Title

Modeling infant cortical tracking of statistical learning in simple recurrent networks

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Xu, Qihui
Sjuls, Guro Stensby
Kalashnikova, Marina
et al.

Publication Date

2024

Peer reviewed

Modeling infant cortical tracking of statistical learning in simple recurrent networks

Qihui Xu (xu.5430@osu.edu)
Department of Psychology, 1845 Neil Ave., Columbus, OH 43210 USA

Guro S. Sjuls (guro.sjuls@gmail.com)
Deptartment of Language & Literature, Norwegian University of Science & Technology,
Dragvoll alle 6, 7049 Trondheim Norway

Marina Kalashnikova (m.kalashnikova@bcbl.eu)
BCBL. Basque Center on Cognition, Brain & Language, 69 Mikeletegi 20009 Donostia-San Sebastián. Gipuzkoa, Spain
Ikerbasque. Basque Foundation for Science, Bilbao, Spain

James S. Magnuson (j.magnuson@bcbl.eu, james.magnuson@uconn.edu)
BCBL. Basque Center on Cognition, Brain & Language, 69 Mikeletegi 20009 Donostia-San Sebastián. Gipuzkoa, Spain
Ikerbasque. Basque Foundation for Science, Bilbao, Spain
Department of Psychological Sciences, University of Connecticut, Storrs, Connecticut 06069 USA

Abstract

Consider a classic statistical learning (SL) paradigm, where participants hear an uninterrupted stream of syllables in seemingly random order. In fact, the sequence is generated by repeating 4 word-like patterns, each comprised of 3 syllables. After brief exposure, adults and infants can discriminate 'words' from the sequence from other syllable sequences ('nonwords' that did not occur in exposure). If syllables have a fixed duration (e.g., 333.3 ms), syllable rate is fixed (e.g., 3/s or 3hz) and so is word rate (e.g., 1hz). If EEG is acquired during exposure, neural phase-locking is observed, initially to the syllable rate, and gradually to the word rate. This has been interpreted as a neural index of word learning. We tested whether two models that can simulate human SL behavior could simulate neural entrainment (Simple Recurrent Networks [SRNs] or multi-layer perceptrons [MLPs, feedforward neural networks]). Both models could, although SRNs provided a better fit to correlations observed between entrainment and behavior. We also discovered that raw input sequences (even for a single syllable) have rhythmic properties that generate apparent 'entrainment' when treated like EEG signals – without learning. We discuss theoretical implications for SL and challenges for interpreting phase-locked entrainment.

Keywords: Cortical tracking; Statistical learning; Computational modeling; Child development; Neural networks

Introduction

From birth, babies are keen observers of their environment, deciphering patterns and regularities in sounds, visual features, and events. This ability to detect statistical regularities equips them with essential tools to understand language, recognize objects, and anticipate actions and events. *Statistical learning* (SL) gives them keys to unlock the vast and intricate world around them, transforming the immense and undifferentiated world into meaningful entities and events. To unlock language, children must learn to combine smaller elements (e.g., phonemes, syllables) into meaningful patterns (e.g., words). At 6-8 months old, infants show an astonishing ability to pick up on patterns in language, in an unsupervised and non-referential manner (e.g., Choi, Batterink, Black, Paller, & Werker, 2020; Saffran, Aslin, & Newport, 1996). They can distinguish potential words from a stream of sounds by learning which syllables systematically co-occur. How do infants develop such 'code-breaking' abilities so early in life? What are the behavioral and neural underpinnings that facilitate this fundamental aspect of learning?

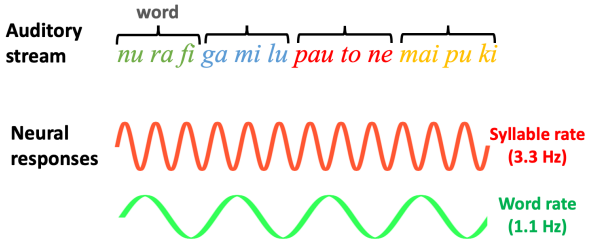

Figure 1: A schematic of cortical tracking of statistical learning (based on Choi et al., 2020). Infant subjects showed neural entrainment synchronized to the driving stimulus, i.e., the syllable rate (3.3Hz). With continued exposure, phase locking also emerged at the word rate (1.1Hz).

SL in infants is conventionally gauged through post-exposure behavioral methods such as measuring differences in looking time for 'words' vs. 'nonwords'. However, Choi et al. (2020), building on Batterink and Paller (2017), have enriched our understanding via EEG measures that appear to index learning continuously (Fig.1). Choi et al. presented syllables at a rate of 3.3Hz, making the word rate 1.1Hz (every 3 syllables). Infants showed initial neural entrainment synchronized to the driving stimulus (syllable rate of 3.3Hz), as indicated by *Inter-Trial Coherence* (ITC[1]). With continued exposure, phase locking also emerged at the word rate (1.1Hz), with the strength of word-rate ITC relative to syllable-rate

[1]See, e.g., Tallon-Baudry, Bertrand, Delpuech, and Pernier (1996). ITC is an index of phase-locked synchrony. While a Fourier analysis assesses power at different frequencies, ITC assess not just whether there is power at critical stimulus frequencies in a signal like EEG but whether the neural signal is syncrhonized (phase-locked) with critical stimulus frequencies.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

ITC increasing with more exposure. The emergence of word-rate ITC is interpreted as a neural index of SL. This inference is bolstered by correlations between word-rate ITC and beahvior: better learners show stronger word-rate ITC.

To what degree do these neural measures predict actual learning? A recent review of studies investigating neural entrainment to statistical patterns in structured speech streams revealed mixed results: only a little over half identified a meaningful link between entrainment and subsequent learning (Sjuls, Harvei, & Vulchanova, 2023). This variability in findings has led to speculation that entrainment may be indicative of a broader auditory processing function, rather than being specific to the segmentation of continuous speech.

Here we extend computational models capable of simulating SL behavior to simulate neural entrainment, with the aim of disentangling to what extent neural entrainment reflects word learning vs. possibly simpler processing. Connectionist models, like humans, can pick up linguistic patterns through SL from data they are exposed to without explicit guidance. Artificial neural networks can attune to sequential regularities through *self-supervised* SL by predicting upcoming elements, such as the next word in a sequence (Elman, 1991, 1990). Learning algorithms that continuously adjust weighted connections between nodes, based on errors between predictions and observed inputs, allow models to gradually attune to statistical regularities. Though controversial, this *prediction principle* could guide how infants learn patterns that enable them to unlock language. While correlations between brain and behavior in adults and analogous patterns in neural network models are increasingly reported (e.g., Goldstein et al., 2022), there are few such examples at early stages of development, when cognitive systems are still maturing and operate with limited resources (for a noteable exception, see Matusevych, Schatz, Kamper, Feldman, & Goldwater, 2023).

We extend simple connectionist models[2] to attempt to simulate infant neural entrainment and behavioral outcomes in SL (Choi et al., 2020). Our aims are : (a) to evaluate two models with differing learning capacities and assess which more closely aligns with observed infant behavioral and neural data, and (b) to examine the rhythmic structure of raw input sequences and investigate if sequence features could partially or wholly drive entrainment patterns typically regarded as diagnostic of word learning *without learning*.

Models and tasks

Models

Our simulation of infant data employs Multi-Layer Perceptrons (MLPs, often labelled more simply as *feedforward networks*) and Simple Recurrent Networks (SRNs). These simple models offer developmentally plausible frameworks suitable for modeling infant learning, and afford interpretability

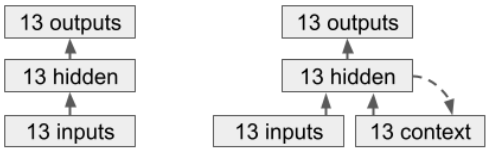

Figure 2: MLP (left) and SRN (right) schematics. Solid arrows indicate full connectivity (every node at the sending layer has a weighted connection to every node at the receiving layer). Dashed line indicates 1-1 connections that copy hidden states from the previous time step to context nodes.

that bridges computational processes with essential cognitive development principles. MLPs trained on next-element prediction can only learn contingencies between adjacent elements, whereas SRNs can also learn nonadjacent and long-distance dependencies (Elman, 1991, 1990). However, Magnuson (in preparation) reports that MLPs simulate human preferences after exposure to embedded-word sequences used by Saffran et al. (1996) and Choi et al. (2020) very well, and just as well as SRNs (although some more complex SL paradigms are not learnable by MLPs). We use both to assess whether simulating the details of neural entrainment will require the additional memory and learning capacities of SRNs.

Our MLPs have 3 fully-connected layers (input, hidden, output) with only feedforward connections. The MLP is trained using backpropagation (Rumelhart, Hinton, & Williams, 1986) based on the difference between observed and desired outputs. Our SRNs are identical, but add a context layer, which contains a copy of the hidden unit activations at the previous time step. Context nodes are fully connected to hidden nodes. Thus, the actual input to the hidden layer is both the bottom-up input (current syllable) as well as top-down information from the previous hidden unit states (which are the sum of multiplying the input by the input-to-hidden weights *and* multiplying the context by the context-to-hidden weights [and an activation function is applied to the sums], making SRNs interactive; see Magnuson & Luthra, under review). The SRN is also trained using backpropagation.

Again, the MLP is limited to adjacent contingencies, but the SRN is not. For instance, in a trisyllabic word like "ABC", MLPs can only retain information about the adjacent pairs "AB" and "BC", but not higher-order contingencies (e.g., A predicts C 2 steps later, and the pattern AB also predicts C). For simplicity and comparability, in the simulations below, both MLPs and SRNs are configured with 13 units in the input and output layers, aligning with the 13 unique syllables used in Choi et al. (2020). Following the heuristic approach of French, Addyman, and Mareschal (2011), we set the hidden size (and context size for SRNs) to the number of inputs (and did not explore the impact of different numbers of hidden nodes). Activation functions were hyperbolic tangent (*tanh*) for hidden layers and softmax for outputs.

Data representation The training and testing stimuli were based on Choi et al. (2020). Training stimuli were 4 tri-

[2]For analytic tractability, we use simple networks capable of simulating SL rather than cutting-edge large language models (Contreras Kallens, Kristensen-McLachlan, & Christiansen, 2023).

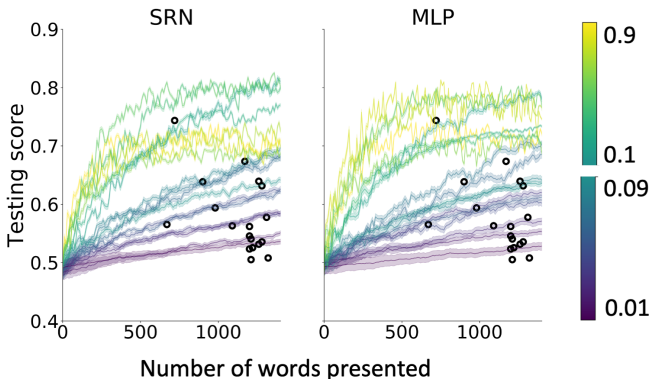

Figure 3: Matching models to infants. Black ovals indicate individual infant data points (total number of syllables presented to one infant in Choi et al. (2020) and their performance in the behavioral test). Lines indicate performance of sets of MLPs (left) or SRNs (right) with the same learning rate (indicated by color). Error bands represent 95% confidence intervals.

syllabic pseudowords (labeled ABC, DEF, GHI, JKL, where each letter stands for a unique syllable; the original items were patterns like /patigu/; Saffran et al., 1996). Words were randomly ordered, but the same word could not immediately repeat. TPs were 1.0 within and 0.333 between words. Testing stimuli included 2 training words (GHI, JKL) and 2 nonword sequences not used during training (BFE, CAM). These were encoded as 1-hot vectors (a localist representation where only 1 element is "hot" and set to 1 while others are set to 0). Given that Choi et al. used 13 unique syllables in total, with 12 used in training and 1 more ("M") used for testing, our input and output vectors have 13 elements (1 per syllable).

Model training All trainable weights (input-to-hidden and hidden-to-output in both MLPs and SRNs, as well as context-to-hidden in SRNs) were initialized to small random values. Both MLPs and SRNs were presented with a series of inputs corresponding to 1-hot syllable vectors corresponding to the embedded word sequence. Models were trained to predict the next syllable of the sequence using backpropagation. Initial outputs would be random, given random weight initialization. But after each input pattern, the actual output is compared to the desired output (the input pattern at the next time step). The backpropagation algorithm assigns credit and blame to each weight. Small (depending on the learning rate) weight changes proportional to each weight's contribution to error ensure that, if the same input were presented immediately, the network would get slightly closer to correct output.

A challenge in modeling Choi et al. (2020) is that infants received variable amounts of exposure (exposure was extended if an infant was not fussy; the range was 600 to more than 1300 'words'). Post-exposure discriminability of words from non-words was not strongly correlated with exposure (i.e., some infants with lower exposure performed much better than some infants with more exposure). Aspects of the ITC results in infants could depend on exposure and speed of learning. To simulate individual variability in learning, we created a large pool of models by varying learning rates (from 0.01 to 0.09 in steps of 0.01, and from 0.1 to 0.9 in steps of 0.1). For each learning rate, we created 30 randomly-initialized MLPs and SRNs (1080 in all; 2 network types x 30 initializations x 18 learning rates). We assessed behavioral performance over training. This allowed us to match individual networks to individual infants (Fig. 3) on the basis of amount of exposure and behavioral performance. For each infant, we selected 1 MLP and 1 SRN from the learning rate set that matched the infants' exposure and test performance. The crucial question is whether the matched networks collectively will simulate infant neural entrainment.

We used an incremental approach for training, updating weights after each input (stochastic gradient descent). While intuitively akin to how infants may learn in real-time (continuously and cumulatively; Thiessen, Kronstein, & Hufnagle, 2013; Siegelman, Bogaerts, Kronenfeld, & Frost, 2018), we have not compared this approach to using larger batch sizes.

Model testing Choi et al. (2020) quantified infants' word vs. nonword preferences during the test phase via preferential looking in response to 2 training words (GHI, JKL) and 2 nonwords (BFE, CAM), scoring each infant on the ratio looking time for nonwords over words plus nonwords (because they predicted and observed a novelty preference, which is common for infants in this task). We linked mean squared error (MSE) in model outputs to this behavioral measure. MSE calculates the average squared discrepancy between model predictions and the actual values, providing a measure of the model's predictive error. We scored models' performance based on the ratio of MSE for nonwords to that for words (Eq. 1). A score above .5 indicates higher error on nonwords than words. This measure should increase as a model learns.

$$\frac{MSE(nonwords)}{MSE(nonwords)+MSE(words)} \quad (1)$$

Simulating behavior

The individual performance of infants in Choi et al. (2020), marked by black ovals in Fig. 3, provides the basis for matching models to infants. From the pools of 540 MLPs and SRNs, we selected 1 MLP and 1 SRN (18 in total) demonstrating identical test performance after the same amount of input as that infant.

Simulating neural entrainment

We used hidden node activations from SRNs and MLPs to simulate neural signals. This aligns with previous work linking recurrent network hidden activations to neural activity (Frank & Yang, 2018; Martin & Doumas, 2017). We did not use output activations because, as a model learns, output activations will increasingly resemble raw input sequences (accurately predicting next syllables within words and activating equally 3 possibilities at word boundaries).

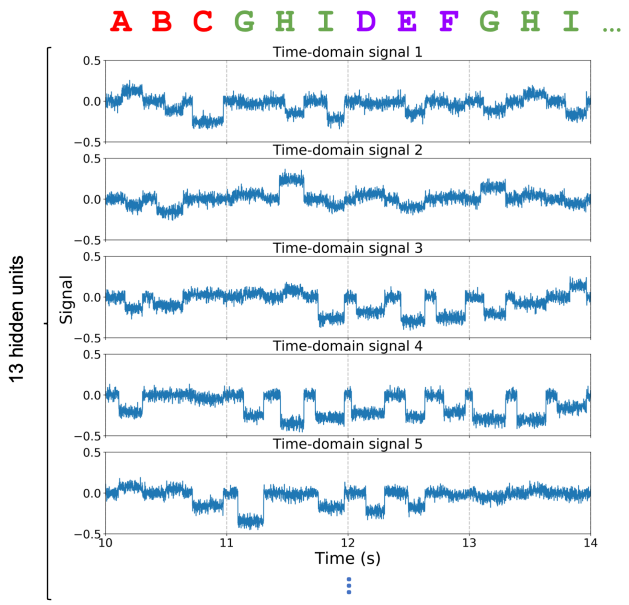

Figure 4: Transforming activations to finer-grained, noisy, EEG-like patterns (see text).

Temporal extension

To bridge the gap between discrete time updates of MLPs and SRNs (1 activation per syllable) and the continuous nature of neural oscillations (EEG is also sampled and digitized, but at a finer scale than once per syllable), we adopted a *temporal extension* method (Frank & Yang, 2018), as detailed next.

Time-domain signal transformation We fixed syllable interval in our simulations to 333 ms (rather than 300ms as in Choi et al., 2020) for simplicity, reflecting word rate of 1hz (1 every 3 syllables). To convert hidden unit activations into a finer-grained time-domain signal, we expanded 3 syllable activations into 1000 samples (going from 3 activations per 'second' to 1000 per second). We next describe further adjustments that make the time series more realistic.

Jittering and noise incorporation To introduce variability and mimic natural irregularities in speech (or latencies in neural responses to speech), we randomly jittered syllable onsets (cf. Frank & Yang, 2018). For example, rather than starting syllables A, B, and C at 0, 333, and 667 ms, temporal jittering might result in A starting at 80 ms, B at 400 ms, and C at 690 ms. To emulate inherent noise in neural activity and recordings, we added Gaussian noise to every sample in our time-domain signal (see Fig. 4).

Simulating EEG signals Our MLPs and SRNs, each with 13 units in the hidden layer, yielded a matrix of 13 by the number of samples (1000 samples per word). This gave us 13 distinct 'channels', with each dimension containing the activity over time of 1 hidden unit (Fig. 4). This matrix is analogous to 13-channel EEG data, and can be analyzed using methods similar to those applied to multi-channel EEG data.

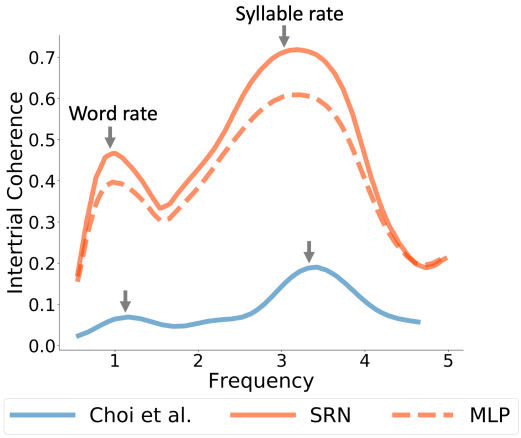

Figure 5: Mean ITC by frequency for SRNs and MLPs, compared to infants from Choi et al. (2020).

Neural entrainment across the training phase

We applied MNE-Python (Gramfort et al., 2013) procedures for measuring EEG entrainment analyses to simluated EEG time series from the models, closely following analyses Choi et al. (2020) conducted on infant EEG data. A 60-Hz notch filter and a band-pass filter from 0.5 to 20 Hz were applied to the simulated data. Next, a continuous wavelet transformation converted the time-course of each of the 13 activation channels into the frequency domain. ITC (Makeig, Debener, Onton, & Delorme, 2004), computed at both syllable (3Hz) and word rates (1Hz), allowed us to assess the phase consistency of model responses. High ITC indicates strong phase locking at the targeted frequencies. Next, these per-dimension ITC values were averaged over the 13 dimensions to obtain ITC at each frequency bin for each model.

Results ITC calculated from hidden-unit activations over training in both MLPs and SRNs exhibits distinctive peaks at word- and syllable-rate frequencies similar to those observed in infants (see Fig. 5).

Neural entrainment time-course analysis

MLPs and SRNs exhibited entrainment peaks at word and syllable frequencies akin to those observed in infants throughout the training phase, but do they follow a learning trajectory similar to that of infants? Choi et al. (2020) documented a logarithmic increase in their *Word Learning Index* (WLI) – the ratio of word-rate ITC to syllable-rate ITC – suggesting an incremental acquisition of word patterns from the input.

Following Choi et al. (2020)'s analyses, we employed a sliding-time-window technique to assess change in the ratio of word to syllable ITC over exposure. We measured ITC for every 10 words and calculated WLI. We fit logarithmic and linear models to WLI over exposure.

Results

In line with Choi et al., we tested whether the slope from the better-fitting model was significantly greater than zero,

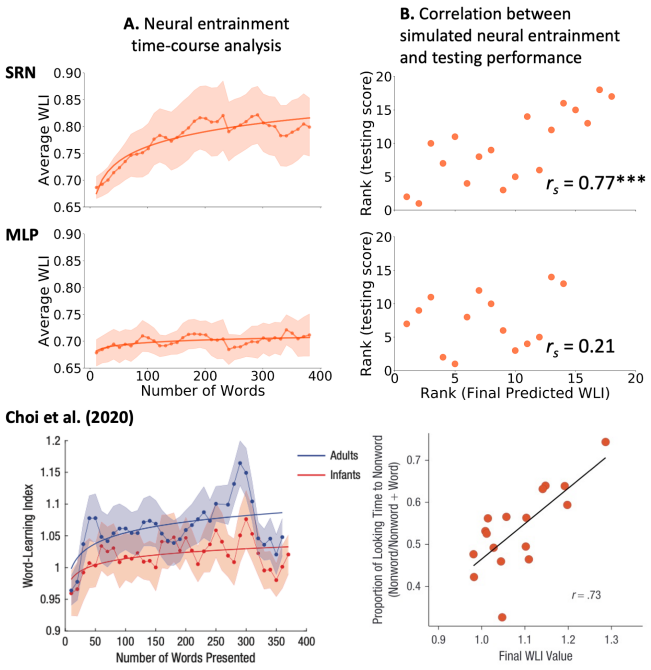

Figure 6: **A**. Word Learning Index (WLI) over exposure (words presented) for SRNs, MLPs, and infants from Choi et al. (2020). Each curve depicts the optimal logarithmic fit (error ribbons indicate standard error). **B**. Testing scores and final WLI values of SRNs, MLPs, and infants. Due to significant deviations from normality, X and Y axes for SRNs and MLPs indicate rank values. Note: *** indicates $p < .001$.

indicating a robust increase in WLI over training. Both MLPs and SRNs exhibited a significant logarithmic relationship between WLI and amount of exposure (number of words) (Fig. 6A), which aligns with trends observed in infant data. For SRNs, individual time-course data were modeled logarithmically because a logarithmic curve fit the average group-level data better than a linear curve ($logR^2 = 0.84, linearR^2 = 0.60$). WLI increased significantly as a function of exposure ($b = 0.04, SE = 0.003, t(17) = 13.80, p < .001, beta = 0.92$). Similarly, the MLP data also favored a logarithmic representation ($logR^2 = 0.37, linearR^2 = 0.29$). WLI increased significantly as a function of exposure ($b = 0.01, SE = 0.002, t(17) = 4.58, p < .001, beta = 0.61$), although the slope is more moderate compared to the SRNs or infants, as we discuss next.

Relating simulated entrainment and behavior

Choi et al. (2020) noted a correlation between WLI and infants' ability to distinguish words from nonwords behaviorally, suggesting a linkage between neural patterns and behavioral outcomes in infants. To test whether a similar correlation exists in our MLPs and SRNs, we computed the correlation between (a) the log-fitted final WLI values of individual model samples, and (b) the testing score based on Eq. 1.

Results We used Spearman correlations for both MLPs and

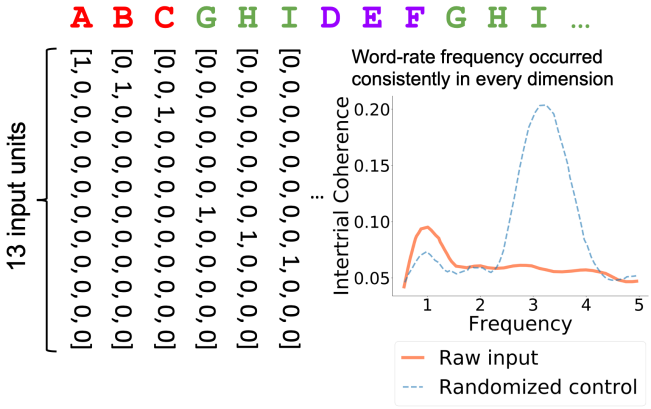

Figure 7: Raw syllable input sequences transformed to EEG-like time series also exhibits strong ITC at the word rate frequency of 1Hz – even if we analyze only a single input channel (i.e., the input pattern for a single syllable).

SRNs due to small numbers of values and marked deviations from normal distributions, as suggested by Q-Q plots. For SRNs, a significant positive correlation was observed between the models' test scores and final WLI values ($R_s = 0.77, p < .001$), mirroring patterns observed in infant data (Fig. 6B). However, the correlation was not significant for MLPs ($R_s = 0.21, p = .474$). In at least this detail, it seems SRNs provide a stronger link to human infants' patterns of behavior and neural entrainment than MLPs. Why this may be requires a deeper level of analysis that is beyond our current scope that we will pursue in our ongoing investigations.

Rhythms in the input?

Here we consider a critical question: does neural entrainment primarily reflect actual learning, low-level feature processing, or a combination of both? This analysis is motivated by the rationale we provided above for tracking hidden unit activations rather than output activations: as the system learns to predict the next syllable, output activations will increasingly resemble the raw input sequence, with the exception that predictions at word boundaries should be distributed among the 3 possible next (first) syllables. This raises a question: what would the implications be if analyses of the input (or output) sequences *also* showed phase locking at the word rate? If they did, this would radically challenge the interpretation of ITC at word rate as a neural index of learning.

We can easily conduct this analysis with our pipeline for creating pseudo-EEG data from patterns of 1s and 0s in 13-element vectors, since input vectors have the same number of elements as hidden vectors. We can apply the same transformation and analysis pipeline to the raw input sequences. We processed the raw input (1-hot vectors representing discrete syllables) (Fig. 7) after applying our temporal extension, jitter, and noise transformations using the ITC pipeline described above. To provide a control comparison, we repeated this but with input syllable order shuffled randomly.

Results Remarkably, even raw syllable input sequences, devoid of any word learning or explicit word markers, exhibited ITC at the 1Hz word rate (Fig. 7), while the randomized control sequence did so to a much lesser extent. This is true even if we analyze only a single channel (the input for 1 syllable). How is this possible?

Consider the constraints on a single syllable such as A from ABC. Because ABC cannot immediately repeat, if A appears in position 1 in the sequence, the next time it can appear is in position 7 (after 3 intervening syllables from another word, e.g., ABCGHIA...). If it does not appear at position 7, the next time it can appear is at position 10 (3 syllables later, e.g., ABCGHIDEFA...). Thus, while the minimal syllable interval between 2 instances of the same syllable is 6 (the equivalent of 2 secs if word rate is 1hz), other possible intervals include 9, 12, 15, etc. Converted to seconds, the possible intervals between 2 instances of the same syllable are 2, 3, 4, 5, 6... – any multiple of 1 greater than 2. This makes the fundamental frequency (f_0) of the sequence 1hz. A robust way to calculate f_0 is to enumerate the intervals between peaks, and determine the *Greatest Common Divisor* (GCD). If the possible intervals are integer multiples of 1 greater than 2, the GCD will be 1 (indeed, so long as the intervals contain at least 1 prime number and others that are not multiples of that prime number, the GCD must be 1).

Note that this finding does not depend on our use of 1-hot input and output patterns. This rhythm is intrinsic to the input sequences. It exists whether we consider all 13 channels simultaneously and would still exist if we created some distributed representation of the inputs (such as phonetic feature vectors rather than discrete syllable patterns).

Since the input sequence has no explicit representation of words, the fact that it contains rhythmic structure that drives ITC at the word rate implies an important caution for interpreting ITC as a neural index of learning. This result suggests that any system that could generate distinct states in response to each syllable would show ITC at word rate for the sequences used by Choi et al. (this assessment would have to be repeated with other input sequences to determine if inputs alone can drive word-rate ITC). We might have more confidence that ITC reflects word learning in a study like this one when changes in ITC converge with other measures of learning (such as the correlation between WLI and post-training word sensitivity observed for infants and SRNs). However, even with such a correlation, substantial ambiguity would remain. It could be that ITC reflects simple shifts to distinct states in response to each syllable rather than a neural signature of emergent lexical representations. It could be that distinct responses to component elements is a prerequisite for word learning, but for now, we cannot conclude that ITC at word rate unambiguously indexes word learning.

Discussion

This study provides the first computational simulations of neural entrainment observed in SL. By comparing 2 models (MLPs, only capable of learning first-order TPs, vs. SRNs, which learn higher-order contingencies as well) with human infant data, we made 3 discoveries. First, both models learn sufficiently from the input patterns to robustly simulate human infant preferences for word-like patterns vs. nonwords (cf. Magnuson, in preparation). Second, both models also exhibit human-like neural entrainment patterns, with phase-locked responses at both syllable and word rates. This required applying a method to transform discrete state changes in the models (1 state change per syllable, i.e., 3 Hz) to a finer-grained time series (1000Hz), and then treating activations over time in hidden unit channels like EEG channels. However, third, there was a quantitative difference in the change of syllable- and word-rate ITC over training, with the SRN showing a more human-like pattern, as well as a difference in the correlation between that change and post-training word sensitivity; the SRN showed a human-like significant correlation, while the MLP did not. This provides suggestive evidence that the more powerful learning possible in an SRN may provide a better model of human SL, though more work is needed to understand what specific emergent computations may drive the observed differences.

We also asked what kind of phase-locking would be predicted by the raw input sequences (i.e., what rhythmic structure would exist in a system that could simply encode [achieve a distinct state for] each syllable?). We had assumed that to detect emergent knowledge of word-like patterns, we would have to track hidden-layer states, which would have to exhibit patterns driven by knowledge of word-like patterns that would emerge over time. However, our ITC analyses of the raw input sequences (in comparison with randomly ordered control sequences) revealed that, indeed, input sequences alone can drive robust ITC at the word rate. Thus, a system that simply came to generate distinct responses to distinct syllables would show word-rate ITC. Settling into element-specific neural states could take some time (and a much simpler form of learning, such as becoming familiar with the repeating elements), precluding interpreting gradual emergence of word-rate ITC as an unambiguous index of word learning.

As we discussed above, correlations between ITC changes and post-training word sensitivity suggest a stronger link between word-rate ITC and learning, but it does not completely resolve the ambiguity. Resolving this ambiguity would require devising SL sequences with more complex statistical structure where the abstractions to be learned (chunking syllables into words, or learning about phrase-like patterns in inputs) are not apparent from ITC analyses applied to raw input sequences. It remains an open question whether any such patterns can be constructed (or whether any patterns used in previous SL studies might already have this characteristic). A crucial next step in our ongoing work will be to investigate more deeply why the word learning index (the ratio of ITC at word rate to the ITC at syllable rate) correlates strongly in humans and SRNs.

Acknowledgements

This work was conducted when Qihui Xu was a postdoctoral researcher at BCBL. This project was supported in part by National Science Foundation grant PAC 2043903 (PI JSM), by the Basque Government through the BERC 2022-2025 program, and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010-S and through project PID2020-119131GB-I00 (PI JSM).

References

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45.

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal infants discover statistical word patterns at similar rates as adults: Evidence from neural entrainment. *Psychological Science*, *31*(9), 1161–1173.

Contreras Kallens, P., Kristensen-McLachlan, R. D., & Christiansen, M. H. (2023). Large language models demonstrate the potential of statistical learning in language. *Cognitive Science*, *47*(3), e13256.

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, *14*(2), 179–211.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*, 195–224.

Frank, S. L., & Yang, J. (2018). Lexical representation explains cortical entrainment during speech comprehension. *PloS one*, *13*(5), e0197304.

French, R. M., Addyman, C., & Mareschal, D. (2011). Tracx: a recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological review*, *118*(4), 614.

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., ... others (2022). Shared computational principles for language processing in humans and deep language models. *Nature neuroscience*, *25*(3), 369–380.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. S. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*(267), 1–13. doi: 10.3389/fnins.2013.00267

Magnuson, J. S. (in preparation). Modeling human statistical learning with feedforward and simple recurrent networks.

Magnuson, J. S., & Luthra, S. (under review). Simple recurrent networks are interactive.

Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends in cognitive sciences*, *8*(5), 204–210.

Martin, A. E., & Doumas, L. A. (2017). A mechanism for the cortical computation of hierarchical linguistic structure. *PLoS biology*, *15*(3), e2000663.

Matusevych, Y., Schatz, T., Kamper, H., Feldman, N. H., & Goldwater, S. (2023). Infant phonetic learning as perceptual space learning: A crosslinguistic evaluation of computational models. *Cognitive Science*, *47(7)*, e13314.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, *323*(6088), 533–536.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining "learning" in statistical learning: What does an online measure reveal about the assimilation of visual regularities? *Cognitive science*, *42*, 692–727.

Sjuls, G. S., Harvei, N. N., & Vulchanova, M. D. (2023). The relationship between neural phase entrainment and statistical word-learning: A scoping review. *Psychonomic Bulletin & Review*, 1–21.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 hz visual responses in human. *Journal of Neuroscience*, *16*, 4240–4249.

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.