repetitions (reps) of the linguistic deviant, as compared to the non-linguistic deviant, and this pattern switched in the second block. The local probability was manipulated within each block. In the low condition two standard stimuli were presented before the linguistic or non-linguistic deviant and in the high condition six standard stimuli were presented before the deviant. Block order was counterbalanced across participants.

2.3.2 Procedure. Participants were instructed to watch a silent animation movie while listening to the auditory streams through a pair of noise-attenuating Cortech ER-2 earphones. Participants performed a visual target detection cover task and were instructed to press a button on a Cedrus Response Pad RB-840 (Cedrus Corporation, San Pedro, CA) as quickly as possible upon seeing a target animation character in the movie. Each experimental block lasted for 17.5 minutes. All the visual and auditory stimuli were presented using Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com).

2.4 EEG Preprocessing

EEG was recorded with a 24-channel mobile EEG system (SMARTING, mBrainTrain, Belgrade, Serbia) which features a sampling rate of 500 Hz, a resolution of 24 bits, and a bandwidth from DC to 250 Hz (SMARTING, www.mbraintrain.com). The amplifier used in this study includes a 3D gyroscope and power supply for several hours use (weight 64 grams; size 82 × 51 × 14 mm). Data is transmitted wirelessly via bluetooth (v2.1) to a nearby paired laptop.

Electrode impedances were kept below 10 kΩ. Recordings were online referenced to electrode FCz and grounded to electrode AFz. All data was saved using Labrecorder software, which is part of the Lab Streaming Layer (LSL).

All continuous data was high-pass filtered at 0.1 Hz, low-pass filtered at 30 Hz, and re-referenced to the mastoids. An Independent Components Analysis (ICA; Delorme, Makeig, & Sejnowski, 2001) was carried out for artifact removal. The components related to eye-movements or muscle activity were identified and removed from the data on the basis of their time-courses, frequency spectra and topographies. The data was then epoched from 100 msec before to 600 msec after stimulus onset and baseline corrected to the 100 msec before stimulus onset. Trials were removed from analysis if the peak-to-peak voltage between 100 ms pre-stimulus and 600 ms post-stimulus exceeded 100 µV for any of the 24 EEG channels. On average, 1.22 components (SD = .40) and 15.11 trials (SD = 12.90) were removed.

2.5 Event Related Potential Analysis

Our main interest is to examine whether the MMN and P3a are modulated by the local and global distributional information of the deviants. The time windows and electrodes of interest were defined by first examining significant differences between all standards and deviants using a cluster-based permutation test within the Mass Univariate Toolbox of Matlab (Groppe, Urbach, & Kutas, 2011). This permutation test is based on a repeated measures t-statistic using every time point at each electrode from 0 to 600 msec post-stimulus, and controls for the family-wise error rate across the full set of comparisons. The cluster-based test corrects for the multiple comparisons problem by first forming clusters of neighboring extreme t-scores and building a null hypothesis distribution from the most extreme cluster statistic (Maris & Oostenveld, 2007). For our cluster-based permutation, neighbors were established as electrodes within approximately 5.44 cm of one another with a family-wise alpha level of .05. This approach capitalizes on the fact that ERP effects are more likely than noise to extend across many adjacent electrodes and time points and is probably the most powerful mass univariate procedure for detecting broadly distributed effects (Groppe et al., 2011; Maris & Oostenveld, 2007).

For the time windows and spatial distributions identified by the omnibus cluster permutation (standards *vs*. deviants), the mean amplitude of the ERP was then averaged across the significant cluster of electrodes for each deviant condition and extracted for each participant using the ERPLAB toolbox in Matlab (Lopez-Calderon & Luck, 2014). The ERP mean amplitudes were then submitted to a linear mixed-effect modeling using the *lmer* function (lme4 version 1.1-20) in R (RStudio Team, 2016). The model included fixed effects for domain (syllable *vs* voice), local probability (high *vs*. low sequence of standards before deviant), and global probability (high *vs*. low frequency of occurrence) with by-subject random intercepts and random slopes for the interaction between local and global probability. Effect sizes are represented by Cohen's *d* and were produced using the lme.dscore function (EMAtools version 0.1.30).

3. Results

3.1 Statistical Learning Behavioral Results

Accuracy on the behavioral statistical learning task was significantly above chance (M = 60.1%, SD = 12.05%, $t(37) = 5.17$, $p < 0.001$). Reaction time slope during the exposure phase did not significantly improve across the course of the exposure window (M = .004, SD = 0.02; $t(37)$ = 0.92, p = 0.36); however, reaction time slope was marginally related to accuracy (R = -0.31, p = 0.055). Individuals who responded more quickly over the course of the exposure phase were more likely to correctly identify words over foils during the test phase.

3.2 ERP Results

We first sought to investigate the overall change in the neural response to auditory deviants in order to define the time windows and electrodes of interest for our main analyses. Therefore, we compared all deviants to all standards between 0 and 600 msec after the stimulus onset. Next, using the time windows and electrodes defined by the first step, we sought to delineate how the properties of auditory deviants (linguistic vs. nonlinguistic, high vs. low global probability, and high vs. low local probability) modulate listeners' neural responses to the deviants.

3.2.1 Deviants compared to standards

The mass univariate analysis resulted in two significant clusters. The deviants elicited a significantly greater negativity than the standards between 60 msec and 214 msec at all electrodes except for F7, T7, and Cz, and between 350 and 598 msec at these same electrodes, with the added exception of P7 (Figure 2). The early effect uncovered by the cluster permutation is similar in temporal and spatial distribution to the canonical MMN (Näätänen, Paavilainen, Rinne, & Alho, 2007). The late effect, distributed widely, shares the similar temporal characteristic as P3a, but with the opposite polarity. Such effect, termed as late discriminative negativity (LDN) or late MMN, has also been reported in passive auditory oddball tasks with speech stimuli (Wetzel & Schroger, 2014; Bishop, Hardiman, & Barry, 2011; Cheour, Korpilahti, Martynova, & Lang, 2001; Korpilahti, Lang, & Aaltonen, 1995).

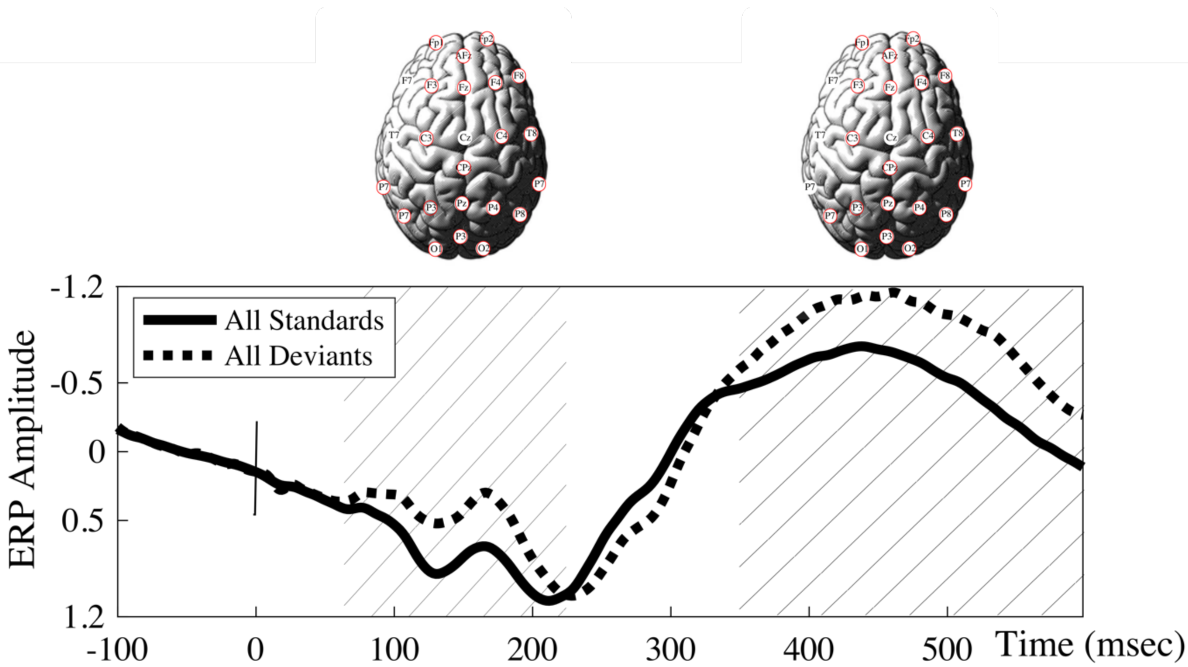

Figure 2. Waveforms recorded in response to standards and deviants. Scalpmaps represent the spatial distribution of significant differences identified by the mass univariate cluster analysis in both the early (60-214 msec) and late (350-598 msec) time windows between deviants and standards. The ERP is a temporal representation of significant differences between deviants and standards collapsed across the electrodes identified as a significant cluster by the mass univariate cluster analysis (cluster-level $p < 0.05$).

3.2.2 Contribution of domain, global probability, and local probability

In order to investigate how the neural responses to auditory deviants are affected by domain (non-linguistic *vs* linguistic), local probability (low *vs*. high), and global probability (low *vs*. high), the average amplitude in the two significant time windows for each participant was submitted to linear mixed effect models.

The linear mixed effect model in the early MMN time window (60-214 msec), shown in Table 1, revealed a main effect of local probability. Participants showed a greater MMN response to deviants with a low local probability (deviants following a longer sequence of standards) than those with a high local probability (deviants following a shorter sequence of standards; Figure 3a). The linear mixed effect model in the late time window, shown in Table 2, revealed a significant main effect of global probability (Figure 3b) and a significant interaction between local and global probability (Figure 4a). Participants showed a greater LDN response to deviants with a low global probability (deviants occurring less frequently) than those with a high global probability (deviants occurring more frequently; Figure 3b). To unpack the interaction between local and global probability, post-hoc pairwise analyses revealed that the group effect shifted from a significant LDN response to a non-significant P3a trend, as the deviants occurred in less frequent probability across global and local contexts. We found a larger and significant LDN response to global probability (low vs. high) for the Local (high) deviants ($p = 0.01$, Bonferroni-corrected $p < 0.05$), as compared to a non-significant global probability effect for the Local (low) deviants ($p = 0.33$). There was also a larger and significant LDN response to the local probability (low vs. high) for the Global (high) deviants ($p = 0.04$), as compared to a non-significant local effect for the Global (low) deviants ($p = 0.17$; see Figure 4b). The effect of domain (non-linguistic vs. linguistic) was not significant in either the early or the late time window.

Table 1. Generalized linear mixed-effects model and effect size (Cohen's *d*) of individual learners' average early ERP amplitude based on Domain, Local Probability and Global Probability.

Dependent Variables	**Predictors**	***β***	**Std. Error**	***t* value**	***p* value**	**Cohen's *d***
Average ERP Amplitude in Early Time Window	(Intercept)	0.13	0.17	0.77	0.44	--
	Domain	-0.43	0.24	-1.83	0.07	-0.23
	Global Probability	-0.34	0.24	-1.40	0.16	-0.19
	Local Probability	**-0.51**	**0.25**	**-2.07**	**0.04**	**-0.31**
	Domain x Global	0.20	0.34	0.61	0.55	0.07
	Domain x Local	0.41	0.34	1.23	0.22	0.15
	Global x Local	0.55	0.30	1.58	0.11	0.23
	Domain x Global x Local	-0.01	0.47	-0.02	0.99	-0.002

Table 2. Generalized linear mixed-effects model and effect size (Cohen's *d*) of individual learners' average late ERP amplitude based on Domain, Local Probability and Global Probability.

Dependent Variables	**Predictors**	β	**Std. Error**	***t* value**	***p* value**	**Cohen's *d***
Average ERP Amplitude in Late Time Window	(Intercept)	-0.07	0.28	-0.25	0.81	--
	Domain	0.03	0.39	0.08	0.94	0.009
	Global Probability	**-0.81**	**0.39**	**-2.10**	**0.04**	**-0.26**
	Local Probability	-0.57	0.39	-1.46	0.14	-0.19
	Domain x Global	0.35	0.55	0.65	0.52	0.08
	Domain x Local	0.18	0.55	0.34	0.74	0.04
	Global x Local	**1.17**	**0.57**	**2.05**	**0.04**	**0.27**
	Domain x Global x Local	-0.39	0.77	-0.50	0.62	-0.06

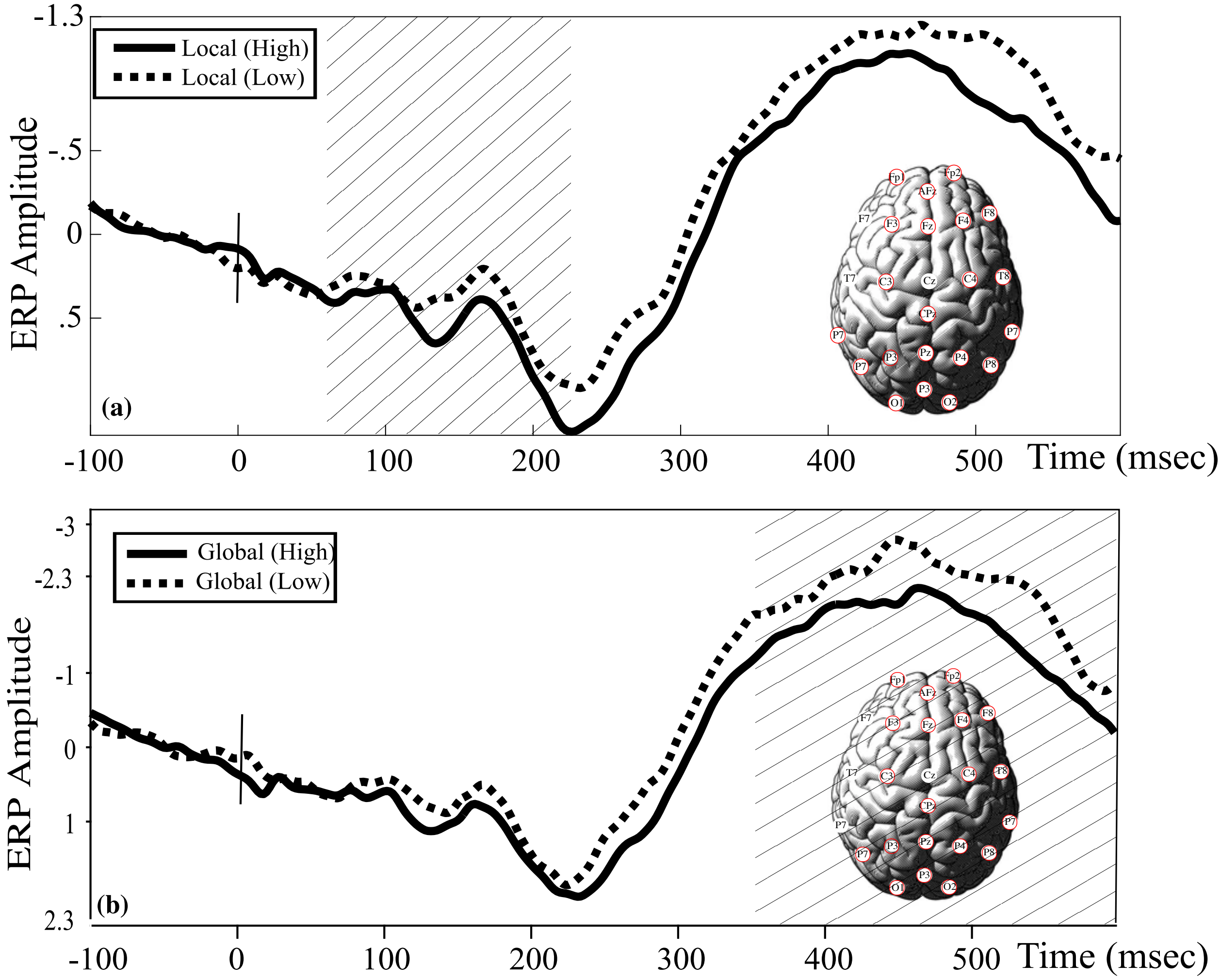

Figure 3. Main ERP effects of a) local probability in the early time window and b) global probability in the late time window. a) Deviants preceded by a longer sequence of standards, the Local (Low) condition, elicited a greater negativity from 60-214 msec (distinguished by hatching) than deviants preceded by a shorter sequence of standards, the Local (High) condition. b) Deviants that occurred less frequently, the Global (Low) condition, resulted in a greater negativity from 350-598 msec (distinguished by hatching) than deviants that occurred more frequently, the Global (High) condition. The scalpmap represents the spatial distribution of the electrodes. Electrodes highlighted in red were identified as significant in the mass univariate analysis comparing all deviants and all standards.

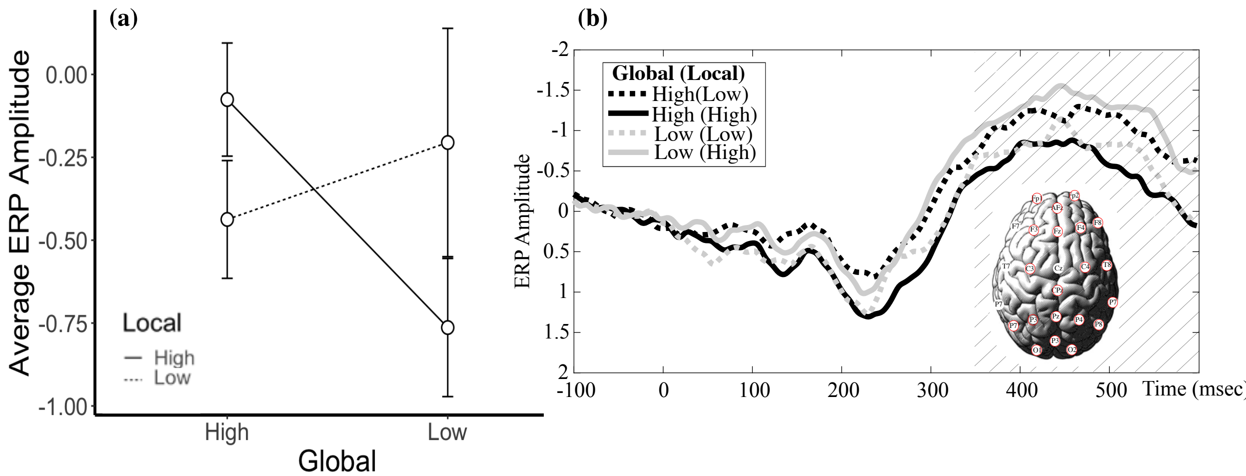

Figure 4. Interaction between local and global probability in the late time window. (a) Average ERP amplitude by local and global probability conditions. (b) ERP waveforms in each condition. The scalpmap represents the spatial distribution of the electrodes. Electrodes highlighted in red were identified as significant in the mass univariate analysis comparing all deviants and all standards in the late time window.

3.3. Correlation Results

To clarify how differences in the neural response during this auditory hierarchical prediction paradigm related to learning of conditional statistics in speech, we extracted the mean amplitude of the effects identified as significant in the early and late time windows based on the above analyses. We then conducted a Spearman correlation analysis between each main ERP effect and the two SL behavioral measures (RT slope and accuracy). We first asked if the pre-attentive neural prediction measured by the MMN in the early time window is related to statistical learning outcomes? Greater magnitude of local probability (low vs. high) is significantly associated with slower online learning, measured by RT slope (*Rho* = -0.33, *p* = 0.04). We next asked whether neural prediction in the late time window measured was related to statistical learning outcomes.

We extracted five ERP measures for each participant: one from the main global probability effect, and four from the significant interaction between the global and local probability reported earlier (the local probability effect within Global (high) and Global (low) conditions, as well as the global probability effect within Local (high) and Local (low) conditions). Among the five correlational analyses, we found one significant result: learners with a larger P3a-like response to local probability for Global (low) deviants exhibited faster RT acceleration during the SL task (*Rho* =-.42, p = 0.009, Bonferroni-corrected p = 0.05; Figure 5a; see Table 3 for all the statistics). None of these ERP markers were related to offline learning measured by accuracy.

Table 3. Spearman correlations between main ERP effects, RT Slope, and accuracy.

MMN ERP Amplitude	**RT Slope**	**Accuracy**
Local Effect	***Rho* = -0.33, *p* = 0.04**	*Rho* = 0.004, *p* = 0.98
LDN/P3a ERP Amplitude	**RT Slope**	**Accuracy**
Global Effect	*Rho* = 0.19, *p* = 0.26	*Rho* = 0.10, *p* = 0.54
Global Effect within Local (High)	*Rho* = 0.20, *p* = 0.24	*Rho* = 0.09, *p* = 0.60
Global Effect within Local (Low)	*Rho* = -0.06, *p* = 0.71	*Rho* = 0.09, *p* = 0.60
Local Effect within Global (High)	*Rho* = -0.19, *p* = 0.25	*Rho* = 0.07, *p* = 0.67
Local effect within Global (Low)	***Rho* = -0.42, *p* = 0.009**	*Rho* = 0.12, *p* = 0.49

To further illustrate the relationship between the neural responses to prediction errors and individual learners' statistical learning performance, we divided the participants based on the polarity of their local probability effect in the Global (Low) condition: 12 participants exhibited an LDN and 26 exhibited a P3a. The P3a group showed significantly faster RT acceleration during the exposure phase as compared to the LDN group ($t(36) = -8.73$, $p < .0001$; Figure 5b). However, the two groups performed similarly during the offline alternative forced-choice task ($t(36) = 0.84$, $p = .41$; Figure 5c).

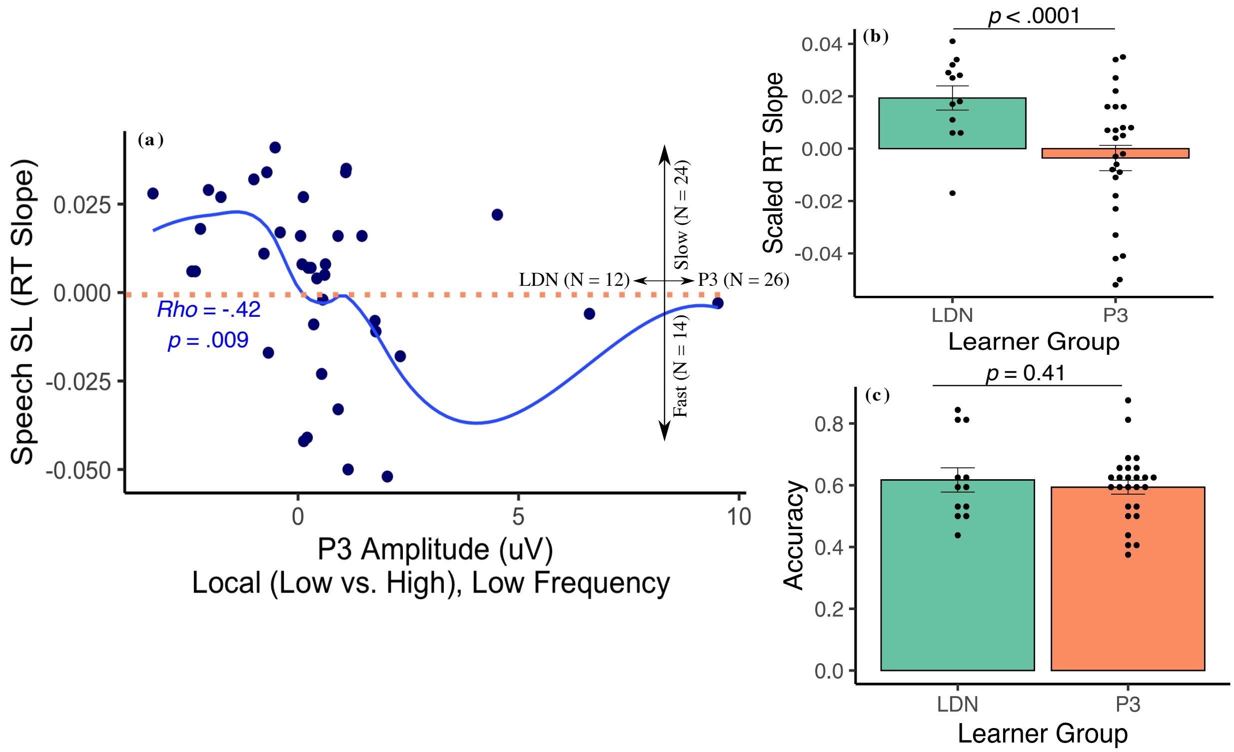

Figure 5. Greater P3a-like response to local probability is related to faster statistical learning. (a) Faster acceleration of RT (more negative slope) is associated with greater P3 response to local probabilities in the Global (Low) condition. (b) The P3a group demonstrated significantly faster online learning of conditional statistics as compared to the LDN group. (c) The P3a and LDN groups showed similar accuracy on the post-training statistical learning task.

4. Discussion

The current study sought to elucidate a paradox in implicit SL research: is an attention-dependent system critical for statistical learning or can learning be achieved without explicit attention towards the stimuli? To address this question, we designed a passive auditory EEG oddball paradigm in which we embedded two levels of distributional information in speech: low and high frequency of occurrence at the local (across neighboring syllables) and global (across the whole speech stream) levels. Our findings indicate that monitoring local probabilities engages an automatic pre-attentive process, reflected as a mismatch negativity (MMN) response, while global distributional information modulates an individual's expectation to the local probability. As auditory oddballs became less frequent and more surprising, participants exhibited a shift from a late discriminative negativity (LDN) response to a P3a response. Individuals who indicated attentive tracking of distributional information were also more likely to show faster RT acceleration in the statistical learning task.

In the early processing stage, we found that local probabilistic information modulated the MMN amplitude, suggesting neural prediction based on transient auditory inputs is a pre-attentive process, graded by the local frequency of occurrence. These results are broadly consistent with the classic view that the elicitation of a MMN represents pre-attentive encoding of short-duration memory traces (Näätänen et al., 2007). In particular, our finding replicated previous reports that increasing repetitions of proceeding standards leads to a progressive enhancement of the MMN (Baldeweg, 2007; Haeascfael, Vernon, Dwivedi, Gruzelier, & Baldeweg, 2005), but extend these findings to the speech domain. Our findings also suggest that unlike local probabilistic information, global probabilistic information was not encoded during this early and pre-attentive process. The blindness of the MMN to global patterns can be attributed to the transient nature of the memory trace represented by the MMN. Reduced MMN amplitude has been reported in response to lengthening the delay between deviants and standards (Pegado et al., 2010). Previous studies with deviants in both local and global contexts reported a lack of MMN responses to deviants that violated global patterns of conditional statistics (e.g. AAAAB (standard) vs. AAAAA (deviant)) (Bekinschtein et al., 2009; Wacongne et al., 2011). Findings from our study confirm that global probabilistic information was not processed during this early window, due to its extended timescale of memory retrieval, rather than the complexity of the pattern.

Our individual differences analyses in the early time window showed an intriguing relationship between the effect of local probability and online learning behavior during the speech SL task. Individuals who were *less* sensitive to local probabilities, as measured by a smaller MMN, showed *faster* RT acceleration during the SL task where the knowledge of transitional probability was established over an extended period of time. Does better learning of global statistical patterns reflect greater inhibition of local probabilistic processing? Our findings are consistent with the global interference effect studied in human visual and auditory perception (Bouvet, Rousset, Valdois, & Donnadieu, 2011; Navon, 1977; Poirel, Pineau, & Mellet, 2008). Across these experiments, accurate identification of the global pattern (e.g., pitch changes over the whole melody) comes at a cost to identifying the local pattern (e.g., pitch changes in a three-tone group). Likewise, hierarchical models of learning (e.g., Mathys et al., 2014) hint that local and global regularities are learned simultaneously, but the long-term updating of the memory trace required to track global probabilities has an overarching impact on updating precision associated with the local predictions being generated (Fitzgerald & Todd, 2018). Our finding might also suggest that processing of local probabilistic information hinders the detection of the global statistical patterns. Such a detrimental effect of local distributional information on global pattern learning has been reported in the context of rule-based learning. Learners failed to generalize global statistical patterns to new items when being overloaded with local distributional information, for example through an extended familiarization phase (Endress & Bonatti, 2007; Peña, Bonatti, Nespor, & Mehler, 2002). Therefore, our finding supports a competitive relationship between local and global probabilistic information, though it does not speak to the direction of the interference.

In the late time window, the presence of a global effect is consistent with previous reports that global probabilistic information is processed later (Chennu et al., 2013; Berkinstein et al., 2009; Wacongne et al., 2011; Marti, Thibault, & Dehaene, 2014). However, unlike these studies, attention was not explicitly directed towards the stimuli in our paradigm. As a result, we observed an LDN rather than a P3 response to global probabilistic information. Research in both children and adults indicates that verbal stimuli elicits a later centrally distributed negativity which reflects the automatic processing of complex auditory, possibly even linguistic, information (Cheour et al., 2001; Hill, McArthur, & Bishop, 2004; Korpilahti et al., 1995). The LDN, sometimes also termed as a late MMN (e.g., Korpilahti, Krause, Holopainen, & Lang, 2001), has been speculated to represent additional (or immature) processing of subtle features of the auditory stimulus (D. V. M. Bishop et al., 2011). Although there has not been a consensus regarding whether the LDN reflects an attention-dependent or attention-independent process (Horváth, Winkler, & Bendixen, 2008; Roeber, Berti, & Schröger, 2003), the fact that the LDN is sensitive to global probability in the context of passive oddball paradigms suggests explicit attention to the stimuli is not necessary (Bishop, 2007; Wetzel & Schröger, 2014).

The interaction between the local and the global effects in the late window provides more evidence about the role of attention in the processing of distributional information in speech. We observed a shift from an LDN to P3a response to probabilistic information as stimuli became less expected, suggesting that the global distributional pattern modulates an individual's expectation to the local stimulus. When encountering a locally rare deviant with a relatively more frequent global probability, participants consistently showed an LDN response. However, when they encountered a locally rare deviant, which is highly unexpected in the global context, the LDN effect was replaced by a P3a-like effect with large variabilities across individuals. We found two-thirds of the participants elicited a P3a response, while the other one-third of the participants elicited an LDN response. Similar shifts from an LDN to a P3a were observed for the effect of global probability. These results provide neural evidence that the mutual interference effect between global and local levels might depend on a top-down attentional process.

Individual difference analyses in the late time window revealed a facilitatory role of attention in statistical learning. Individuals with a larger P3a response, which indexes a greater involuntary attentional shift modulated by the global context, demonstrated more rapid online statistical learning. It is likely that flexible top-down control of attention towards salient prediction errors allows for more efficient learning during the statistical learning task, through prioritizing the global patterns embedded in the speech stream (Zhao, Al-Aidroos, & Turk-Browne, 2013). Although selective attention to stimuli is known to boost statistical learning (e.g. Toro, Sinnett, & Soto-Faraco, 2005; Turk-Browne, Jungé, & Scholl, 2005), our findings emphasized the relationship between endogenous attentional control and real-time statistical learning. However, attention is apparently not necessary for statistical learning. Participants, regardless of their ERP profiles (P3a or LDN), were equally successful in recognizing the familiar words. These findings align well with the view that both the attention-independent and the attention-dependent learning mechanisms are at play, both within and across individuals (Batterink et al., 2015; Conway, 2020). Walk and Conway (2016) proposed that implicit learning is sufficient for learning unimodal sequential regularities (i.e., sequential dependencies between items in the same perceptual modality) but that additional cognitive resources such as selective attention or working memory maybe required to learn cross-modal sequential patterns. From the developmental perspective, a bottom-up implicit-perceptual learning system develops early in life and encodes the surface structure of input, while a second system that is dependent on attention, develops later in life, and relies to a greater extent on top-down information to encode and represent more complex patterns (Daltrozzo & Conway, 2014). Even though endogenous attentional control is not mature in children, children show adult-like performance in speech statistical learning measured by offline recognition tasks (Raviv & Arnon, 2018; Saffran et al., 1996; Shufaniya & Arnon, 2018).

Processing-based measures, such as ERPs and reaction time, will have the potential to unveil the developmental shift of weights from automatic to attention-dependent learning mechanisms.

Taken together, the current study provides important evidence about the facilitatory role of attention in statistical learning: while the neural adaptation to local and transient distributional information is automatic, computing the global distributional information over a continuous stream of stimuli may engage involuntary attentional shift, a process possibly deployed by more successful statistical learners.

Acknowledgements:
We thank An Nguyen and Violet Kozloff for their contribution in experiment implementation and data collection. We thank Tyler Perrachione for his advice on stimuli construction and experimental design. We thank Sara Beach and Elizabeth Norton for their contribution in auditory stimuli recording. We thank Yoel Sanchez Araujo and Wendy Georgan for their contribution in the design of the web-based speech statistical learning task.

Funding: This work was supported by NARSAD Young Investigator Award (#24836) sponsored by Brain & Behavior Research Foundation and National Institute on Deafness and Other Communication Disorders (R21DC017576).

5. References

Alho, K., Woods, D. L., & Algazi, A. (1994). Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. *Psychophysiology*, *31*(5), 469–479. https://doi.org/10.1111/j.1469-8986.1994.tb01050.x

Aslin, R. N., & Newport, E. L. (2014). Distributional Language Learning: Mechanisms and Models of Category Formation. *Language Learning* , *64*(2), 86–105. https://doi.org/10.1111/lang.12074

Baldeweg, T. (2007). ERP repetition effects and mismatch negativity generation: A predictive coding perspective. *Journal of Psychophysiology*. https://doi.org/10.1027/0269-8803.21.34.204

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78. https://doi.org/10.1016/j.jml.2015.04.004

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(5), 1672–1677. https://doi.org/10.1073/pnas.0809667106

Bishop, D. (2007). Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: Where are we, and where should we be going? . *Psychological Bulletin*, *133*(4), 651–672. Retrieved from https://psycnet.apa.org/doiLanding?doi=10.1037%2F0033-2909.133.4.651

Bishop, D. V. M., Hardiman, M. J., & Barry, J. G. (2011). Is auditory discrimination mature by middle childhood? A study using time-frequency analysis of mismatch responses from 7 years to adulthood. *Developmental Science*, *14*(2), 402–416. https://doi.org/10.1111/j.1467-7687.2010.00990.x

Bouvet, L., Rousset, S., Valdois, S., & Donnadieu, S. (2011). Global precedence effect in audition and vision: Evidence for similar cognitive styles across modalities. *Acta Psychologica*, *138*(2), 329–335. https://doi.org/10.1016/j.actpsy.2011.08.004

Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibáñez, A., … Bekinschtein, T. A. (2013). Expectation and attention in hierarchical auditory prediction. *Journal of Neuroscience*, *33*(27), 11194–11205. https://doi.org/10.1523/JNEUROSCI.0114-13.2013

Cheour, M., Korpilahti, P., Martynova, O., & Lang, A.-H. (2001). Mismatch Negativity and Late Discriminative Negativity in Investigating Speech Perception and Learning in Children and Infants. *Audiology and Neuro-Otology*, *6*(1), 2–11. https://doi.org/10.1159/000046804

Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. https://doi.org/10.1016/j.neubiorev.2020.01.032

Daltrozzo, J., & Conway, C. M. (2014, June 18). Neurocognitive mechanisms of statistical-sequential learning: What do event-related potentials tell us? *Frontiers in Human Neuroscience*. Frontiers Media S. A. https://doi.org/10.3389/fnhum.2014.00437

Daltrozzo, J., Emerson, S. N., Deocampo, J., Singh, S., Freggens, M., Branum-Martin, L., & Conway, C. M. (2017). Visual statistical learning is related to natural language ability in adults: An ERP study. *Brain and Language*, *166*, 40–51. https://doi.org/10.1016/j.bandl.2016.12.005

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1-12. doi:10.3758/s13428-014-0458-y.

Delorme, A., Makeig, S., & Sejnowski, T. (2001). Automatic artifact rejection for EEG data using high-order statistics and independent component analysis. In *Proc. of the 3rd International Workshop on ICA* (Vol. 457, p. 462).

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(3), 357–374. https://doi.org/DOI: 10.1017/S0140525X00058027

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*(2), 247–299. https://doi.org/10.1016/j.cognition.2006.09.010

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*. https://doi.org/10.1016/j.dr.2015.05.002

Fiser, J, & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504.

Fiser, József, & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458.

Fitzgerald, K., & Todd, J. (2018). Hierarchical timescales of statistical learning revealed by mismatch negativity to auditory pattern deviations. *Neuropsychologia*, *120*, 25–34. https://doi.org/10.1016/j.neuropsychologia.2018.09.015

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015, March 1). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in*

Cognitive Sciences. Elsevier Ltd. https://doi.org/10.1016/j.tics.2014.12.010

Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N. A., Cook, K. F., & Nowinski, C. J. (2013). NIH toolbox for assessment of neurological and behavioral function. *Neurology*, *80*(11 Suppl 3), S2. https://doi.org/10.1212/wnl.0b013e3182872e5f

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*. Blackwell Publishing Inc. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Haeascfael, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *Journal of Neuroscience*, *25*(45), 10494–10501. https://doi.org/10.1523/JNEUROSCI.1227-05.2005

Hill, P. R., McArthur, G. M., & Bishop, D. V. M. (2004). Phonological categorization of vowels: A mismatch negativity study. *NeuroReport*, *15*(14), 2195–2199. https://doi.org/10.1097/00001756-200410050-00010

Horváth, J., Winkler, I., & Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biological Psychology*. Retrieved from https://psycnet.apa.org/record/2008-12958-001

Kaufman, A. S., & Kaufman, N. L. (2004). KBIT2 : Kaufman Brief Intelligence Test . Pearson/PsychCorp.

Korpilahti, P., Lang, H., & Aaltonen, O. (1995). Is there a late-latency mismatch negativity (MMN) component? *Electroencephalography and Clinical Neurophysiology*, *95*(4), P96. https://doi.org/10.1016/0013-4694(95)90016-g

Korpilahti, Pirjo, Krause, C. M., Holopainen, I., & Lang, A. H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language*, *76*(3), 332–339. https://doi.org/10.1006/brln.2000.2426

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*(1 APR). https://doi.org/10.3389/fnhum.2014.00213

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=17517438&site=ehost-live

Marti, S., Thibault, L., & Dehaene, S. (2014). How does the extraction of local and global auditory regularities vary with context? *PLoS ONE*, *9*(9). https://doi.org/10.1371/journal.pone.0107227

Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in Human Neuroscience*, *8*, 825. https://doi.org/10.3389/fnhum.2014.00825

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007, December). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, Risto, Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001, May 1). "Primitive intelligence" in the auditory cortex. *Trends in Neurosciences*. Elsevier. https://doi.org/10.1016/S0166-2236(00)01790-2

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. *International Journal of Psychophysiology*, *88*(2), 109–123. https://doi.org/10.1016/j.ijpsycho.2013.03.015

Pegado, F., Bekinschtein, T., Chausson, N., Dehaene, S., Cohen, L., & Naccache, L. (2010). Probing the lifetimes of auditory novelty detection processes. *Neuropsychologia*, *48*(10), 3145–3154. https://doi.org/10.1016/j.neuropsychologia.2010.06.030

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-Driven Computations in Speech Processing. *Science*, *10*, 1. Retrieved from http://search.epnet.com/login.aspx?direct=true&db=aph&an=7738404&lang=es

Poirel, N., Pineau, A., & Mellet, E. (2008). What does the nature of the stimuli tell us about the Global Precedence Effect? *Acta Psychologica*, *127*(1), 1–11. https://doi.org/10.1016/j.actpsy.2006.12.001

Polich, J. (2007, October). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*. NIH Public Access. https://doi.org/10.1016/j.clinph.2007.04.019

Qi, Z, Araujo, YS, Georgan, WC, Gabrieli, JDE, & Arciuli, J (2019): Hearing matters more than seeing: a cross-modality study of statistical learning and reading ability. *Sci Stud Read* 23: 101–115.

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: modality-based differences in the effect of age. *Developmental Science*, *21*(4), e12593. https://doi.org/10.1111/desc.12593

Roeber, U., Berti, S., & Schröger, E. (2003). Auditory distraction with different presentation rates: An event-related potential and behavioral study. *Clinical Neurophysiology*, *114*(2), 341–349. https://doi.org/10.1016/S1388-2457(02)00377-2

Saffran, J R, Aslin, R. N., & Newport, E. L. (1996a). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928.

Saffran, J R, Aslin, R. N., & Newport, E. L. (1996b). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Saffran, Jenny R, Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/http://dx.doi.org/10.1016/S0010-0277(98)00075-4

Saffran, Jenny R, Newport,', E. L., Aslin, R. N., Tunick, R. A., & Barrueco3, S. (1997). *INCIDENTAL LANGUAGE LEARNING: Listening (and Learning) out of the Comer of Your Ear*. Retrieved from https://journals.sagepub.com/doi/pdf/10.1111/j.1467-9280.1997.tb00690.x

Schneider, J. M., Hu, A., Legault, J., & Qi, Z. (2020). Measuring statistical learning across modalities and domains in school-aged children via an online platform and neuroimaging techniques. *Journal of Visualized Experiments*, *2020*(160), 1–21. https://doi.org/10.3791/61474

Seitz, A., Kim, R., Wassenhove, V. van, Perception, L. S.-, & 2007, undefined. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Journals.Sagepub.Com*, *36*(10), 1445–1453. https://doi.org/10.1068/p5843

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*(10), 1391–1400. https://doi.org/10.1038/nn1549

Shufaniya, A., & Arnon, I. (2018). Statistical Learning Is Not Age-Invariant During Childhood:

Performance Improves With Age Across Modality. *Cognitive Science*, *42*(8), 3100–3115. https://doi.org/10.1111/cogs.12692

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, *49*(2), 418–432. https://doi.org/10.3758/s13428-016-0719-z

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401. https://doi.org/10.1016/0013-4694(75)90263-1

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science (New York, N.Y.)*, *150*(3700), 1187–1188. https://doi.org/10.1126/science.150.3700.1187

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, *139*(4), 792–814. https://doi.org/10.1037/a0030801

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2). https://doi.org/10.1016/j.cognition.2005.01.006

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The Automaticity of Visual Statistical Learning. https://doi.org/10.1037/0096-3445.134.4.552

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945.

Van Zuijen, T. L., Simoens, V. L., Paavilainen, P., Näätänen, R., & Tervaniemi, M. (2006). *Implicit, Intuitive, and Explicit Knowledge of Abstract Regularities in a Sound Sequence:*

An Event-related Brain Potential Study.

Vuong, L. C., Meyer, A. S., Christiansen, M. H., & Kau, K. (2016). Concurrent Statistical Learning of Adjacent and Nonadjacent Dependencies. *Language Learning*, *66*(1). https://doi.org/10.1111/lang.12137

Wacongne, C., Labyt, E., Van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(51), 20754–20759. https://doi.org/10.1073/pnas.1117807108

Walk, A. M., & Conway, C. M. (2016). Cross-Domain Statistical–Sequential Dependencies Are Difficult to Learn. *Frontiers in Psychology*, *7*, 250. https://doi.org/10.3389/fpsyg.2016.00250

Wetzel, N., & Schröger, E. (2014). On the development of auditory distraction: A review. *PsyCh Journal*, *3*(1), 72–91. https://doi.org/10.1002/pchj.49

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, *24*(5), 667–677. https://doi.org/10.1177/0956797612460407