Insights into Early Word Comprehension: Tracking the Neural Representations of Word Semantics in Infants.

A Preprint

Rohan Saha*
Department of Computing Science
University of Alberta
rsaha@ualberta.ca

Jennifer Campbell*
Department of Psychology, UBC Language Sciences
University of British Columbia
jenncscampbell@gmail.com

Janet F. Werker
Department of Psychology
University of British Columbia
jwerker@psych.ubc.ca

Alona Fyshe
Department of Computing Science and Psychology
University of Alberta
alona@ualberta.ca

October 28, 2021

Abstract

Infants start developing rudimentary language skills and can start understanding simple words well before their first birthday [Bergelson and Swingley, 2012]. This development has also been shown primarily using Event Related Potential (ERP) techniques to find evidence of word comprehension in the infant brain [Parise and Csibra, 2012, Friedrich and Friederici, 2010]. While these works validate the presence of semantic representations of words (word meaning) in infants, they do not tell us about the mental processes involved in the manifestation of these semantic representations or the content of the representations. To this end, we use a *decoding* approach where we employ machine learning techniques on Electroencephalography (EEG) data to predict the semantic representations of words found in the brain activity of infants. We perform multiple analyses to explore word semantic representations in two groups of infants (9-month-old and 12-month-old). Our analyses show significantly above chance decodability of overall word semantics, word animacy, and word phonetics. As we analyze brain activity, we observe that participants in both age groups show signs of word comprehension immediately after word onset, marked by our model's significantly above chance word prediction accuracy. We also observed strong neural representations of word phonetics in the brain data for both age groups, some likely correlated to word decoding accuracy and others not. Lastly, we discover that the neural representations of word semantics are similar in both infant age groups. Our results on word semantics, phonetics, and animacy decodability, give us insights into the evolution of neural representation of word meaning in infants.

Keywords EEG · Semantics · Brain imaging · Infant Decoding

1 Introduction

Recent work has demonstrated that infants begin to learn words by six months of age [Bergelson and Swingley, 2012, Tincoff and Jusczyk, 2012]). This astonishing ability of young infants has raised many questions about early cognitive development including how deep infants' early word understanding is. Each word in a language has a *meaning* associated with it. This meaning is usually referred to as *semantics* and represents the conceptual idea attached to each word (e.g., cat is an animal, apple is a fruit and is edible). These semantics can be modelled computationally using

---

*Authors had equal contribution

mathematical models, which are obtained from statistical modeling of large corpora of text. The trained semantic models of text express various characteristics of words such as gender, plurality, etc. [Mikolov et al., 2013a]. These semantic models have been beneficial for studying language processing in the brain by empowering us to better understand the pattern of neural analysis that the brain is performing while processing words. Such an analysis can be performed using a *decoding* approach where a machine learning model is used to predict the stimuli from recorded brain activity.

Previous studies conducted on brain-imaging data recorded from adults have shown success in decoding word semantics [Papadimitriou et al., 2018, Fyshe et al., 2019, Honari-Jahromi et al., 2021]; we extend this idea to brain-imaging data recorded from infants. Specifically, we analyze the change in neural response patterns as the infants hear single words. Infants in two age groups (9-month-old and 12-month-old) listened to single words spoken by a native English speaker while their EEG data was recorded. We then trained a machine learning model to decode the semantic representations of the words from the brain activity.

Our analyses on brain-imaging data collected from infants provide evidence that,

- EEG data recorded from 9-month-old and 12-month-old infants can be used with a machine learning approach to reveal semantic information.
- Nine- and twelve-month-old infants show high decoding accuracy of words immediately after word onset.
- Individual word stimuli properties such as phonetics can be decoded from infants of both age groups.
- The neural responses of 9-month-old and 12-month-old infants show similar representations of word semantics.

2 Background

Compared to the plethora of semantic decoding research conducted with neuroimaging data collected from adults, there is a scarcity of research where machine learning techniques have been applied to decode semantics from infant neuroimaging data. Semantic research in young infants has traditionally been carried out using behavioral tasks, involving looking time paradigms [Bergelson and Swingley, 2012], or, for neuroimaging data, Event Related Potentials (ERP), a characteristic of EEG data [Friederici, 2005]. The ERP component used in prior infant research is the N400, a high amplitude negativity that emerges in adults at around 400ms (and later in infants) to a mismatch between a spoken or written word and its referent [Kutas and Federmeier, 2000]. This component, which has been widely used in adult studies, has been shown to emerge reliably by one year of age in infants [Friedrich and Friederici, 2010], see [Junge et al., 2021] for a systematic review of N400 work with infants). For example, study conducted in 2012 by [Parise and Csibra, 2012] used the N400 with 9-month-old infants to show that the participants can detect a mismatch between an object and a word label that preceded the object. Infants aged 9-month-old were presented with live audio stimulus from their mother in one experimental condition and from an experiment in another condition. Then a congruent or an incongruent object relative to the preceding audio stimulus was shown on a screen placed in front of infants. The results indicated that infants best represent the object features associated with familiar words in the mother-speech condition indicating that the infants *understood* their mother's speech. Still, all of these studies limit researchers to probe infants' understanding of a small number of words at a time.

Increasingly, machine learning is being applied to the rich neural activity data obtained from various brain-imaging techniques to reveal the featural composition of words in the adult lexicon. Several studies have demonstrated the potential of machine learning to study adult word comprehension. A cognitive state classification conducted by Mitchell in 2004 [Mitchell et al., 2004] showed that machine learning methods can be employed for decoding semantic categories of written stimuli from whole brain fMRI (functional magnetic resonance imaging) data collected from adults. In this study, university students participated in various tasks while their fMRI data was collected. Then a comparative analysis was performed using different machine learning algorithms (Gaussian Naïve Bayes, Support Vector Machines and k-Nearest Neighbors) on the fMRI data to classify cognitive states. This indicated that machine learning models can be successfully applied to brain-imaging data .

Other studies that required participants to perform a language related task, also showed that the lexical (word level) semantic information is encoded in the brain. A study conducted in 2014 by Fyshe et. al. built vector space models of words by incorporating semantic information found in the brain [Fyshe et al., 2014]. Eighteen participants took part in the study and their MEG and fMRI data were recorded while they viewed sixty concrete nouns. To represent these sixty nouns, *vector space models* that encode semantic information were built. A simple L_2 regularized linear regression model was used to predict these word vectors from brain-imaging data. It was observed that when brain-imaging data and text were leveraged to build the word vectors, the regression model outperformed in predicting the word vectors compared to predicting them when the word vectors were built using only text. The finding that the word vectors built by leveraging brain-imaging data in addition to text were better at representing word semantics than text alone shows that the brain-imaging data had semantics encoded in it.

Another study conducted by Sudre in 2012 used MEG data where machine learning was used to decode perceptual and semantic features of 60 concrete nouns [Sudre et al., 2012] using a classification approach. By decoding perceptual and semantic features, an MEG based classifier was able to determine two different concrete nouns not seen during training. The temporal analyses conducted provided insights into the difference between the time course of MEG magnitude and the decodable semantic information; it showed that perceptual information can be decoded before semantic information. This study showed that applying machine learning to timeseries data can help us discover dynamic patterns that vary with time.

Machine Learning has also been used with MEG data to decode phrase stimuli semantics from brain-imaging data. For example, [Fyshe et al., 2019] showed that a simple linear model (ridge regression) was able to track the neural decoding of adjective and noun phrase semantics from MEG data. A total of thirty adjective-noun phrases were present in the stimuli set, out of which twenty-eight were used for training and two for model performance evaluation. The decoding model also predicted the semantics of the adjective-noun phrases, indicating that neural responses contain information pertaining to the stimuli much later after stimuli offset. All in all, machine learning methods are a powerful approach for semantic decoding research and also for studies involving temporal analyses of brain-imaging data.

As we have seen with brain-imaging data recorded from adults, machine learning can also be applied to brain-imaging data collected from infants. In fact, classical and modern machine learning techniques have been used to investigate EEG data in developmental studies. For example, Gibbon used Support Vector Machine (SVM) and Convolutional Neural Network (CNN) to show that rhythmic stimuli can be accurately classified from EEG data collected from infants of 8-months old [Gibbon et al., 2021]. Another study used SVM on EEG data obtained from 6-month-old infants to accurately classify their risk of developing language related disorders [Zare et al., 2016]. These findings demonstrate that machine learning can be used on brain-imaging data collected from infants in downstream tasks and using computational models can also provide us with valuable insights into the dataset over traditional methods such as Event Related Potential (ERP).

Other techniques such as the Multivariate Pattern Analysis (MVPA) technique have also been used to decode the infant mind. Studies such as [Emberson et al., 2017] used functional near-infrared spectroscopy (fNIRS) brain-imaging data (a technique that measures changes in cortical blood oxygenation) in a task that classifies a set of stimuli into audio or visual categories. Using fNIRS and fMRI techniques is ideal when topographical neural information is vital, but it comes at the cost of temporal resolution. Our study uses EEG for a language related task as it is ideal for observing changes in neural responses over time. Specifically, we collect EEG data while infants are listening to words with various semantic and phonetic characteristics, and then use machine learning to identify what properties of the words the infants are able to detect, represent, and understand.

## 3 Materials and Methods

3.1 Participants

A total of 46 infants participated in our study. Of this initial sample, 21 nine-month-old (10 female, 11 male) infants were included in the final analysis. Seven additional 9-month-olds were tested but excluded from the final sample due fussiness (*n*=5), poor electroencephalogram (EEG) impedance (*n*=1), and excessive body movements (*n*=1). Fourteen 12-month-old (7 female, 7 male) infants were included in the final sample. An additional four 12-month-old infants were tested but were excluded due to language criteria (*n*=1) and fussiness (*n*=3). All infants were native English learners, had no cognitive impairments and were born full-term.
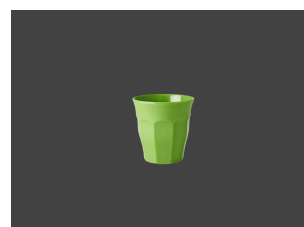
3.2 Stimuli

The stimuli presented to the participants consisted of two parts, an audio stimulus and a visual stimulus. The audio stimuli represented single words, which were spoken into a microphone by a native English female speaker (the experimenter). The visual stimuli represented images shown on a screen placed in front of the participants.

The word stimuli set consisted of sixteen total words equally divided into animate and inanimate words. The stimuli words were selected using the reported average age of acquisition from the MacArthur–Bates Communicative Development Inventory (CDI) [Fenson, 1993], and previous experimental studies [Bergelson and Swingley, 2012, Tincoff and Jusczyk, 1999, 2012, Bergelson and Swingley, 2015, 2018]. Table 1 shows the list of sixteen words grouped by animacy. It is useful to note that each spoken word lasts a different duration, but requires less than one second to speak.

For each audio stimuli, an image followed. Sixteen images matching the auditory stimuli were presented to the participants. Each image was publicly available on Google Images and had black backgrounds. The images represented the real-life object equivalent of the corresponding word label and not toys. For example, the image for *dog* showed a
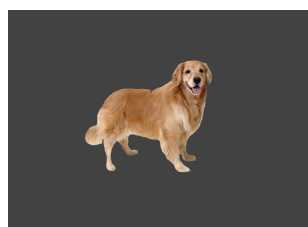
(a) Dog as an animate image in the stimuli set.

(b) Cup as an inanimate image in the stimuli set.

Figure 1: An example of an animate and inanimate image presented to the participants after the word stimuli.
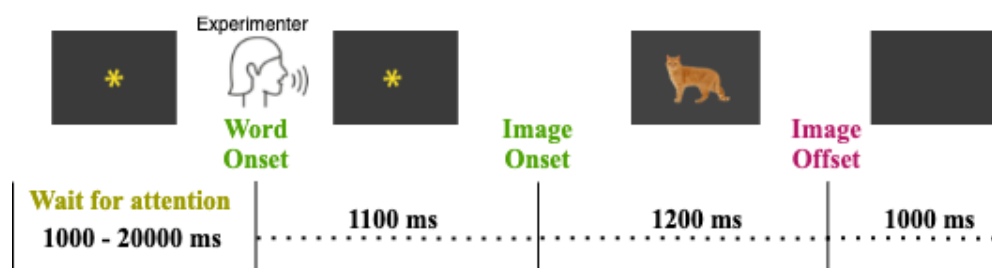

Figure 2: Timing diagram showing stimulus presentation design. The word stimulus was in audio format and the image was shown on a screen. Before each trial a dynamic video was shown with a rotating asterisk that changed colors. After word onset, the rotating asterisk was shown for at least 1100ms. Next, an image appeared on the screen lasting 1200ms. After the image offset, a blank screen followed that lasted 1000ms..

real-life dog and not a toy dog. Figure 1 shows an example of an animate and an inanimate image presented to the infants.

Stimuli Words	
Animate	**Inanimate**
baby	banana
bear	bottle
bird	cookie
bunny	cracker
cat	cup
dog	juice
duck	milk
mom	spoon

Table 1: Stimuli words used in the study.

Before each trial, the computer displayed a dynamic video as an attention getter on the screen. The dynamic attention getter video displayed a small rotating asterisk that changed colors. This video was used to ensure the participants' attention based on the criteria that they were either looking at the rotating asterisk or the experimenter. The experimenter spoke the word marking the word stimuli onset, following which the rotating asterisk was shown for 1100ms. Next, the image was presented on the screen that lasted for 1200ms. Finally, a blank screen was shown for 1000ms (see figure 2 for details). On average, the time between image offset and the next word stimulus varied from 2100ms to 2500ms because participants' attention was required, and the reaction time of the speaker was accounted for to the next word stimulus provided by a computer software. A timing diagram is shown in figure 2.

For each trial, the Eprime software [Schneider et al., 2002] provided a stimulus word through the speaker's headphones, such that the infants and the parents could not hear the word. Then, as the speaker said the word, the microphone in front of the speaker registered the word stimulus and would signal the Eprime software to present an image after, at least 1100ms after the word onset. For the computer to register the onset of the spoken word stimulus, a threshold level of 65dB was used.

An image was shown for 1200ms on the screen in front of the participant after each word stimulus. For half of the trials, the image matched the word label (for example, an image of *dog* was shown for the preceding word label *dog*). For the other half of the trials, the image did not match the preceding word label (for example, an image of a *cup* was shown for the preceding word label *banana*). Half of the mismatched images matched the preceding word for its animacy category, and the remaining half mismatched the animacy category of the word. This resulted in four cases of word-image pairs. These were:

1. Animates congruent.
2. Animates incongruent.
3. Inanimates congruent.
4. Inanimates incongruent.

In the congruent case, the image matches the animacy group of the word (e.g., animate word *baby* and an image of the animate object *cat*), and in the incongruent case, the animacy does not match the preceding word stimulus (e.g., inanimate word *cup* and an image of the animate object *dog*). The images were presented in a semi-random order such that there could not be more than two presentations of the same word type (animate or inanimate) in a sequence.

3.3 Data Collection and Preprocessing

Throughout the study, infant brain imaging data were collected using Electroencephalography (EEG) at a western Canadian university. EEG is a brain-imaging technique that involves placing electrodes on the scalp to record the electrical activity of the brain. It is a non-invasive and affordable method that provides excellent temporal resolution. EEG is ideal for our study as we wanted high temporal precision to analyze the changes in brain activity over time. We used a Geodesic 64 channel EEG cap which had a data capturing frequency of 1000 hertz. Channels 61-64 were removed during pre-processing since these channels corresponded to cheek channels which were not present on all caps. The experimental procedure was carried out using Eprime software with Chronos hardware [Schneider et al., 2002] and data was captured using 5.4.2 EGI Netstation software.

During the study, infants were seated on their parent's lap in front of a computer monitor. Standing next to the screen was a native female English speaker who produced the stimulus words during the study. At the start of each trial, a fixation asterisk would appear in the centre of the screen. The fixation asterisks rotated and changed colors while staying fixed in the middle of the screen. In this initial part of the trial, the speaker waited for the infant's attention, defined by the infant looking at either the screen or speaker. This period of time was on average 1000 to 20000ms. While the speaker waited for the infant's attention, the stimulus word was played through headphones such that the infants and the parents could not hear it. Once the infant was attending and relatively still, the speaker said the word and a microphone in front of the speaker registered the word stimulus and and signalled the Eprime software to present an image after a 1100ms delay. For the computer to register the onset of the audio stimulus, a threshold level of 65dB was used. The image would remain on the screen for 1200ms after which a blank screen was presented for 1000ms.

For our analyses, we processed the EEG data by removing trials where the baby was not paying attention or was not looking at the screen or the speaker before each trial. We also removed any trials that contained movement artifacts which were found by visual inspection of the EEG signal and video feed of the infant's face. In the end, we had a total of 1026 samples of EEG data for the 9-month-old babies and a total of 683 samples of EEG data for the 12-month-old babies. The data was sampled at a rate of 1000Hz and was not downsampled. Data was filtered with a high pass filter of .1Hz and a low pass filter of 50Hz. The data was re-referenced using a vertex channel and then epochs were extracted from -200 to 1000ms around the onset of each word utterance. The data was also baseline corrected using the 200ms prior to word onset.

3.4 Word vectors

To obtain a vector representation of word semantics, we used word vectors obtained from the Word2Vec model [Mikolov et al., 2013b] pretrained on the Google News Dataset (an internal Google dataset containing news articles and comprising about a hundred billion tokens[2]). These word vectors represent the semantic properties of the stimuli. Each word vector has 300-dimensions. For our study, we used the vectors acquired from the skip-gram algorithm, a neural network trained to predict the context for a given word. We chose vectors from Word2Vec because the existing studies show that word vector representations of stimuli can be decoded from brain-imaging data in adults [Ruan et al., 2016, Kivisaari et al., 2019, Sudre et al., 2012, Murphy et al., 2011, Foster et al., 2021, Honari-Jahromi et al., 2021].

---

[2] https://code.google.com/archive/p/word2vec/

3.5 Data Subsets and Averaging

The EEG dataset collected is represented by $X \in \mathbb{R}^{N \times p}$, where N is the number of samples and $p = sensors \times time$, is the number of features. A machine learning model learns a function h that transforms the input EEG data X into a different m dimensional space Y. This m dimensional space consists of the 300-dimensional word vectors that represent the semantics of the word stimuli $Y \in \mathbb{R}^{N \times m}$.

Our data has a total of p = $60 \times 1000 = 60000$ features for 60 sensors and 1000 time points for each sensor. We divided our data into equal length chunks of 100ms windows and used a sliding step of 10ms. Therefore, for each window, we train a ridge regression model which takes in a data matrix of size $N \times t$ where $t = 60 \times 100 = 6000$ is the total number of features (60000 total data points for 1 second of data chunked into 100ms windows). So our original data matrix $X^{N \times p}$ is now $X^{N \times t}$ ($t < p$) on which the model is trained to predict the 300-dimensional word vectors. The data matrix $X^{N \times t}$ is then divided into training and testing subsets X_{train} and X_{test}. The word vector representations Y is similarly divided into Y_{train} and Y_{test}. The training subset X_{train} along with Y_{train} is then used to train multiple ridge regression models to learn the mapping from the brain imaging data (X_{train}) to the word vectors (Y_{train}). For our test set X_{test}, we average all the trials for a stimulus word within an age group resulting in 16 total samples (one sample of each stimulus word) to increase the single-to-noise ratio. For example, if the test set contains five EEG samples for the word 'baby', we average these five samples to obtain one sample. This process is carried out for each word in the stimuli set, giving us 16 EEG samples for X_{test}.

3.6 Prediction Model

Our primary analyses use a simple machine learning algorithm. Specifically we use the Ridge Regression model from scikit-learn [Pedregosa et al., 2011], which was trained on the EEG data to predict the word vectors.

To train the model, we learn a mapping h from X to Y. h takes the brain-imaging data X and predicts word vectors $\hat{Y}$.

$$h(X) = \hat{Y} \tag{1}$$

or

$$h(X) = X \times \hat{\mathbf{w}} = \hat{Y} \tag{2}$$

To learn the mapping h, we estimate the weights $\hat{\mathbf{w}}$ by minimizing the loss function:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} ||X\mathbf{w} - Y||_2^2 + \lambda \mathbf{w}^T \mathbf{w} \tag{3}$$

The term $\lambda \mathbf{w}^T \mathbf{w}$ is called the ridge or the L2 regularizer and the hyperparameter λ controls the regularization strength.

We used monte-carlo nested cross-validation procedure with negative mean squared error to optimize the hyperparameter λ for the ridge regression model (see eq. 3). During each sampling iteration of the monte carlo procedure, we randomly sampled 80% of the dataset which constituted as X_{train} and the rest 20% constituted as the test set. For the inner loop of the nested cross validation procedure, we used a k-fold split of the training set(X_{train}) with k=5 resulting in 20% of X_{train} being used as the validation set for the hyperparameter tuning. We utilize sklearn's GridSearchCV[3] to accomplish hyperparameter optimization. For the EEG data from the 9-month-old infants, this resulted in 821 samples for X_{train} and 205 samples for X_{test} for each sampling iteration. About 164 samples from X_{train} were used for selecting the hyperparameter inside each cross-validation fold. For the EEG collected from the 12-month-old infants, there were 547 samples for X_{train} and 136 samples for X_{test}. About 109 samples from X_{train} were used for hyperparameter tuning. We set the number of sampling iterations for the monte-carlo procedure to 50.

3.7 2 versus 2 Test

To evaluate the performance of the model, we used the 2 vs 2 test procedure. The test procedure works as follows. We first consider two instances from the predictions $(\hat{y}_i, \hat{y}_j)$ and two instances from the ground truth vectors (y_i, y_j). We then use a distance metric to calculate the distance between the two pairs of vectors. We use the cosine distance criterion $d(u, v)$ where u and v are the word vectors. The chance accuracy of the 2 vs 2 test is 50% because there are two possible assignments of the predicted vectors to the ground truth vectors.

The 2 vs 2 test passes if:

$$d(y_i, \hat{y}_i) + d(y_j, \hat{y}_j) < d(y_i, \hat{y}_j) + d(y_j, \hat{y}_i) \tag{4}$$

The 2 vs 2 test is depicted in figure 3.

Since our test set contains the average of all the trials for a stimulus word, the total number of instances in our test set is 16. And the total number of pairs evaluated by the 2 vs 2 test are $\binom{16}{2} = 120$.

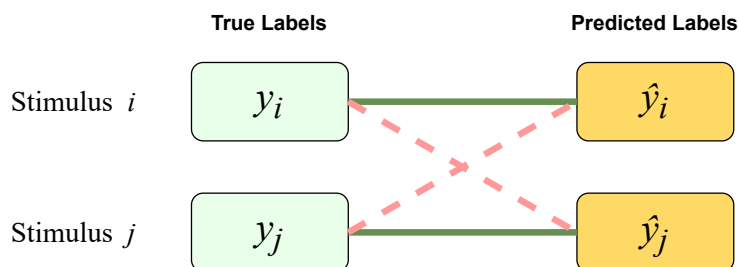[3]https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

Figure 3: 2 vs 2 test showing pairwise comparisons of true and predicted labels. The test passes if the sum of the cosine distance between matching word vector pairs (represented by solid green lines) is less than the sum of the distance between the non-matching word vector pairs (represented by red dashed lines).

3.8 Testing for Above Chance Accuracy

To determine the analysis windows with above chance accuracy, we compared our results to random chance accuracy using the permutation test [Ojala and Garriga, 2010]. The permutation test is performed by randomizing the assignment of stimuli word vectors to their corresponding EEG data so that each word vector for a word stimulus is randomly matched to any sample of the EEG data. This mimics a situation where no association exists between the input EEG data X and the output word vectors Y. We ran our decoding framework on the dataset for 100 permutation iterations. As expected, since there is no relation between the input EEG data and the output word vectors, the accuracy when running the prediction framework on this shuffled dataset is chance (50% for 2 vs 2 test). We then fit a gaussian kernel density function to the obtained 2 vs 2 accuracies from the permuted dataset to generate a null distribution. Finally, we compared the non-permuted accuracies to the accuracies obtained from the permutation test to establish statistical significance.

Once the null distribution is obtained, a p-value is calculated for the original non-permuted accuracy. To correct the p-values for multiple comparisons over time ($\alpha < 0.01$), we use the Benjamini-Hochberg-Yekutieli False Discovery Rate (FDR) correction method with no dependency assumption [?].

3.9 Testing for Difference Between Conditions

After we trained our machine learning model to decode word semantics from EEG data, we used a different significance testing method called the non-parametric cluster permutation test [Maris and Oostenveld, 2007] to compare the significant difference in accuracy values between different experimental conditions. The two conditions in our analyses are the two age groups (9-month-old and 12-month-old infants). The test is conducted as follows. First, we obtain the 2 vs 2 accuracies for each condition, resulting in fifty accuracy values for each sampling iteration (see 3.6). Next, we calculate a test statistic between the accuracy values for the two conditions, which gives us the observed test statistic. We then permute these accuracies by putting them into a single set, from which two random subsets of accuracies are drawn; this is called a random partition. We repeat this process 10000 times. We calculate the test statistic between the two subsets for each random partition and compare it to the observed test statistic, resulting in a p-value. Finally, we report the largest cluster of time points containing at least 3 or more consecutive windows with a significant difference in accuracy values (p < 0.01).

4 Results

4.1 Analysis 1: Decoding Animacy Category from EEG

In the first analysis, we sought to explore whether infants' neural representations include complex properties such as animacy. To this end, we used our pre-processed EEG data to predict whether a word labelled an animate object or an inanimate object. If infants have learned an object's animacy and have mapped that property to the object's label, then we would expect successful decoding using the infant EEG signal.

In Analysis 4.1, we found that the prediction framework was able to decode the word's animacy from 9-month-old's EEG data but not from 12-month-old's EEG data. For 9-month-old infants, accuracy was significantly above chance at multiple time points over the course of the trial. This provides evidence that 9-month-old infants' representations of word labels includes information about the animacy of the corresponding objects. This is consistent with prior research showing that infants' ability to distinguish animate from inanimate objects emerges very early in infancy [Rakison and

Poulin-Dubois, 2001] and guides their reasoning about actions, defining features, mental states, and more (see [Opfer and Gelman, 2011] for a review).

Interestingly, we did not find any success in the ability to predict word animacy using only the 12-month-olds' EEG data. This could have been due to a lack of data from the 12-month-old infants to allow for successful model training in combination with a weak signal of animacy in the neuroimaging data. It is also noteworthy, that even in the 9-month-olds, we found that the accuracy of predictions fluctuated across the trial which may further suggest a weak signal of animacy.
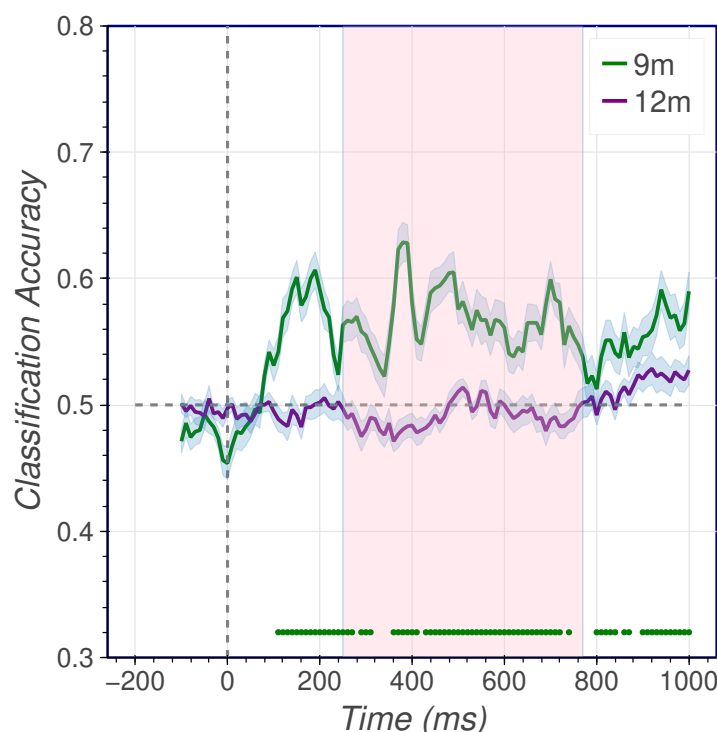

Figure 4: Accuracy for decoding word animacy category from the stimuli word. Green curve - 9-month-old infants; Purple curve - 12-month-old infants. Each point on the graph represents an accuracy value for the model trained on a 100ms window (100ms to the left of the accuracy point) with a 10ms sliding step. Green dots show significantly above chance accuracy points for 9-month-olds ($p < 0.01$, FDR corrected for multiple comparisons over time). No significantly above chance accuracy points were obtained for 12-month-infants. The shaded (pink) area shows significant difference in accuracies between the two age groups using non-parametric statistic test Maris and Oostenveld [2007].

4.2   Analysis 2: Representations of Word Semantics

In the second analyses, our goal was to find evidence of word comprehension in the infant neural signal. Here, we obtained the word vector representations for each word in our stimuli set from the pretrained Word2Vec model and ran the prediction framework to predict each dimension of the word vector from the EEG data. If the word meaning is encoded in the brain, we expect to obtain a successful prediction and significantly above chance accuracy.

In Analysis 4.2, we found that the prediction framework was able to decode the stimuli word vectors from the EEG data collected from 9-month-old and 12-month-old infants. This provides evidence about the presence of word semantic information in infant neural data. In addition to previous studies [Parise and Csibra, 2012, Friedrich and Friederici, 2010] which showed that babies can detect semantic mismatch of the stimuli, this analysis shows that infants can *understand* the meaning of the stimuli word as well. Since our analyses is carried out over 100 milliseconds chunks of EEG data with 10 millisecond sliding step, it allowed us to explore individual sections of the EEG data that encode word semantics and observe the results over time.

Interestingly, for 9-month-old infants, all accuracy values from the onset of the word till 950ms after word onset were found to be above chance and reliable (green dots in figure 5). For 12-months-old infants, we also observed multiple accuracy values after word onset but only until around 700 milliseconds, after which the accuracy dropped.
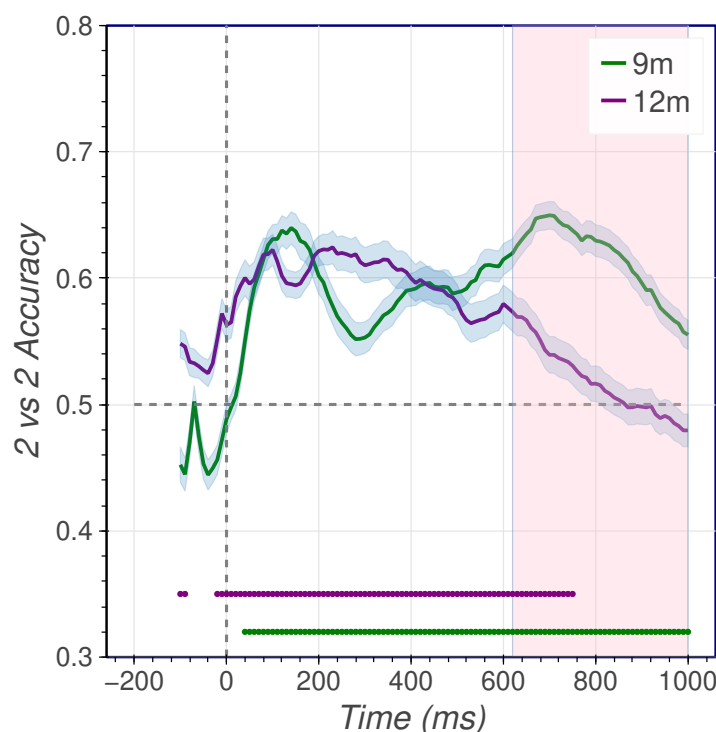
Figure 5: 2 vs 2 accuracy for predicting pretrained Word2Vec word vectors from EEG data collected from 9-month-old and 12-month-old infants. Each point on the graph represents an accuracy value for the model trained on a 100ms window (100ms to the left of the accuracy point) with a 10ms sliding step. The green dots above the x-axis represent the points where the accuracy is reliable for 9-month-old infants, and the purple dots represent reliable above chance accuracy for 12-month-old infants ($p < 0.01$, FDR corrected for multiple comparisons over time)

After conducting Analysis 4.2, we understood that the evoked neural responses in babies from both age groups were correlated with the stimuli words and the results also provides evidence that a machine learning approach is suitable for associating infant word semantics comprehension with the neural data. But we were unsure about whether the high decoding accuracy can be attributed to the semantics only or were there any confounding variables or features in the EEG data that helped the model decode the stimuli. Since babies of such age groups are very sensitive to phonetic differences in known words [Jusczyk and Aslin, 1995, Parise and Csibra, 2012], in the next analysis we removed the phonetic components of the word stimuli from their word vector representations.

4.3 Analysis 3: Presence of Phonetic representations of Word Stimuli

In Analysis 4.2, we decoded the overall word semantic representations from the EEG data. However, what constitutes this semantic understanding is unclear. In this analysis, we try to investigate one aspect of semantic understanding of words: word phonetics.

We know that early infants build up phonological memory structures and use them to recognize known words [Jusczyk and Aslin, 1995], and infants from birth are very sensitive to phonetic differences in speech sounds [Werker and Tees, 1992]. Extending this idea, we investigated the contribution of word phonetics to semantic understanding by predicting a vector representation of word phonemes from the infants' neural responses.

We carry out this analysis as follows. Each word in our stimuli set comprises multiple letters (ranging between three to six letters), and each letter is represented as a phoneme. We map each letter of the word stimuli to its respective phoneme, where each phoneme was represented in its IPA notation[4]. Whenever possible, the Canadian Pronunciation was used, and the General American (GA) pronunciation otherwise, defined in Dictionary.com[5] and Wiktionary[6]. Then,

---

[4]https://en.wikipedia.org/wiki/International_Phonetic_Alphabet
[5]https://www.dictionary.com/
[6]https://www.wiktionary.org/

we map each phoneme to a 36-dimensional vector obtained from [Mielke, 2012]. We concatenated each phoneme vector into one long vector P and zero-padded all the concatenated vectors to make them of equal size, resulting in a 216-dimensional long vector. Finally, we predict these 216-dimensional vectors from the EEG data using a ridge regression model. The process of creating these concatenated phoneme vectors is shown in 6.

The presence of phonetic components in the neural responses will be indicated by significantly above chance accuracy. Figure 7 shows the results where we predict the phoneme vectors from the neural responses.
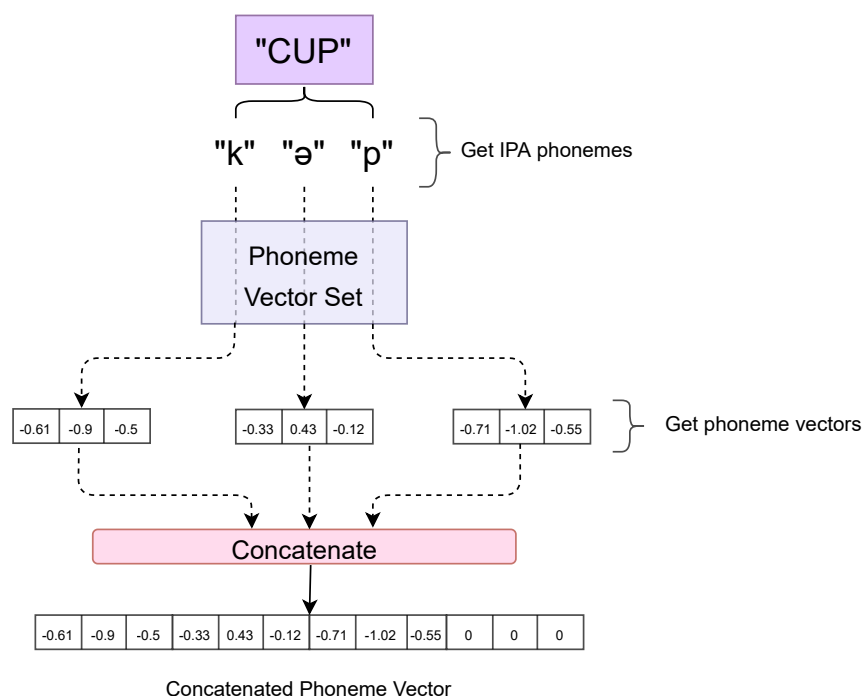

Figure 6: Phoneme vector creation process for a sample word 'cup' from the stimuli set. The stimulus word is broken down into its individual IPA phoneme. For each IPA phoneme, a vector is retrieved from [Mielke, 2012]. Finally, a long vector is created by concatenating individual vectors. Zero-padding is used to obtain equal length vectors. We show 3-dimensional phoneme vectors for simplicity. Actual vectors are 36 dimensional long.

It is interesting to see that for 9-month-old infants (orange curve in figure 7a), we were able to decode phonetic information from EEG data with significant above chance accuracy. This indicates that phonetic information is encoded in the neural responses of 9-month-old infants. We observed significantly above chance accuracy immediately after word onset, which lasted till 1000ms. On the other hand, when we observe the accuracy for the 12-month-old infants (black curve in figure 7a), we notice an initial rise in decoding accuracy around word onset, which lasts till around 380ms and quickly dropping after that. This suggests that even though the 12-month-old infants encode phonetic information, it is transient and not as long lasting as the phonetic representations for the 9-month-old infants. The pink shaded area shows regions with significant differences between decoding accuracy for phoneme vectors for 9-month-old infants and 12-month-old infants. Starting around 400ms, the neural representations differ greatly, resulting in a *divergence* of the accuracy curves.

We also compare the results between the current phoneme decoding analysis and the semantic decoding analysis described in Analysis 4.2, in figure 7b for 9-month-old infants and in figure 7c for 12-month-old infants.

In figure 7b, we see that for 9-month-old infants, the accuracy curve for the phoneme decoding analysis (orange) traces close to the accuracy curve for the semantic decoding analysis (green). On the flip side, when we compare the phoneme decoding analysis and semantic decoding analysis for 12-month-old infants, we see that in figure 7c, the curve for phoneme decoding analysis (black) and semantic decoding analysis (purple) trace each other closely till around 400ms after which the phoneme decoding accuracy drops sharply but semantic decoding accuracy persists till about 750ms. This effect is depicted by the pink shaded area showing a significant difference in accuracy from around 370ms to 1000ms.

In this analysis, we tried to understand the contribution of word phonemes to neural responses. We observed that infants in both age groups have phonetic representations of words, but the neural responses in the two age groups differ significantly. The next section will investigate how similar the neural responses are across the two age groups for our stimuli set.

4.4   Analysis 4: Shared Representations of Word Semantics between age groups

Analysis 4.4 tries to investigate if the evoked neural responses observed for a stimulus word contain similar information between 9-month-old infants and 12-month-old infants. Like previous experiments, the analysis is for all time windows of the EEG data. Similar to experiments 4.1 and 4.2, a ridge regression model is trained on the EEG data from one group to learn the mapping between the EEG data and word vectors, but during the test time, we predict the word vector representations from the EEG data obtained from the other age group.

The accuracy in word decoding in the previous experiments showed generally similar patterns between the 9-month-old and 12-month-old infants. To more directly measure whether there are commonalities in 9- and 12-month-old's representations we ran two analyses wherein we trained on one age group to predict the data of the other age group. Specifically, we wanted to know if the information present in the neural data from one age group can predict the word vectors from the neural data obtained from the other age group.

As seen in Figure 8, training on 9-month-old data was able to significantly predict the word vectors using a 2 vs 2 analysis of the 12-month-old infants in a window between 130ms to 610ms. Similarly, when we trained using 12-month-old data we were able to predict the word vectors of 9-month-old infants immediately after onset until 800ms after word onset. However, when we trained on 12-month-old infants to predict word vectors from EEG data obtained from 9-month-old infants, accuracy was not above chance for the entire period from word onset to 800ms and accuracy fell below chance around 200ms and 480ms; this is shown in figure 8.

Our result that both age groups were able to predict the word vectors of the other age group during large segments of the test trials, suggests that there are shared aspects in the semantic representations of these word vectors across age groups. This supports the idea that although there may be developmental changes across age groups, there is nonetheless an overwhelming similarity between them as well. Training on the older infants' EEG data was able to make predictions of the young infants' word vectors early in the test trial and over many windows in the test trial. The pattern was very different when we trained on younger infants to predict the older group. This difference could be due to noisier data in the younger infants possibly stemming from less developed word representations as compared to 12-month-old infants. Prior work has found changes in early word comprehension between 9 and 12 months of age (Bergelson and Swingley, 2017). It is possible then, that our work further speaks to the idea that older infants had significantly more advanced representations than the younger infants; however, more work is needed to explore this idea.

4.5   Analysis 5: Decoding fine-tuned word vectors from EEG

In Analysis 4.2, we ran our prediction framework to predict the word vectors obtained from the pretrained Word2Vec model. Since the pretrained Word2Vec model is trained on the Google News Corpus, the word vectors will likely be more oriented to have semantic representations to that of adults. One might also ponder upon improving these word vectors to better represent the semantic representation of words in infants. This thought stems from the idea that word vectors characteristic of infants' representation of word semantics may lead to better infant neural data predictions than pretrained word vectors. To this end, we fine-tuned the pretrained Word2Vec model on two datasets, the children's book test dataset introduced in [Hill et al., 2016] and the Childes corpus dataset [MacWhinney, 2000]. For the fine-tuning process, we used one thousand iterations of gradient descent with default $start$ and end learning rates[7]. Finally, we proceeded to decode these fine-tuned word vector representations from the EEG data for which the results are shown in 9.

In figure 9a we see that fine-tuned word vectors can be decoded with significantly above chance accuracy from the EEG data recorded from both groups of infants. However, we see a clear difference when comparing the results with the decoding accuracy of pretrained word vectors. In figure 9b, we notice that the decoding accuracy for pretrained word vectors (green) is high in the later stages of the time course compared to the decoding accuracy for the fine-tuned word vectors (orange). The pink shaded area shows the region where the accuracy values are significantly different. Similarly, for 12-month-old infants, we see that in figure 9c the decoding accuracy for pretrained word vectors (purple) is generally higher than the accuracy of fine-tuned word vectors (blue). Since using fine-tuned vectors did not perform superior to their pretrained counterpart, we used pretrained word vectors for our analyses.

---

[7]https://radimrehurek.com/gensim/models/word2vec.html#gensim.models.word2vec.Word2Vec.train
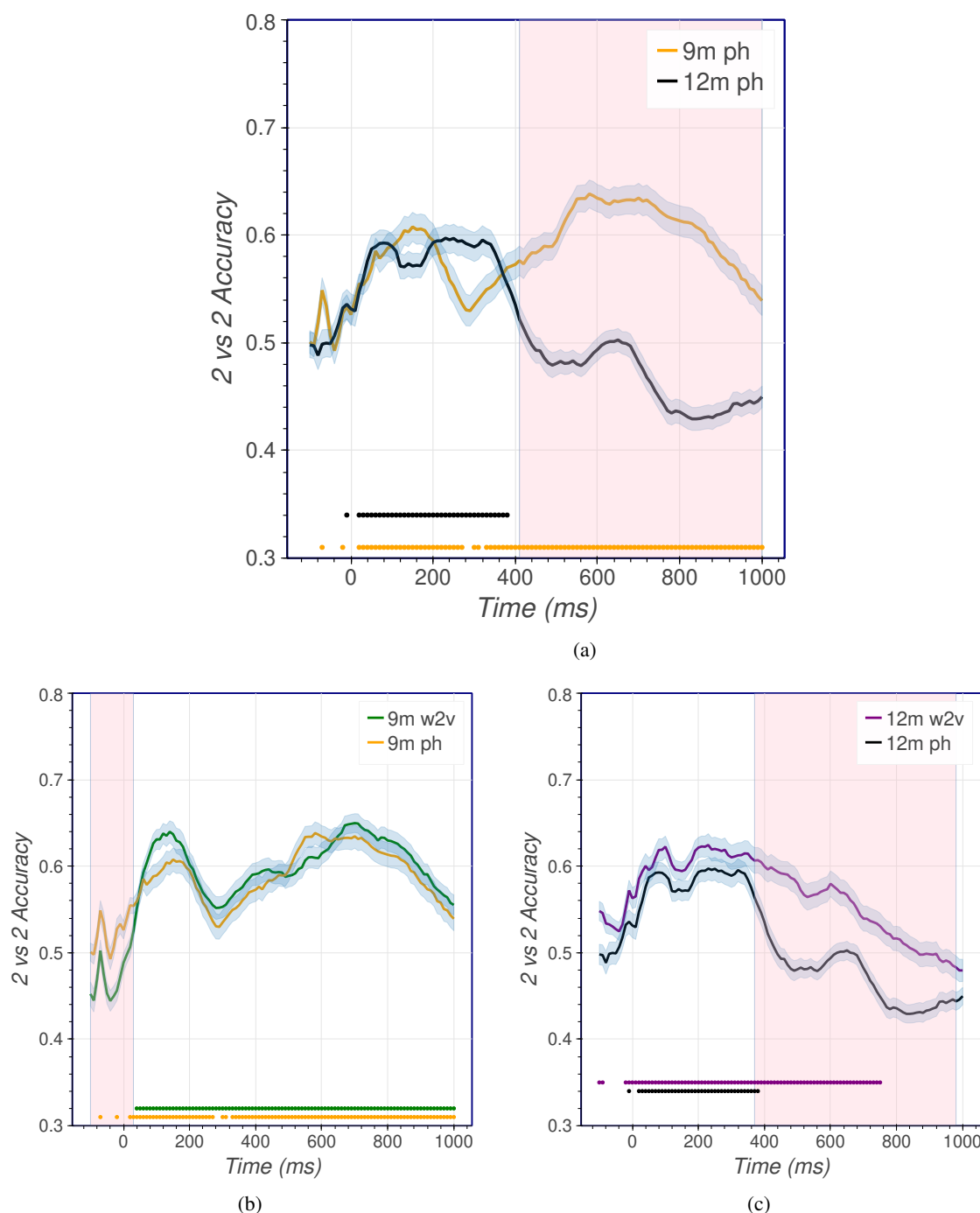
(a)

(b)

(c)

Figure 7: Decoding accuracy for predicting phonetic components of word stimuli. 7a shows the decoding accuracy for predicting phonemes from EEG data recorded from 9-month-old and 12-month-old infants. Figure 7b shows the comparison of accuracy for predicting phonemes (orange) and Word2Vec semantic vectors (green) from EEG data recorded from 9-month-old infants. Figure 7c shows the comparison of accuracy for predicting phonemes (black) and Word2Vec semantic vectors (purple) from EEG data recorded from 12-month-old infants. Shaded areas show significant differences in accuracies, and dots above the x-axis show points of significantly above chance accuracy (p-value < 0.01).
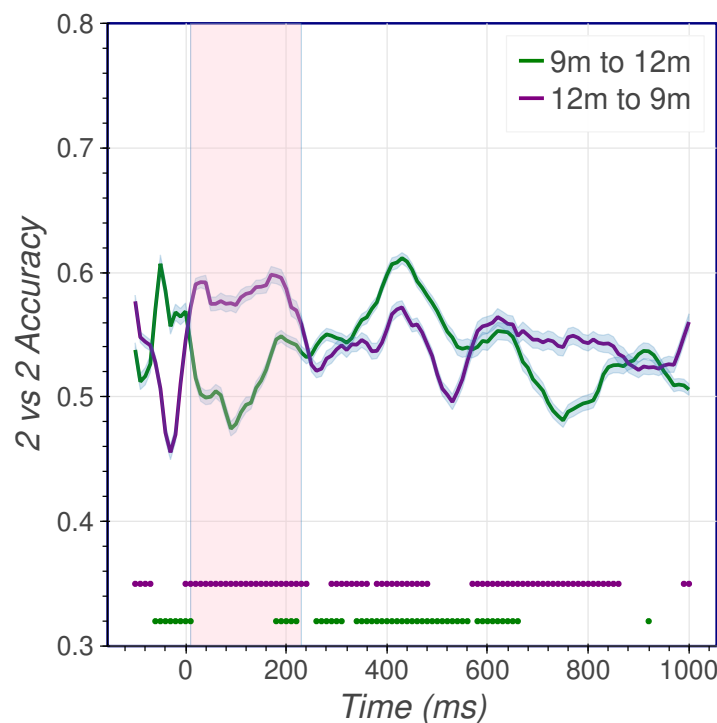
Figure 8: 2 vs 2 accuracy for predicting pretrained Word2Vec word vectors from EEG across age groups. Each point on the graph represents an accuracy value for the model trained on a 100ms window (100ms on the left of the accuracy point) with a 10ms sliding step. The green dots denote above chance accuracy for the model trained on 9-month-old and tested on 12-month-old infants, and purple dots denote above chance accuracy for the model trained on 12-month-old infants and tested on 9-month-old infants ($p < 0.01$, FDR corrected for multiple comparisons over time). The shaded area shows where the accuracy curve is significantly different for the two conditions.

5   Discussion

This is the first work to report successful decoding of words from neural activity from infants before or around their first birthday. Prior work examining semantic representations in infants, 12-months-old and younger, commonly approached this topic by either using looking-while-listening paradigms (e.g., [Bergelson and Swingley, 2012]) or relying on ERP analyses (e.g., [Friedrich and Friederici, 2005]). These methods relied on examining infants' ability to match a label to a picture by comparing trials with correctly matched label-image pairs and mismatched pairings. Although the prior work establishes that infants have an early understanding of words, it cannot tell us about the semantic understanding of these early learned words only whether infants detected a mismatch on some dimension. Specifically, the prior work with young infants is limited in that it can tell us whether infants attribute categorical meaning to an object word but cannot tell us much more about the semantics of early infant words. Our work presented here begins to address this question of early semantic comprehension through a novel method by showing that the infant neural signal triggered in response to auditory words can be modeled using adult semantic word vectors.

It is significant that infants' representations of words in the first year of life are complex and similar enough to the adult semantic representation for an adult based model (i.e., Word2Vec) to fit infant neuroimaging data. In Analysis 4.1, we predicted word animacy from the EEG data, which provides evidence that infants as young as 9-month-old can distinguish words based on their animacy category. In Analysis 4.2, we used 300-dimensional word vectors constructed from adult writing samples to model infant neuroimaging data after infants heard a word. We were able to decode the words infants were hearing from EEG data from both 9-month-old and 12-month-old infants. Our models achieved significantly accuracy almost immediately after word onset and it maintained this accuracy for over 700ms. There is some possibility that the decoding success early in trials for the 9-month-old infants appears to be somewhat correlated to the phonemes of the words. As shown in figure 7, when we decoded phoneme information from the EEG data, we obtained significantly above chance accuracy appeared approximately almost immediately after word onset in infants of both age groups. While the decoding accuracy for the 9-month-old infants closely traced the word vector decoding
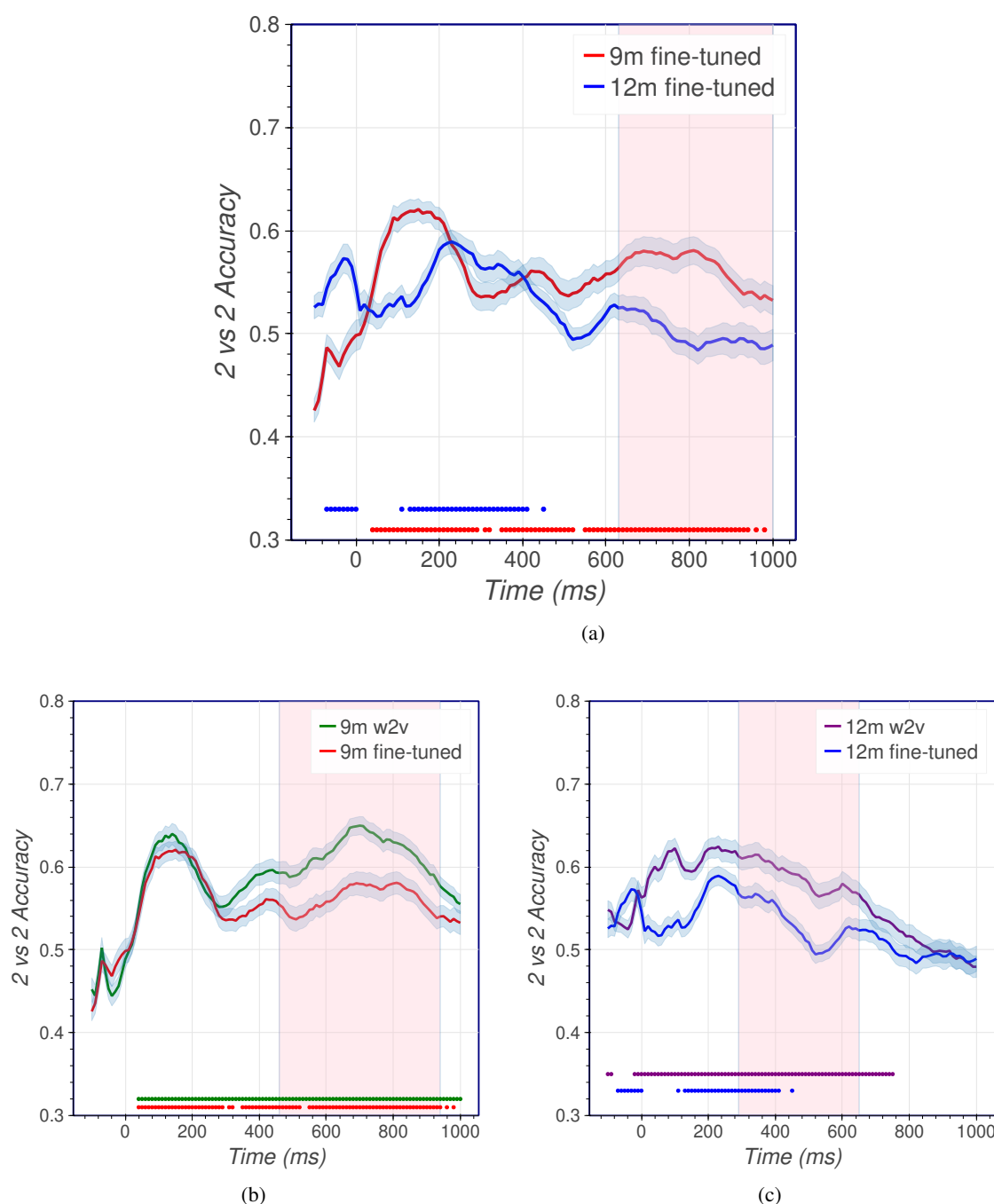
13

(a)

(b)

(c)

Figure 9: 9a shows the decoding accuracy for predicting fine-tuned word vectors from EEG data. 9b shows the accuracy comparison for predicting fine-tuned word vectors (orange) and pretrained word vectors (green) from EEG data collected from 9-month-old infants. 9c shows the accuracy comparison for predicting fine-tuned word vectors (blue) and pretrained word vectors (purple) from EEG data collected from 12-month-old infants. Shaded areas show significant differences in accuracies, and dots above the x-axis show points of significantly above chance accuracy (p-value < 0.01, FDR corrected for multiple comparisons over time).

accuracy, it is also interesting to observe that the phoneme decoding accuracy for 12-month old dropped quickly at around 400ms but the overall word semantic decoding accuracy lasted until around 750ms, indicating that the latter half of the semantic decoding trial is unlikely to be correlated to word phonetics only.

The general pattern of accuracy achieved neuroimaging data to predict word vectors was similar across age groups. We conducted an analysis to directly compare these age groups, wherein we found that the data from either age group was able to predict the word vectors of the other (Analysis 4.4). This result suggests that the representations between the age groups may be very similar to one another. Although the predictions made using the 12-month-old data appeared to be longer lasting they were not stable across the whole trial so it is difficult to conclude if the older infants had a more mature representation. Additionally, the peak accuracy values obtained while training on either age group were similar, further supporting that neither age group *were better in* predicting semantic word vectors.

In the first set of experiments, we focused on trying to predict adult word vectors; however, our accuracy values obtained may have been lower because of a mismatch in the representations of child semantic knowledge and adult semantic knowledge. Yet, when we fine-tuned our adult word vectors using child-directed databases, the accuracy values in both age groups suffered. This was a surprising result and may be a result of the child-directed speech data being quite small to conduct a modelling experiment. It is also possible that by using speech directed at older children rather than exclusively infants, we introduced unnecessary noise by now having word vectors that unintentionally further mismatched the infant experience. Further studies would be needed to disentangle this result.

6 Conclusion

In this study, we explored whether words could be decoded from young infants' neuroimaging data and the temporal nature of infants' early word representations. In both 9-month-old and 12-month-old infants, we were able to decode words from their EEG data early after word onset. Moreover, we discovered that both age groups appeared to have similarities in their representations. It is particularly impressive that infant EEG data was able to decode adult word vectors, suggesting that infants' early representations may be more similar to adults' than previously thought. Finally, this work opens up a new area of research by demonstrating for the first time that words can be decoded from the infant brain signal prior to or around infants' first birthday.

References

E. Bergelson and D. Swingley. At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9):3253–3258, feb 2012. doi:10.1073/pnas.1113380109.

Eugenio Parise and Gergely Csibra. Electrophysiological Evidence for the Understanding of Maternal Speech by 9-Month-Old Infants. *Psychological science*, 23(7):728–733, July 2012. ISSN 0956-7976. doi:10.1177/0956797612438734. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4641316/.

Manuela Friedrich and Angela D. Friederici. Maturing brain mechanisms and developing behavioral language skills. *Brain and Language*, 114(2):66–71, August 2010. ISSN 1090-2155. doi:10.1016/j.bandl.2009.07.004.

Ruth Tincoff and Peter W. Jusczyk. Six-month-olds comprehend words that refer to parts of the body. *Infancy*, 17(4):432–444, 2012. ISSN 1532-7078(Electronic),1525-0008(Print). doi:10.1111/j.1532-7078.2011.00084.x.

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic Regularities in Continuous Space Word Representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 746–751, Atlanta, Georgia, June 2013a. Association for Computational Linguistics. URL https://aclanthology.org/N13-1090.

Angeliki Papadimitriou, Nikolaos Passalis, and Anastasios Tefas. Decoding Generic Visual Representations From Human Brain Activity using Machine Learning. *arXiv:1811.01757 [cs, q-bio]*, November 2018. URL http://arxiv.org/abs/1811.01757. arXiv: 1811.01757.

Alona Fyshe, Gustavo Sudre, Leila Wehbe, Nicole Rafidi, and Tom M. Mitchell. The lexical semantics of adjective–noun phrases in the human brain. *Human Brain Mapping*, 40(15):4457–4469, jul 2019. doi:10.1002/hbm.24714.

Maryam Honari-Jahromi, Brea Chouinard, Esti Blanco-Elorrieta, Liina Pylkkänen, and Alona Fyshe. Neural representation of words within phrases: Temporal evolution of color-adjectives and object-nouns during simple composition. *PLOS ONE*, 16(3):e0242754, March 2021. ISSN 1932-6203. doi:10.1371/journal.pone.0242754. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0242754.

Angela D. Friederici. Neurophysiological markers of early language acquisition: from syllables to sentences. *Trends in Cognitive Sciences*, 9(10):481–488, October 2005. ISSN 1364-6613. doi:10.1016/j.tics.2005.08.008. URL https://www.sciencedirect.com/science/article/pii/S136466130500241X.

Marta Kutas and Kara D. Federmeier. Electrophysiology reveals semantic memory use in language comprehension. 4 (12):463–470, dec 2000. doi:10.1016/s1364-6613(00)01560-6.

Caroline Junge, Marlijne Boumeester, Debra L. Mills, Mariella Paul, and Samuel H. Cosper. Development of the N400 for Word Learning in the First 2 Years of Life: A Systematic Review. *Frontiers in Psychology*, 12:2420, 2021. ISSN 1664-1078. doi:10.3389/fpsyg.2021.689534. URL https://www.frontiersin.org/article/10.3389/fpsyg.2021.689534.

Tom M. Mitchell, Rebecca Hutchinson, Radu S. Niculescu, Francisco Pereira, Xuerui Wang, Marcel Just, and Sharlene Newman. Learning to Decode Cognitive States from Brain Images. *Machine Learning*, 57(1):145–175, October 2004. ISSN 1573-0565. doi:10.1023/B:MACH.0000035475.85309.1b. URL https://doi.org/10.1023/B:MACH.0000035475.85309.1b.

Alona Fyshe, Partha P Talukdar, Brian Murphy, and Tom M Mitchell. Interpretable semantic vectors from a joint model of brain and text based meaning. The 52nd Annual Meeting of the Association for Computational Linguistics , Baltimore, Maryland. 2014, 2014. URL https://www.cs.cmu.edu/~afyshe/papers/acl2014/jnnse_acl2014.pdf.

Gustavo Sudre, Dean Pomerleau, Mark Palatucci, Leila Wehbe, Alona Fyshe, Riitta Salmelin, and Tom Mitchell. Tracking neural coding of perceptual and semantic features of concrete nouns. *NeuroImage*, 62(1):451–463, August 2012. ISSN 1095-9572. doi:10.1016/j.neuroimage.2012.04.048.

Samuel Gibbon, Adam Attaheri, Á. N. Choisdealbha, Sinead Rocha, Perrine Brusini, Natasha Mead, Panagiotis Boutris, Helen Olawole-Scott, Henna Ahmed, Sheila Flanagan, Kanad Mandke, M. Keshavarzi, and U. Goswami. Machine learning accurately classifies neural responses to rhythmic speech vs. non-speech from 8-week-old infant EEG. *Brain and Language*, 220, 2021.

Marzieh Zare, Zahra Rezvani, and April A. Benasich. Automatic classification of 6-month-old infants at familial risk for language-based learning disorder using a support vector machine. *Clinical Neurophysiology*, 127(7):2695–2703, July 2016. ISSN 1388-2457. doi:10.1016/j.clinph.2016.03.025. URL https://www.sciencedirect.com/science/article/pii/S1388245716300074.

Lauren L. Emberson, Benjamin D. Zinszer, Rajeev D. S. Raizada, and Richard N. Aslin. Decoding the infant mind: Multivariate pattern analysis (MVPA) using fNIRS. *PLOS ONE*, 12(4):e0172500, April 2017. ISSN 1932-6203. doi:10.1371/journal.pone.0172500. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0172500.

Larry Fenson. *MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. Cengage Learning, July 1993. ISBN 9781565932593. Google-Books-ID: A9iGKgAACAAJ.

Ruth Tincoff and Peter W. Jusczyk. Some Beginnings of Word Comprehension in 6-Month-Olds. *Psychological Science*, 10(2):172–175, March 1999. ISSN 0956-7976. doi:10.1111/1467-9280.00127. URL https://doi.org/10.1111/1467-9280.00127.

Elika Bergelson and Daniel Swingley. Early Word Comprehension in Infants: Replication and Extension. *Language learning and development : the official journal of the Society for Language Development*, 11(4):369–380, 2015. ISSN 1547-5441. doi:10.1080/15475441.2014.979387. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4671511/.

Elika Bergelson and Daniel Swingley. Young Infants' Word Comprehension Given an Unfamiliar Talker or Altered Pronunciations. *Child development*, 89(5):1567–1576, September 2018. ISSN 0009-3920. doi:10.1111/cdev.12888. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5741549/.

Walter Schneider, Amy Eschman, and Anthony Zuccolotto. *E-Prime: User's Guide. Reference Guide. Getting Started Guide*. Psychology Software Tools, Incorporated, 2002. URL http://www.pstnet.com/eprime.cfm.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. January 2013b.

Yu-Ping Ruan, Zhen-Hua Ling, and Yu Hu. Exploring Semantic Representation in Brain Activity Using Word Embeddings. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 669–679, Austin, Texas, November 2016. Association for Computational Linguistics. doi:10.18653/v1/D16-1064. URL https://www.aclweb.org/anthology/D16-1064.

Sasa L. Kivisaari, Marijn van Vliet, Annika Hultén, Tiina Lindh-Knuutila, Ali Faisal, and Riitta Salmelin. Reconstructing meaning from bits of information. *Nature Communications*, 10(1):927, February 2019. ISSN 2041-1723. doi:10.1038/s41467-019-08848-0. URL https://www.nature.com/articles/s41467-019-08848-0.

Brian Murphy, Massimo Poesio, Francesca Bovolo, Lorenzo Bruzzone, Michele Dalponte, and Heba Lakany. EEG decoding of semantic category reveals distributed representations for single concepts. *Brain and Language*, 117 (1):12–22, April 2011. ISSN 0093-934X. doi:10.1016/j.bandl.2010.09.013. URL https://www.sciencedirect.com/science/article/pii/S0093934X10001811.

Chris Foster, Chad C. Williams, Olave E. Krigolson, and Alona Fyshe. Using EEG to decode semantics during an artificial language learning task. *Brain and Behavior*, jun 2021. doi:10.1002/brb3.2234.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Markus Ojala and Gemma C. Garriga. Permutation Tests for Studying Classifier Performance. *Journal of Machine Learning Research*, 11(62):1833–1863, 2010. URL http://jmlr.org/papers/v11/ojala10a.html.

Eric Maris and Robert Oostenveld. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190, August 2007. ISSN 0165-0270. doi:10.1016/j.jneumeth.2007.03.024.

D. H. Rakison and D. Poulin-Dubois. Developmental origin of the animate-inanimate distinction. *Psychological Bulletin*, 127(2):209–228, March 2001. ISSN 0033-2909. doi:10.1037/0033-2909.127.2.209.

John E. Opfer and Susan A. Gelman. Development of the animate-inanimate distinction. pages 213–238. Wiley-Blackwell, 2011. ISBN 9781405191166.

Peter W. Jusczyk and Richard N. Aslin. Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29(1):1–23, 1995. ISSN 1095-5623. doi:10.1006/cogp.1995.1010.

Janet F. Werker and Richard C. Tees. The organization and reorganization of human speech perception. *Annual Review of Neuroscience*, 15:377–402, 1992. ISSN 1545-4126. doi:10.1146/annurev.ne.15.030192.002113.

Jeff Mielke. A phonetically based metric of sound similarity. *Lingua*, 122(2):145–163, January 2012. ISSN 0024-3841. doi:10.1016/j.lingua.2011.04.006. URL https://www.sciencedirect.com/science/article/pii/S0024384111000891.

Felix Hill, Antoine Bordes, Sumit Chopra, and Jason Weston. The Goldilocks Principle: Reading Children's Books with Explicit Memory Representations. *arXiv:1511.02301 [cs]*, April 2016. URL http://arxiv.org/abs/1511.02301. arXiv: 1511.02301.

B MacWhinney. The childes project: Tools for analyzing talk. Third edition. *Mahwah, NJ: Lawrence Erlbaum Associates*, 2000.

Manuela Friedrich and Angela D. Friederici. Phonotactic knowledge and lexical-semantic processing in one-year-olds: Brain responses to words and nonsense words in picture contexts. *Journal of Cognitive Neuroscience*, 17(11): 1785–1802, 2005. ISSN 1530-8898(Electronic),0898-929X(Print). doi:10.1162/089892905774589172.