# Modeling Sensitivity to Infant-Directed Sentences Using Spiking Neural Networks: A Neuromorphic AI Approach to Early Language Learning

**Valentina Simon**
Yale University
valentina.simon@yale.edu

## Abstract

This study models infant-directed sentence sensitivity using a biologically plausible spiking neural network (SNN). Drawing from the Brent corpus in CHILDES, it was examined whether spike-timing-dependent plasticity and rate-coded sentence inputs can yield categorical structure, sensitivity to contrast, and generalization. Findings suggest that SNNs are a promising tool for developmental psycholinguistic modeling.

## 1 Introduction

Spiking neural networks (SNNs) are a biologically inspired class of neural networks that encode and process information through discrete spike events rather than continuous activations. Unlike traditional artificial neural networks, which operate on real-valued signals across continuous time steps, SNNs mimic temporal dynamics observed in biological neurons. Information is represented in both the rate and timing of spikes, making these models particularly suitable for studying neural coding in perceptual and cognitive tasks.

In recent years, SNNs have been leveraged to model sensory learning, motor control, and cognitive processes in ways that mirror the developmental trajectory of humans and animals (Wang et al., 2022, Hasegan et al., 2022). One promising domain is early language learning, where infants acquire phoneme and sentence structures by passively listening to naturalistic input. Computational models have traditionally employed vector-based methods for phoneme categorization and word segmentation (Poli et al., 2023), but the discrete temporal nature of language processing in the brain motivates the use of SNNs to explore these phenomena.

This report explores whether SNNs can capture infant-like generalization and contrast sensitivity in spoken language.

## 2 Neural Networks as a Model for Language Learning

Neural networks have become an increasingly valuable tool in computational psycholinguistics, offering scalable and learnable architectures that approximate cognitive functions such as perception, categorization, and memory. In the study of language acquisition, neural models allow researchers to simulate learning processes from raw input data, evaluate representational development, and compare model behavior to psycholinguistic benchmarks across development.

A variety of modeling approaches have been applied to infant language learning. Early models, such as Gaussian Mixture Models (GMMs), were used to approximate phoneme category formation from distributions over acoustic features (Schatz et al., 2021). While useful for capturing probabilistic clustering, these models often relied on strong assumptions about input representations and lacked temporal dynamics. More recent work has employed vector space embeddings and deep neural networks to model phonetic and lexical acquisition (Matusevych et al., 2020, Dupoux et al., 2018). These models improved performance on tasks like phoneme discrimination and word segmentation, but they typically abstract away from real-time processing constraints and lack biological interpretability.

Spiking Neural Networks (SNNs), by contrast, represent a class of models that incorporate biologically plausible spatiotemporal dynamics. Instead of relying on continuous activation values, SNNs use discrete spikes to encode information over time, mimicking how neurons process signals in the brain. Prior applications of SNNs have demonstrated success in modeling auditory feature processing (Deng et al., 2023) and unsupervised category learning using spike-timing-dependent plasticity (STDP) (Nessler et al., 2013). Despite

this, their use in modeling higher-level language constructs such as full-sentence processing remains rare.

This study contributes to this emerging area by exploring the feasibility of using SNNs to model sensitivity to sentence-level structure in infant-directed speech. By grounding learning in biologically motivated mechanisms such as STDP and rate coding, I aim to investigate whether SNNs can capture generalization and contrast sensitivity patterns observed in early linguistic development.

## 3 Research Questions

This study investigates:

1. Can an SNN cluster similar sentences based on output spike activity?
2. Does repeated exposure stabilize output patterns, akin to learning?
3. Can the SNN discriminate minimally different sentences?
4. Will the SNN generalize learned structure to novel but similar inputs?

## 4 Methods

I use the Brent subcorpus of the CHILDES database, a widely used collection of English mother-infant interactions. A total of 211 .cha files were loaded, and utterances marked as spoken by the mother (participant "MOT") were extracted. Sentences were tokenized and filtered to retain 3–10 word utterances, focusing on the 100 most frequent words to limit input dimensionality.

Each sentence was transformed into a firing-rate vector via rate coding: background words fired at 5 Hz and active words at 50 Hz. These vectors were passed into a spiking neural network implemented in the Brian2 simulator. The network comprises three biologically inspired components:

**Input Layer (PoissonGroup)**: The input layer consists of a PoissonGroup, where each neuron corresponds to one of the top 100 most frequent words in the corpus. For each sentence, the firing rate of each input neuron is determined by word presence, with spikes emitted according to a Poisson distribution. This rate-coded scheme captures the probabilistic variability of neural responses in biological systems and introduces temporal jitter characteristic of natural spike trains.

**Output Layer (Leaky Integrator Neurons)**: The output layer contains 10 leaky integrate-and-fire (LIF) neurons. These neurons integrate incoming synaptic inputs over time, with their membrane potentials decaying exponentially in the absence of stimulation. When a neuron's potential crosses a predefined threshold, it emits a spike and resets. This mechanism allows for temporal integration of input and models biological neural behavior.

**Plastic Synapses (STDP-based updates)**: Connections between the input and output layers are mediated by plastic synapses governed by spike-timing-dependent plasticity (STDP). STDP is a biologically realistic learning rule that modifies synaptic strength based on the relative timing of pre- and post-synaptic spikes. Synapses are strengthened (long-term potentiation) when a presynaptic spike precedes a postsynaptic spike and weakened (long-term depression) when the order is reversed. This rule enables the network to learn from repeated temporal correlations without supervision.

Spikes were recorded using `SpikeMonitor` and processed into spike count vectors for downstream analysis, including dimensionality reduction and sentence-level comparisons.

### 4.1 Overview of Spiking Network Dynamics

To better understand the computational principles underlying spiking neural networks (SNNs), Figure 1 illustrates the basic operation of a feed-forward spiking architecture. This schematic visualizes how temporally distributed input spikes, passed through synaptic connections, drive the activity of a postsynaptic neuron over time.

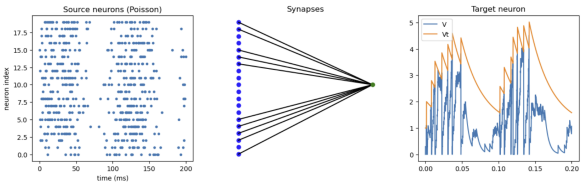


Figure 1: A schematic of spiking neural dynamics. (Left) A group of 20 source neurons generates spike trains following a Poisson distribution. (Center) These spikes are transmitted via synapses to a single target neuron. (Right) The postsynaptic membrane potential $V$ (blue) integrates incoming spikes and crosses the dynamic threshold $V_t$ (orange) to emit output spikes.

In the left panel, the source neurons generate spikes stochastically, emulating the rate-coded input used in the sentence encoding. Each dot represents a spike event for a given neuron at a specific

time.

The center panel shows how these inputs converge onto a single postsynaptic neuron via synapses. In this simulation, such connections are plastic and undergo modification through spike-timing-dependent plasticity (STDP), reinforcing temporal patterns over learning.

The right panel shows how the membrane potential $V$ of the target neuron evolves. As incoming spikes arrive, $V$ rises. When it surpasses the threshold $V_t$, a spike is emitted and the potential resets. This interaction of spike timing, integration, and thresholding enables the neuron to encode temporal patterns in its input—an essential feature of learning in spiking models.

#### 4.1.1 Leaky Integrate-and-Fire Neuron Dynamics

The target neurons in this model are leaky integrate-and-fire (LIF) neurons, whose membrane potential $V(t)$ evolves according to:

$$\tau_m \frac{dV(t)}{dt} = -V(t) + RI(t)$$

where $\tau_m$ is the membrane time constant, $R$ is the membrane resistance, and $I(t)$ is the synaptic input current. When $V(t)$ crosses a threshold $V_{\text{th}}$, the neuron fires a spike and $V(t)$ is reset to a resting potential $V_{\text{reset}}$.

#### 4.1.2 Spike-Timing-Dependent Plasticity (STDP)

Synaptic weights $w_{ij}$ between a presynaptic neuron $i$ and a postsynaptic neuron $j$ evolve based on the relative timing of their spikes. In the additive STDP rule, the change in synaptic weight $\Delta w$ depends on the time difference $\Delta t = t_{\text{post}} - t_{\text{pre}}$ between post- and presynaptic spikes:

$$\Delta w = \begin{cases} A_+ e^{-\Delta t/\tau_+} & \text{if } \Delta t > 0 \text{ (LTP)} \\ -A_- e^{\Delta t/\tau_-} & \text{if } \Delta t < 0 \text{ (LTD)} \end{cases}$$

where $A_+$ and $A_-$ are learning rates for long-term potentiation (LTP) and depression (LTD), and $\tau_+$, $\tau_-$ are time constants. This rule allows the network to strengthen connections that reliably predict output spikes and weaken those that do not.

Together, these mechanisms allow the network to adaptively encode temporal features of the input, facilitating emergent representations from raw spike-based data—a biologically inspired basis for language learning.

## 5 Experiments and Results

### 5.1 Sentence Clustering via Spike Patterns

The objective of this experiment was to determine whether the network can distinguish sentence types, evidenced by clustering similar sentence structures together. To test this, 500 in-vocabulary sentences were passed through the SNN. The resulting spike vectors were reduced to 2D with PCA.

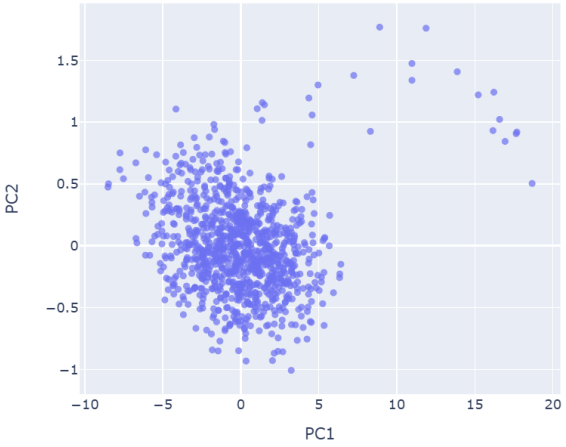


Figure 2: Emergent clusters from PCA on spike vectors. The X and Y axes have the two principal components after PCA. Each dot represents a sentence embedded in that space. While only a screenshot of the plot has been provided here, an interactive html of the graph is provided for the reader to experiment with in the Github. On the interactive version, placing the cursor over a dot results in the corresponding sentence being displayed. Additionally, it is possible to zoom in and out to inspect the clustering more closely.

The resulting 2D plot (Figure 2) reveals sentence clusters with visible separation. Structurally similar or lexically repetitive sentences (e.g., "do you want it", "do you want this") formed tight clusters, indicating that the network encodes similarity in activation space. Of note, all of the sentences that fall outside the cluster are ungrammatical. For instance, some of these include: “do whats going” and “did you wanna down”.

This shows that unsupervised STDP in a simple SNN can result in emergent representations that preserve semantic or syntactic similarity. The result supports the hypothesis that infants may rely on temporal spike-based encodings to differentiate utterances passively.

### 5.2 Repeated Exposure and Learning Stabilization

The objective of this experiment was to determine if repeated input reduces the variability of the response. That is, does a SSN simulate learning? In order to test this, the sentence "do you want it" was shown 50 times, and the neuron activations were observed, as shown in the Figure below.

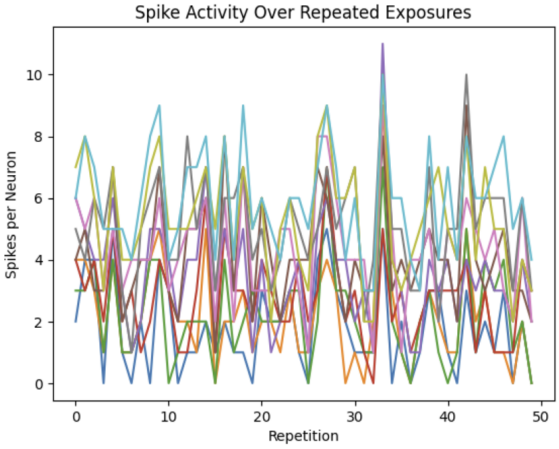


Figure 3: The x-axis represents the repetition index from 0 to 9, while the y-axis denotes the number of spikes generated by a given output neuron during each repetition. Each colored line corresponds to one of the network's output neurons, tracing how its activity fluctuates across exposures.

Contrary to expectations of convergence, the spiking patterns remain highly variable and do not exhibit stabilization over time. This ongoing fluctuation suggests that the model fails to develop consistent representations for the repeated input, raising questions about whether category-like attractor dynamics are being learned. The variability may result from continued plasticity via spike-timing-dependent mechanisms or from stochasticity in the rate-coded Poisson input. As such, the figure highlights a potential limitation in the current setup: without a mechanism to consolidate or freeze learned synaptic weights, the network does not settle into a stable response pattern.

Future work should consider distinguishing a training phase—during which learning is active—from a testing phase where synapses are fixed, in order to evaluate representational stability more clearly. Additional steps such as noise reduction or response smoothing could also help assess whether consistent encoding is developing beneath the observed variability.

### 5.3 Sensitivity to Minimal Sentence Contrasts

To investigate whether the spiking neural network exhibits contrast sensitivity—a hallmark of early language processing—pairs of minimally different sentences were fed into the model. Each pair varied by a single word, such as a possessive pronoun or noun, while maintaining an otherwise identical syntactic structure. The resulting spike count vectors for each sentence were compared across the 10 output neurons, and cosine similarity was used to quantify the overlap between activity patterns. Three representative examples are shown in Figures 4–6.

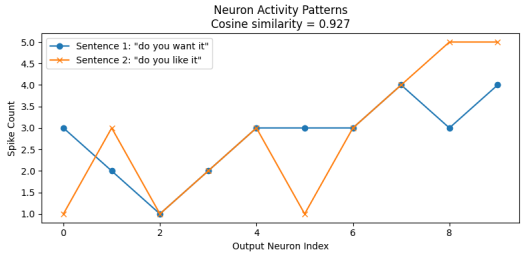


Figure 4: Comparison of "that is my shoe" and "that is your shoe". Cosine similarity = 0.949. The x-axis indicates neuron index; y-axis shows spike count.

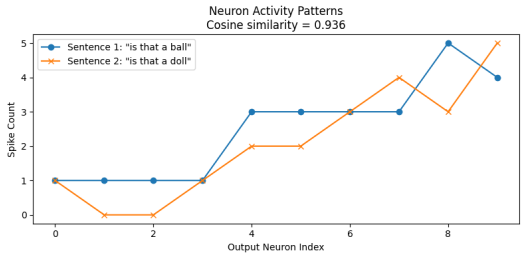


Figure 5: Comparison of "look at the car" and "look at the cat". Cosine similarity = 0.837. The x-axis indicates neuron index; y-axis shows spike count.

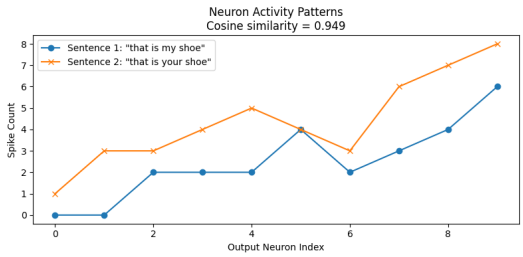


Figure 6: Comparison of "do you want it" and "do you like it". Cosine similarity = 0.927, The x-axis indicates neuron index; y-axis shows spike count.

Note, in the context of a spiking neural network, each output neuron does not represent a specific lin-

guistic unit (such as a word or concept), but rather reflects a distributed response based on the accumulated input activity and learned synaptic weights. The neurons function as a population code; meaning is not encoded in any single neuron, but rather in the pattern of activity across the ensemble. When comparing two spike patterns, I examined whether their distributed representations are similar—an indicator that the network is treating them as functionally related inputs. The figures capture how much each neuron was activated in response to a given sentence, and the resulting pattern is shaped by the input encoding (rate-coded words), synaptic weights shaped by STDP, and the network's architecture. Over time, certain neurons may become more responsive to particular syntactic or semantic features, but this is emergent rather than explicitly designed.

Across all sentence pairs, the model produced highly similar spike patterns for semantically and syntactically related inputs, with cosine similarity values ranging from 0.837 to 0.949. Notably, the pair "look at the car" versus "look at the cat" yielded the lowest similarity, suggesting that even minor phonological changes in noun identity can be reflected in downstream neural activity. This aligns with infant psycholinguistic findings that early learners are sensitive to both lexical and phonemic contrasts.

The relatively high similarity in all cases indicates that the model encodes sentence structure in a distributed but robust fashion. Neurons exhibit systematic tuning to small changes, particularly in sentence-final words. These results suggest that the spiking model supports the emergence of structured representations and may be capable of differentiating semantically distinct utterances in a manner analogous to human learners.

All 20 sentence pairs tested can be found in the .html file on the Github. The three sentence pairs presented in this section are representative of the overall trends observed in the larger dataset.

### 5.4 Generalization from Similar Input Structures

To evaluate whether the spiking neural network could generalize its representations to novel but structurally similar inputs, I conducted a test in which the model was exposed repeatedly to two training sentences—"look at that" and "look at this"—and subsequently tested on the novel input "look at those." These three sentences differ only in the final demonstrative pronoun, allowing a focused probe of lexical generalization.

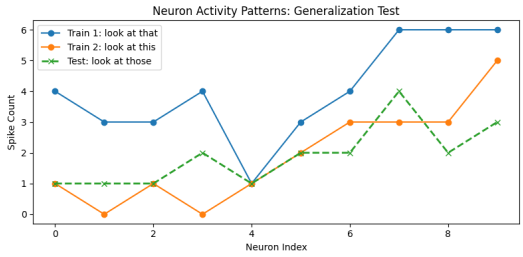


Figure 7: Spike count distributions for "look at that" and "look at this" (training inputs) and "look at those" (test input). The x-axis indicates neuron index; y-axis shows spike count.

As shown in Figure 7, the spike pattern elicited by the test sentence "look at those" (green dashed line) partially aligns with the profiles of both training inputs. The shared structure "look at" likely contributes to consistent early neuron activations, while divergence emerges near the sentence-final word. This mirrors behavioral findings in infants, where generalization is supported by familiar structural frames but influenced by novelty in lexical content.

While the spike counts for "look at those" are generally lower in magnitude—potentially due to lack of synaptic strengthening during training—the shape of the activation profile retains some similarity to the learned patterns. This suggests that the SNN encodes local similarities in sentence structure and may be capable of generalization within related lexical sets. However, the partial divergence indicates room for improvement in the system's ability to abstract away from specific exemplars.

These results support the hypothesis that spiking models can exhibit sensitivity to shared syntactic scaffolding, an important feature of early sentence processing and generalization during infant language acquisition.

## 6 Limitations and Future Work

While this study demonstrates the potential of spiking neural networks (SNNs) for modeling infant-directed sentence sensitivity, several limitations constrain the current implementation and point to valuable future directions.

First, the network architecture is a simple feed-forward model without recurrence or lateral inhibition. As a result, it cannot capture dynamic temporal dependencies beyond the integration window of

leaky neurons. Real-world infant language processing likely involves recurrent feedback and working memory, which are critical for handling longer sequences and dependencies. Incorporating recurrent SNNs or reservoir computing may better simulate incremental sentence parsing and retention.

Second, input representation was based on a simplified rate-coded model using word identity alone, without phonological, prosodic, or acoustic detail. This abstraction omits many cues known to influence infant speech perception, such as stress, intonation, and phoneme boundaries. Future work could integrate low-level auditory models or cochleagrams to drive input spikes, enhancing ecological validity.

Third, the results of the repetition experiment indicate unstable spiking patterns across exposures. This suggests that while spike-timing-dependent plasticity (STDP) supports adaptive learning, it may require either additional consolidation mechanisms (e.g., synaptic normalization, homeostasis) or a fixed testing phase to ensure stability. Distinguishing between online learning and evaluation phases could clarify how learned categories solidify over time.

Fourth, while sentence clustering and contrast sensitivity were promising, this model does not yet ground these distinctions in downstream behavioral metrics. Future work should compare model outputs directly to empirical infant data (e.g., looking-time paradigms or discrimination thresholds) to validate representational alignment.

Lastly, the current model assumes full observability of tokens, bypassing issues of segmentation and word discovery that are central to early language acquisition. Expanding the model to learn from raw speech or to jointly acquire segmentation and categorization could extend its developmental relevance.

## 7 Conclusion

This pilot study explored the use of spiking neural networks (SNNs) to model sentence-level sensitivity in infant-directed speech. Leveraging the Brent subcorpus of CHILDES and implementing biologically inspired mechanisms such as spike-timing-dependent plasticity and rate-coded inputs, I evaluated the model's ability to cluster sentences, respond to repeated exposures, detect minimal contrasts, and generalize to novel utterances.

Results demonstrate that even a simple SNN can yield structured spike-based representations of linguistic input, showing sensitivity to both syntactic frames and lexical variation. Sentence clustering emerged without supervision, contrastive pairs elicited distinct but similar patterns, and novel sentences triggered activity partially aligned with training data, suggesting a primitive capacity for generalization.

However, the network also showed limitations in stability over time, with spike responses failing to converge across repeated exposures. This highlights the need for more sophisticated training-testing phase distinctions and synaptic consolidation mechanisms to better simulate learning dynamics.

Overall, this work demonstrates the potential of SNNs as a cognitively and biologically plausible framework for studying early language learning. It lays foundational groundwork for future research bridging computational neuroscience and developmental psycholinguistics through temporally grounded models.

## 8 Code

https://github.com/ScienceFair2018/SpikingNeuralNetworks-InfantLanguageLearning.git

## References


[1] Schatz, T., Feldman, N. H., Goldwater, S., Cao, X.-N., Dupoux, E. (2021). Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input. *Proceedings of the National Academy of Sciences, 118*(7), e2001844118. https://doi.org/10.1073/pnas.2001844118

[2] Dupoux, E. (2018). Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition, 173*, 43–59. https://www.sciencedirect.com/science/article/abs/pii/S0010027717303013

[3] Matusevych, Y., Schatz, T., Kamper, H., Feldman, N. H., Goldwater, S. (2020). Evaluating computational models of infant phonetic learning across languages. *arXiv preprint arXiv:2008.02888*. https://arxiv.org/abs/2008.02888

[4] Deng, B., Fan, Y., Wang, J., Yang, S. (2023). Auditory perception architecture with spiking neural network and implementation on FPGA. *Neural Networks, 165*, 31–42. https://doi.org/10.1016/j.neunet.2023.05.026

[5] Nessler, B., Pfeiffer, M., Buesing, L., Maass, W. (2013). Bayesian computation emerges in generic cortical microcircuits through

spike-timing-dependent plasticity. *PLoS Computational Biology, 9*(4), e1003037. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003037

[6] Wang, Y., Zeng, Y. (2022). Multisensory concept learning framework based on spiking neural networks. *Frontiers in Systems Neuroscience, 16*, 845177. https://doi.org/10.3389/fnsys.2022.845177

[7] Hasegan, D., Deible, M., Earl, C., D'Onofrio, D., Hazan, H., Anwar, H., Neymotin, S. A. (2022). Training spiking neuronal networks to perform motor control using reinforcement and evolutionary learning. *Frontiers in Computational Neuroscience, 16*, 1017284. https://doi.org/10.3389/fncom.2022.1017284

[8] Poli, M., Chemla, E., Dupoux, E. (2023). Improving spoken language modeling with phoneme classification: A simple fine-tuning approach. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics. https://aclanthology.org/2023.emnlp-main.302