Journal of Neuroscience Methods 403 (2024) 110036

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Decoding speech information from EEG data with 4-, 7- and 11-month-old infants: Using convolutional neural network, mutual information-based and backward linear models

Mahmoud Keshavarzi [*], Áine Ní Choisdealbha, Adam Attaheri, Sinead Rocha, Perrine Brusini [1], Samuel Gibbon, Panagiotis Boutris, Natasha Mead, Helen Olawole-Scott, Henna Ahmed, Sheila Flanagan, Kanad Mandke, Usha Goswami

Centre for Neuroscience in Education, Department of Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK

ARTICLE INFO

Keywords:
EEG
Infant
Speech decoding
Backward linear model
Convolutional neural network
Mutual information

ABSTRACT

Background: Computational models that successfully decode neural activity into speech are increasing in the adult literature, with convolutional neural networks (CNNs), backward linear models, and mutual information (MI) models all being applied to neural data in relation to speech input. This is not the case in the infant literature.

New method: Three different computational models, two novel for infants, were applied to decode low-frequency speech envelope information. Previously-employed backward linear models were compared to novel CNN and MI-based models. Fifty infants provided EEG recordings when aged 4, 7, and 11 months, while listening passively to natural speech (sung or chanted nursery rhymes) presented by video with a female singer.

Results: Each model computed speech information for these nursery rhymes in two different low-frequency bands, delta and theta, thought to provide different types of linguistic information. All three models demonstrated significant levels of performance for delta-band neural activity from 4 months of age, with two of three models also showing significant performance for theta-band activity. All models also demonstrated higher accuracy for the delta-band neural responses. None of the models showed developmental (age-related) effects.

Comparisons with existing methods: The data demonstrate that the choice of algorithm used to decode speech envelope information from neural activity in the infant brain determines the developmental conclusions that can be drawn.

Conclusions: The modelling shows that better understanding of the strengths and weaknesses of each modelling approach is fundamental to improving our understanding of how the human brain builds a language system.

1. Introduction

Studies in adult auditory neuroscience have shown that human speech perception relies in part on neural tracking and encoding of the speech amplitude envelope (Giraud and Poeppel, 2012). Further, there are temporal modulation patterns nested in the speech envelope, which are processed at different timescales simultaneously by the adult brain, and which appear to relate to different levels of linguistic information (Ghitza, 2012; Ghitza and Greenberg, 2009; Gross et al., 2013). For the adult brain listening to adult-directed speech (ADS), amplitude modulations in the envelope at frequencies corresponding to the oscillatory theta band (4 – 8 Hz) appear to be particularly important for speech intelligibility (Ghitza, 2012). Further, acoustic "landmarks" (amplitude rise times) in the theta range provide perceptual markers that are critical for intelligibility: if these rise times are removed, speech becomes unintelligible (Doelling et al., 2014). The infant brain also shows neural tracking of the amplitude envelope of speech (Jessen et al., 2019; Kalashnikova et al., 2018; Ortiz Barajas et al., 2021), but whether this tracking is functionally important for speech processing is currently unknown (Jessen et al., 2021). In contrast to adults, studies with children show that for the developing brain, amplitude modulations in the speech envelope that correspond to the oscillatory delta band (1 – 4 Hz)

* Corresponding author.
E-mail addresses: mk919@cam.ac.uk, mahmoud.keshavarzi.ir@ieee.org (M. Keshavarzi).
[1] Current address: Institute of Psychology, Health and Society, University of Liverpool, Waterhouse Building, Block B, Brownlow Street, Liverpool, L69 3GF, UK.

https://doi.org/10.1016/j.jneumeth.2023.110036
Received 18 April 2023; Received in revised form 11 December 2023; Accepted 15 December 2023
Available online 19 December 2023

may govern individual differences in language acquisition (Keshavarzi et al., 2022a; Goswami, 2022). For example, children with developmental dyslexia, who show linguistic impairments regarding the development of phonology (the sound structure of speech), exhibit impaired encoding of speech envelope information between 0 – 2 Hz when listening to sentences (Power et al., 2016). Furthermore, they show atypical delta-band oscillatory synchronisation when listening to stories (Keshavarzi et al., 2022; Molinaro et al., 2016), and atypical phase entrainment in the delta band when presented with rhythmic speech (repetition of the syllable "ba" at a 2 Hz rate; Keshavarzi et al., 2022a; Power et al., 2013). However, phase entrainment in the theta band is not different between children with dyslexia and control children in the rhythmic speech paradigm (Keshavarzi et al., 2022a; Power et al., 2013). Accordingly, delta band speech information may play a critical role in the development of a language system, as individual differences in delta entrainment are those related to language development for child populations. Regarding infants, a recent backward linear modelling study using EEG gathered from 55 infants listening to natural speech is supportive of this possibility (Attaheri et al., 2022a).

In the adult speech reconstruction literature, neural tracking in auditory cortex is dominant in the delta, theta and gamma (35 Hz+) frequency bands, which are thought to yield different types of linguistic information (Giraud and Poeppel, 2012). Theta band cortical tracking identifies the onsets of syllables, which may contribute to speech parsing (Ding and Simon, 2014; Di Liberto et al., 2015; Keshavarzi et al., 2020; Keshavarzi and Reichenbach, 2020). Cortical activity in the delta band tracks phrasal and discourse-level information, contributing to encoding syntactic and semantic information in the speech signal (Broderick et al., 2018; Ding et al., 2016; Weissbart et al., 2020), and gamma band activity is thought to be related to phoneme-level processing (Giraud and Poeppel, 2012). Adult speech reconstruction/decoding studies have relied on a range of computational methods to find the best approximation of the acoustic stimulus from the population of evoked neural activity. Speech reconstruction was originally proposed as a method to study the representational properties of the neural populations, enabling intuitive interpretation given that the cognitive features of the target (the meaning of the speech inputs) were known. Early applications of decoding methods led to novel insights, for example the responses in auditory cortex in the neural theta band were shown to distinguish between individual sentences heard by listening adults (Luo and Poeppel, 2007). Latterly, speech reconstruction studies have progressed to being able to decode the brain activity generated during imagined word recognition (Pei et al., 2011) and even silent reading (Martin et al., 2014). A range of inverse mapping techniques have been employed to find the best approximation of the acoustic stimulus from the population of evoked neural activity (Crosse et al., 2015; Mesgarani et al., 2009; O'Sullivan et al., 2015; Pasley et al., 2012). The early literature was focused on mutual information (MI) and forward linear models (for example, see Gross et al., 2013; Cogan and Poeppel, 2011; Di Liberto et al., 2018a, 2018b; Di Liberto et al., 2015; Haufe et al., 2014; Mesgarani et al., 2014). Although informative and now highly popular, linear models have not offered the quality of speech reconstruction that would be required to build a brain-computer-interface for use by individuals who are unable to communicate as a result of neurological impairments (Anumanchipalli et al., 2019). Accordingly, a range of deep learning methods are currently being utilised for adult speech reconstruction studies, typically via creating recurrent neural network models that either use all the neural frequency bands (delta, theta, beta, alpha, low and high gamma) to estimate the parameters of a speech vocoder directly (Akbari et al., 2019), or recurrent networks that decode neural activity into representations of articulatory movement, and then transform these representations into speech acoustics (Anumanchipalli et al., 2019).

The developmental literature has not kept pace with these technical advances. Despite a recent tutorial regarding the use of linear speech reconstruction methods with infants, contrasting different linear models using EEG data from 10 infants (Jessen et al., 2021), other modelling techniques utilised with adults have yet to be applied to infant data. For example, convolutional neural networks (CNNs) have not yet been applied to infant data prior to our project (the Cambridge UK Baby-Rhythm project, see Gibbon et al., 2021). Studies aiming to find the best approximation of a speech stimulus from the population of evoked neural activity with children and infants typically use either simple cross-correlation techniques to estimate speech envelope tracking (in which amplitude peaks in the speech envelope are correlated with peaks in the broadband neural response (Ortiz et al., 2021; Abrams et al., 2009; Power et al., 2012), or linear modelling techniques in which the speech stimulus is used to estimate the neural response (Jessen et al., 2019; Kalashnikova et al., 2018; Power et al., 2016; Di Liberto et al., 2018a, 2018b). Regarding infants, there are some published studies employing linear methods with natural speech, such as Jessen et al. (2019) and Kalashnikova et al. (2018) which both used forward linear models with 7-month-old infants learning either German (55 infants) or English (12 infants), Attaheri et al. (2022a) which used a backward linear model with 55 English-learning infants, and Jessen et al. (2021) which applied both forward and backward linear models to EEG data from 10 German-learning infants. There is also an EEG study with newborn infants exposed to either French, English or Spanish (Ortiz et al., 2021), however this study only measured neural tracking of the broadband speech envelope via a simple cross-correlation method. All these studies concluded that neural tracking of natural speech had been demonstrated for infants. However, none of these prior studies included longitudinal data to assess potential developmental changes. Further, nuanced linguistic conclusions were not possible, as analyses were based on the broadband speech envelope, with no consideration of the different frequencies nested in the envelope, which are thought to contain different types of linguistic information.

In the current report, we apply novel speech decoding methods regarding infant neural data (CNN; MI), and also apply a backward linear model as in Attaheri et al. (2022a), but selecting different portions of data for testing and training the model and different parameters, in order to see whether such choices affect model outcomes. In each case we compute and compare speech information in two neural frequency bands (delta and theta). The backward linear and CNN models were chosen because infant data is typically noisy and recording sessions must be relatively short, thus it may be useful for researchers to know if either the deep learning or the linear approach is superior for working with such data. Whereas these two approaches aim to reconstruct the stimulus envelopes from the neural data, the MI uses the actual stimulus envelopes and neural responses to estimate the amount of common information between them. Furthermore, the MI model does not require any training or testing procedures, parameter-tuning, or mathematical problem optimisation. The primary research questions were whether infants' neural activity represents speech in a form that encodes different kinds of linguistic information (delta-band and theta-band speech envelope information), and whether the different methods for estimating speech encoding will give converging results. Based on our prior infant neural speech research (Attaheri et al., 2022a; Di Liberto et al., 2023), we expected that the delta-band model outputs would be significantly more accurate than the theta-band outputs. We also tentatively predicted a developmental decrease with age regarding delta-band speech information, and a developmental increase with age regarding theta-band speech information (Attaheri et al., 2022a; Di Liberto et al., 2023). Although it is not possible to directly compare the models due to their different output units, we anticipate convergent results, with the above predictions holding across models. Furthermore, we expect that infant datasets that rank highly in terms of decoding accuracy in one model, will also rank highly in the other models. This would suggest that the models are all picking up on a common neural feature of speech tracking. Similarly, if the different computational approaches give similar results regarding potential age-related changes in representing speech information, this will enable reliable conclusions to

be drawn about the earliest developmental factors involved in creating a linguistic brain. Any divergence in results may indicate that one model outperforms the others, or that they are each using different features to decode the neural speech response.

It should be noted that the analysis of high frequency bands such as the EEG gamma band data can be technically challenging due the high level of noise in infant data, hence we focused on the frequency ranges that are more readily accessible and reliably measured in this age group. Our concentration on delta and theta frequencies allows us to investigate neural processes that are more relevant to the initial stages of language acquisition and comprehension, which may not involve gamma oscillations.

2. Methods

2.1. Participants

EEG was recorded from 50 infants participating in a longitudinal study investigating the relation between neural rhythmic entrainment and language acquisition. There were no exclusion criteria for this community convenience sample, and the 50 participants included were those whose EEG data was available for modelling when this study began. The study was reviewed by the Psychology Research Ethics Committee of the University of Cambridge. Parents gave written informed consent after a detailed explanation of the study and families were repeatedly reminded that they could withdraw from the study at any point during the repeated appointments. For the current modelling, some participants' EEG recordings were excluded due to missed appointments or unusable data. Thirty-five infants provided data at all three timepoints, eleven at two timepoints, and four at one timepoint only. Forty-three infants were included in the 4-month sample, 42 in the 7-month sample, and 45 in the 11-month sample. Reasons for data exclusion include technical issues (e.g. stimulus information not marked in the EEG file), infants sitting for fewer than 2 repetitions of the nursery rhymes, and infants providing too few trials after preprocessing (fewer than half of the 83 nursery rhymes phrases).

2.2. Acoustic stimuli and materials

83 nursery rhyme phrases such as "Baa baa black sheep have you any wool?" with a sampling rate of 4800 Hz were analysed as linguistic stimuli for each participant. The length of each stimulus was between 3.5 s and 6 s (mean length ± SD: 4.23 s ± 0.88).

2.3. EEG signal acquisition, EEG signal processing and acoustic stimuli pre-processing

Parents were seated in an electrically shielded room and either held their infants (4-month recordings) or the infant was seated in an infant chair. Both infant and parent were presented acoustic stimuli and EEG data were collected concurrently using a 64 channel EGI Geodesic Sensor Net system. We excluded 4 facial electrodes, leaving us 60 EEG channels for the analyses. The sampling rate for the data acquisition was 1000 Hz. The MATLAB EEGLAB toolbox (Delorme and Makeig, 2004) was then used to pre-process EEG data. For each participant, EEG data were band-pass filtered between 0.5 Hz and 45 Hz (using a zero phase FIR filter, low cutoff (−6 dB): 0.25 Hz, high cutoff (−6 dB): 45.25 Hz). Probability and kurtosis (built-in functions available in EEGLab toolbox) were used to detect bad channels and were interpolated if they were 3 SD away from the average. The data was then referenced to the global average (of the 60 channels) and epoched into the 83 individual trials. Bad channel detection and interpolation were again performed per epoch. The average number of interpolated channels were approximately 7 for 4 months, 6 for 7 months, 7 for 11 months. The data was next band-pass filtered to extract either delta band (using a zero phase IIR filter, low cutoff (−3 dB): 1 Hz, high cutoff (−3 dB): 4 Hz, order: 6) or theta band (using a zero phase IIR filter, low cutoff (−3 dB): 4 Hz, high cutoff (−3 dB): 8 Hz, order: 6). On the other hand, stimuli envelopes were computed as the absolute value of the analytical signal (that was obtained through the Hilbert transform) of the stimuli. It should be noted that the amplitude envelope was only calculated for the auditory stimuli (nursery rhymes phrases). The envelopes were then band-pass filtered in frequency range 1 – 8 Hz (using a zero phase IIR filter, low cutoff (−3 dB): 1 Hz, high cutoff (−3 dB): 8 Hz, order: 6). Both EEG and stimuli envelopes were downsampled to 50 Hz.

3. Computational models

Here three computational models were used to analyse the data:

3.1. Backward linear model

The first model was based on a linear mapping between stimuli and neural responses. This model reconstructs the stimuli envelopes using the backward linear model (Crosse et al., 2016) which is given by:

$$\widehat{s}(t) = \sum_n \sum_\tau r(t+\tau, n) g(\tau, n) \tag{1}$$

where $\widehat{s}(t)$ is the reconstructed stimulus, $r(t+\tau, n)$ is the neural response at channel n and time lag τ, $g(\tau, n)$ is a decoder representing the linear mapping from the neural response to the corresponding stimulus for time lag τ and channel n. The decoder was also estimated by minimizing the mean square error between actual and reconstructed stimuli.

The minimum and maximum time lags were $\tau_{\min} = -100ms$ and $\tau_{\max} = 300ms$, respectively. The validation approach was the "leave-one-out" cross-validation (using mTRFcrossval function from the mTRF Toobox, Crosse et al., 2016) in which each trial (stimulus-response) is "left out" or used for testing and the remainder are used to train the model and this procedure is repeated across all trials. This validation approach was exclusively employed to determine the optimal ridge parameter (λ) from a range of candidates ($\lambda = 10^0, 10^1, \ldots, 10^{10}$). The optimal value for the ridge parameter was selected based on the one that yielded the highest average correlation score during this process. Subsequently, this identified optimal value was utilised to train the model. Here about 80% of data were used to train the model and the rest of data was employed for testing it. The decision to use 80% of the data for training the model was based on a considered trade-off between model performance and data preservation. Allocating too much data for training might increase the risk of overfitting, where the model becomes overly specialised to the training data and performs poorly on new, unseen data. By reserving 20% of the data for testing, we aimed to achieve a balance that helps prevent overfitting while still allowing the model to learn effectively. After reconstructing stimuli for participant j, the average correlation score $Corr_{av}^{(j)}$ was calculated by:

$$Corr_{av}^{(j)} = \frac{1}{M_j} \sum_{m=1}^{M_j} \rho_m^{(j)} \tag{2}$$

where $\rho_m^{(j)}$ is the correlation value between the mth reconstructed stimulus and the mth actual stimulus for the participant j. It should be noted that we set $\rho_m^{(j)}$ to zero if the correlation between the mth reconstructed envelope and the corresponding actual envelope was negative. Infant EEG data is known to be particularly susceptible to various sources of noise such as movement artifacts. Negative correlations in our analysis could potentially be attributed to such noise rather than representing meaningful relationships between the EEG data and the speech envelope. By setting these negative correlations to zero, we aimed to focus our analysis on the stronger, more reliable associations, thus enhancing the quality and robustness of our results.

3.2. CNN

Over the past few decades, artificial neural networks (ANN) have been widely used to achieve significant improvements in many tasks in vision, hearing, neuroscience, and language domains. ANNs can be generally categorised into three main architectures (Keshavarzi et al., 2018): feed-forward deep neural networks, recurrent neural networks, and CNN. Among these, CNNs have achieved the best performance to process two-dimensional data such as image and EEG/MEG data in tasks like recognition, segmentation, detection, and retrieval (Karpathy et al., 2014).

The CNN here consisted of three main layers: (1) Two-dimensional (2D) convolutional layer (with 30 filters of size [4 4] and a stride of [2 2], ReLU activation function and dropout); (2) max pooling layer (with pool size [2 2] and stride [2 2]); (3) fully connected layer (with 75 units). Fig. 1 shows the schematic diagram of the CNN algorithm. Here we consider EEG data as a series of 2D matrices, where each matrix serves as a snapshot capturing the brain's electrical activity. Within these matrices, one dimension represents the progression over time, creating a timeline of recorded brain signals. Simultaneously, the other dimension corresponds to the various EEG channels. This transforms the EEG data into an image-like structure, where each single point in time represents a spatially interconnected pattern of electrical activity. This conceptual framework is very effective as it allows us to explore the spatial correlations that dynamically exist between EEG channels over the recorded time interval. We chose 2D CNN models for our EEG data because they can effectively capture both temporal and spatial information. The term "spatial" in this context is specifically related to the 2D matrices which represent the EEG data. 2D CNNs excel in capturing spatial dependencies, a crucial aspect of EEG data for predictive performance. These models have a track record of success in EEG-related tasks, including brain-computer interfaces and EEG analysis (such as Schirrmeister et al., 2017; Lawhern et al., 2018; Gibbon et al., 2021).

Our primary goal was to estimate the speech envelope using the model applied to EEG data. In this task, we aimed to capture temporal dependencies and patterns across the EEG channels to achieve accurate estimations. We used an iterative approach to fine-tune our model parameters. We particularly explored various hyperparameter settings, such as network architecture, filter sizes, batch sizes, number of epochs, pooling strategies, and learning rates. While our focus in this study was on tuning these crucial parameters, consistent with previous studies, the impact of channel order was not investigated. However, it is important to note that factors such as selection of channels in a specific order may capture particular relationships and localised features within the EEG data. Accordingly, different channel orders could potentially lead to different spatial representations, which could be explored in future work. A MATLAB toolbox named "Deep Learning Toolbox" was used to construct, train, and test the CNN. The resilient back- propagation algorithm "RMSprop" (Riedmiller and Braun, 1993) was employed as the optimizer function to minimize the mean square error in the training algorithm. The learning rate was also initialized to 0.01 and decreased by a factor of 0.9 after each training run (a run was based on using whole training data once, Keshavarzi et al., 2018). The batch size was 4, and 50 training runs were performed.

The CNN took features (short frames of filtered EEG in this study) extracted from the neural responses as its input and predicted the corresponding stimulus envelop as the target. The actual stimulus envelope (as the output of network) was segmented into frames with a duration of 1.5 s (75 samples) and with no overlapping between successive frames. In pilot work, it was found that CNN had a better performance to estimate stimuli envelopes when the frame duration was 75 samples as compared to 25 samples, 50 samples, 100 samples. The neural responses (as the input of network) were also windowed into frames with the duration of 2 s (100 samples) and an overlap of 25% (25 samples) between successive frames. The purpose of this overlapping was to compensate the delay between the stimulus and its corresponding

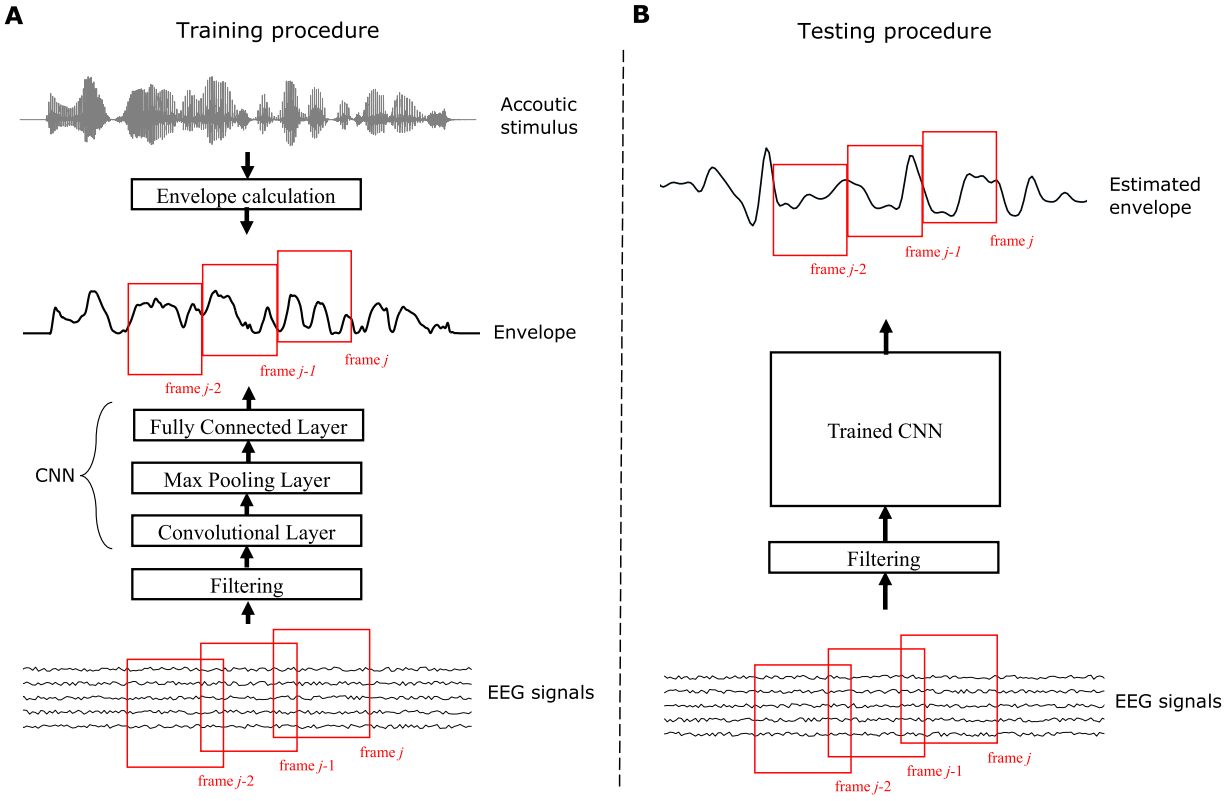

Fig. 1. Schematic diagram of the CNN algorithm and the envelope-reconstruction framework. Panels A and B show the training and testing procedures, respectively. The network consists of three main layers. It takes frames of pre-processed neural responses and predicts the envelope of the corresponding acoustic stimulus.

response.

The EEG data for each participant was divided into training, validation, and testing datasets. Accordingly, 70% of data were used to train the network, 10% was employed for validation, and the remaining 20% for testing the network. Again, the choice of these percentages was based on the trade-off between model performance and the preservation of data. Finally, the average correlation score was calculated using Eq. (2). It is important to note that these percentages were applied separately for data from each infant, and a single model was built separately for each infant.

3.3. MI

The MI between two random variables is defined as a measure of the amount of information that one random variable contains about another random variable (Cover, 1999). It is nonlinear, non-negative, and is zero if and only if the variables are statistically independent. The MI between two random variables $R = \{r_1, r_2, \ldots, r_T\}$ and $S = \{s_1, s_2, \ldots, s_T\}$ is mathematically given by (Cover, 1999):

$$I(S;R) = \sum_{r\in R}\sum_{s\in S} P(s,r)\log\frac{P(s,r)}{P(s)P(r)} \tag{3}$$

where $P(s)$ and $P(r)$ are the marginal distributions of variables S and R, respectively, and $P(s,r)$ is the joint distribution of these variables. Here marginal and joint distributions were estimated using the Gaussian kernel estimator (Qiu et al., 2009):

$$P(s) = \frac{1}{\sqrt{2\pi b^2}T}\sum_{t=1}^{T} e^{-\frac{1}{2b^2}(s-s_t)^2} \tag{4}$$

$$P(r) = \frac{1}{\sqrt{2\pi b^2}T}\sum_{t=1}^{T} e^{-\frac{1}{2b^2}(r-r_t)^2} \tag{5}$$

$$P(s,r) = \frac{1}{\sqrt{2\pi b^2}T}\sum_{t=1}^{T} e^{-\frac{1}{2b^2}[(s-s_t)^2+(r-r_t)^2]} \tag{6}$$

where b is called bandwidth which acts as the parameter tuning the kernel function (Qiu et al., 2009). As the MI is an expectation value with respect to joint distribution of S and R, it can be estimated using data samples drawn from densities (Thomas et al., 2014):

$$\widehat{I}(S;R) = \langle\log\left(\frac{P(s,r)}{P(s)P(r)}\right)\rangle = \frac{1}{T}\sum_{t=1}^{T}\log\left(\frac{P(s_t,r_t)}{P(s_t)P(r_t)}\right) \tag{7}$$

In this study, MI was used as a metric to measure the amount of information that neural responses give about the actual stimuli envelopes. To this end, we obtained the average MI for each individual using three steps: (1) Calculating the MI score between each stimulus and each channel (electrode) of its neural response independently; (2) Calculating the MI score for each pair of stimulus-response by averaging across all channels; (3) Calculating the mean MI by averaging across all pairs of stimulus-response. The mean MI score for participant j was given by:

$$I_{av}^{(j)} = \frac{1}{M_j N_j}\sum_{m=1}^{M_j}\sum_{n=1}^{N_j}\widehat{I}\left(S_m^{(j)};R_{mn}^{(j)}\right) \tag{8}$$

where M_j and N_j are the number of stimuli and the number of EEG channels, respectively, $S_m^{(j)}$ is the mth stimulus, and $R_{mn}^{(j)}$ is the mth neural response at channel n.

3.4. Chance level calculation by participant

To assess whether the accuracy scores obtained from computation methods were above chance, null models were computed separately for each method and for each frequency band. To this end, the neural response data were first permuted across different trials separately for each individual infant, age, band and method, while the stimuli envelopes were kept as before. We then calculated the permuted decoding scores separately for each computational method and for each infant at each age in each frequency band. This procedure was performed 100 times for each infant, age, band and method, yielding 100 random "accuracy" scores corresponding to each respective infant-band-age-method. The mean individual chance level per participant for each infant-band-age-method was finally obtained by averaging across these 100 scores in each case. These real versus random scores for each participant were used as a basis for creating the LMEM models, which included all theoretical factors of interest (age and frequency band) for each model, enabling a complete comparison of the modelling approaches. This approach also allows for each model to be evaluated using a single statistical test, in contrast to the more typical group-level approach (see below) which requires a separate test for each combination of age and frequency band.

3.5. Chance level calculation by group

To evaluate the group statistical significance of decoding accuracy for each band-age-method combination, we again calculated the null distribution associated to each such combination. We determined the critical decoding score within each null model, corresponding to a significance level of $p = 0.05$. This score, representing the boundary for chance in each case, was compared to the average group decoding scores obtained from the actual data for that particular age-band combination. This enabled us to compare group-level performance (4-months, 7-months, 11-months) against chance. As noted in Section 3.4, to fully compare the three models, group-level statistical significance was not considered when running the participant-level LMEM analyses, as the LMEM models utilise random data generated for each participant.

3.6. Statistical analysis

Linear mixed effects regressions were run on the output values from each model, using both the real data and the matched chance-level values that were generated for each participant (see Section 3.4). These tests examined the effects of frequency band (delta or theta), age (4, 7 or 11 months) and data type (real or chance level) on the decoding accuracy values of each model, as well as the interactions between each of these factors, at the level of individual infants rather than the group. There was a random intercept on participant identity and a random slope on age. The regressions were performed in R using the lmerTest package (Kuznetsova et al., 2017).

We also examined similarities between the decoding models. The output values from each model are not directly comparable because they rely on different principles. Nonetheless, if the models are using the same features of the data to decode it, we would expect to find that the infant whose data had the highest decoding values in one model, would also be among the highest values in the other models. Likewise, the infant with the lowest decoding value in one model would also have low decoding values in other models. To test this, we used Spearman rank order correlations. We examined each model pairing and each age group separately. To limit the number of comparisons, we used delta band values only, as we hypothesised that this band would have the highest reconstruction values in the linear mixed effects regressions (see Attaheri et al., 2022a).

4. Results

Linear mixed effects regressions were used to analyse the participant-level data (see Section 3.4) using lme4 and lmerTest in R software. As this is hypothesis-driven work, all factors expected to have an effect on the results were included in these linear mixed effects models, namely age, frequency band, and data type (real or chance level), as well as their interactions, with random intercepts on participant identity. Base cases for all models were the random data, the theta band, and the four-month

age group. Three models were run, one for each decoding approach. Before reporting the simple effects of these models, we report the results of Satterthwaite-corrected ANOVAs run on the models, to show whether each variable made a significant contribution to each of the three models. These results are reported in Table 1. Also reported in Table 1 are the results of Chi-square tests comparing whether the statistical models describe the data significantly better than an equivalent model containing only random effects (i.e. the intercept on participant identity). As can be seen, all Chi-square tests were significant, indicating that all 3 models performed better than a random effects-only model. Both the type of data (real or chance-level) and the frequency band (delta vs theta) significantly affected model fit.

4.1. Backward linear model

The backward linear model reconstructed the stimulus envelopes (1 – 8 Hz) using the neural responses filtered in either a 1 – 4 Hz band or a 4 – 8 Hz band, henceforth "delta" and "theta" bands respectively. Please note that the delta band filter previously used with these infants in Attaheri et al. (2022a) was 0.5 – 4 Hz. To assess the statistical significance of decoding accuracy by group for each band-age separately, we computed the null distribution specific to each band-age using the permuted data (see Section 3.5). Subsequently, we determined the correlation value in the null model corresponding to $p = 0.05$ (see Section 3.5). The statistical significance for each band-age is presented in Fig. S1 (see Supplementary Information). The results revealed that at the group level, decoding accuracy was significantly higher than chance only for the 4-month and 11-month age groups in both the delta and theta bands. The 7-month data did not exceed chance values.

For a complete factorial comparison of all models, we retained the 7-month data for the participant-level LMEM analyses. We selected different portions of data for testing and training the model. The resulting correlation values provide an estimate of how accurately the stimulus envelope could be reconstructed from the neural response, corresponding to the correlation between the filtered envelopes of the nursery rhyme stimuli and the reconstructed envelopes. To obtain these accuracy scores, up to six backward linear models were created for each infant separately – one for each of the frequency bands, at each of three different timepoints – 4 months, 7 months, and 11 months of age. The data are shown in Fig. 2 and the model estimates in Table 2. In Fig. 2, the horizontal red lines illustrate the mean chance level obtained from the permutation test (as described in Section 3.4) for each age and each band. As can be seen, at the participant level the average accuracy score obtained from the real data is higher than the average chance level at all ages tested, including the 7-months infants.

The model estimates (see Table 2) indicate that overall, the delta band data showed significantly higher decoding values than the theta band data, $p < 0.0001$. There was also an interaction between band and data type ($p = 0.023$), showing that real data were decoded more accurately than random data, but only in the delta band. Accordingly, the backward linear model could successfully decode the envelope from the EEG data, but only from the delta band EEG and not from the theta band EEG. Taken together with the group analyses (Fig. S1 in Supplementary Information), the conservative conclusion is that delta-band decoding using an mTRF approach is only reliable at 4 and 11 months.

Table 1
Satterthwaite-corrected ANOVA results illustrating whether each factor made a significant contribution to the statistical model. Final row shows results of chi-square test comparing the statistical model to a random effects-only model.

Variable	Backward Linear	CNN	MI
Age	F(2, 54) = 3.13	F(2, 45) = 0.714	F(2, 46) = 1.397
Band	F(1, 423) = 974.21***	F(1, 381) = 70.05***	F(1, 381) = 14,1150***
Type (real or chance)	F(1, 423) = 22.68***	F(1, 381) = 56.55***	F(1, 381) = 561.95***
Age * Band	F(2, 423) = 2.05	F(2, 381) = 0.28	F(2, 381) = 16.392***
Age * Type	F(2, 423) = 1.51	F(2, 381) = 3.03	F(2, 381) = 0.514
Band * Type	F(1, 423) = 5.99*	F(1, 381) = 0.33	F(1, 381) = 167.92***
Age*Band*Type	F(2, 423) = 0.57	F(2, 381) = 0.02	F(2, 381) = 0.109
Chi-square	561.21***	120.16***	2759***

** $p < 0.01$.
* $p < 0.05$.
*** $p < 0.001$

4.2. CNN

The second algorithm applied to the data was a CNN. The target outcome for the CNN was to predict the envelopes of the stimuli filtered to 1 – 8 Hz. The input was the neural responses filtered in either the delta (1 – 4 Hz) or theta (4 – 8 Hz) band. The correlation between the actual envelopes and those estimated from each infant's neural data was calculated for both delta and theta bands at all three time-points, and for both real and random data (see Fig. 3 and Section 3.4). To assess the statistical significance of decoding accuracy for each band-age grouping separately, we again computed the null distribution specific to each band-age using the permuted data (see Section 3.5). The statistical significance for each band-age is presented in Fig. S2 (see Supplementary Information). For the CNN models, decoding accuracy was significantly higher than chance for all band-age combinations excepting the 4-month delta band data.

The results of the linear mixed effects model at the participant level (Section 3.4) are reported in Table 3. There was a significant effect of data type, indicating that the CNN decoded the EEG data significantly better than the randomly permuted data, that is, it was decoded at an above-chance level ($p = 0.047$, see Table 3). This was true across the sample, as the 4-month data were used as the base case in this comparison and although the estimates on the interactions between 7-month and 11-month age groups and the real data type were positive (indicating a better performance relative to chance at these ages), they were non-significant ($p = 0.915$ and $p = 0.095$ respectively). There was a significant effect of frequency band, with the model showing higher decoding values for the delta band than for the theta band data ($p = 0.0005$). There was also an interaction between delta band and real data, $p = 0.023$. This shows that the CNN produces higher decoding accuracy scores for the delta band compared to the theta band, for the real data only. Whereas the backward linear model was only able to decode delta band data at an above-chance level, the CNN decodes both delta and theta at an above chance level – and is also more accurate for delta than for theta. Taken together with the group analyses (Fig. S2 in Supplementary Information), the conservative conclusion is that delta-band decoding of the speech signal using a CNN approach is reliable at 7 and 11 months, while theta-band decoding is reliable using a CNN approach at all ages.

4.3. MI

Finally, we calculated the MI between the actual stimulus envelopes filtered in the frequency range of 1 – 8 Hz and the neural responses filtered in either the delta (1 – 4 Hz) or theta (4 – 8 Hz) band for the 4-, 7-, and 11-month infants respectively. To assess the statistical significance of decoding accuracy for each band-age grouping separately, we again computed the null distribution specific to each band-age using the permuted data (see Section 3.5). The statistical significance for each band-age is presented in Fig. S3 (see Supplementary Information). The results revealed that decoding accuracy in the MI model was significantly greater than chance at the group level for all band-age combinations.

The LMEM ANOVAs (see Table 4) for the MI model showed higher decoding values for real than random data overall (see Fig. 4), indicating that the model could decode both delta and theta band information

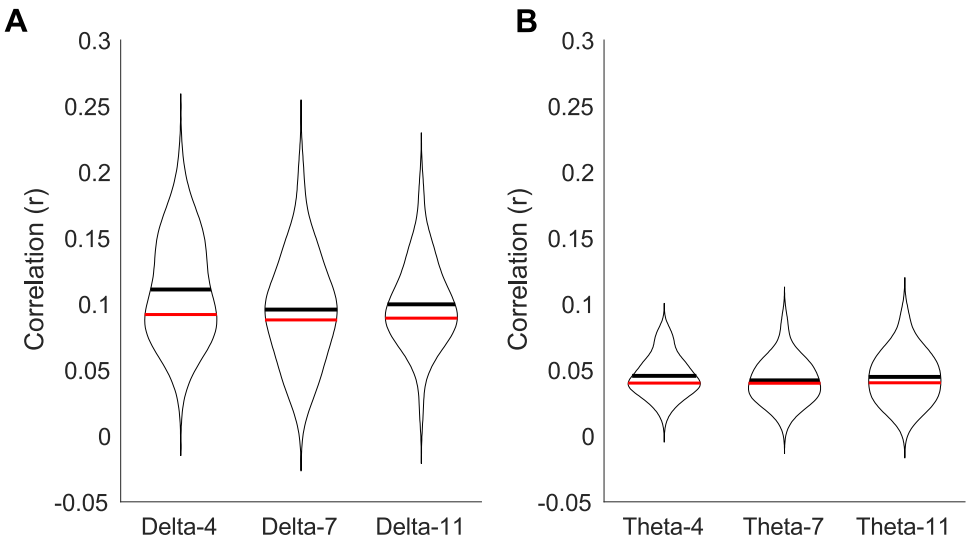

Fig. 2. The backward linear model accuracy scores for the three ages 4-, 7- and 11-month infants and for both delta and theta bands. The violin plots show the mean accuracy scores (Pearson correlation values). The horizontal black lines denote the average values and horizontal red lines denoted the mean chance level values. Please note that the discrepancy between the tail of the violin plot appearing below zero while all the scores are positive is a result of the visualization method used and does not indicate negative values in the data. Violin plots are constructed by mirroring and stacking density plots, which show the distribution of data values. In our case, there were no negative values in the dataset. The appearance of the tail below zero is attributable to the way the density estimation is presented in the plot.

Table 2
The results of the mixed effect linear regression for the backward linear model.

Variable	Beta coefficient	Standard error	*t*-value	*p*-value
7 months (vs 4 months)	0.00003	0.004	0.006	0.995
11 months (vs 4 months)	0.0004	0.004	0.091	0.928
Delta (vs Theta)	**0.052**	**0.004**	**12.248**	**<0.0001**
Real (vs Rand)	0.005	0.004	1.278	0.202
7 months*Delta	-0.004	0.006	-0.656	0.512
11 months*Delta	-0.003	0.006	-0.503	0.615
7 months*Real	-0.003	0.006	-0.54	0.59
11 months*Real	-0.001	0.006	-0.172	0.863
Delta*Real	**0.014**	**0.006**	**2.275**	**0.023**
7 months*Delta*Real	-0.008	0.009	-0.94	0.348
11 months*Delta*Real	-0.008	0.008	-0.902	0.368
Intercept	0.04	0.003	12.919	<0.0001

Table 3
The results of the mixed effect linear regression for the CNN model.

Variable	Beta coefficient	Standard error	*t*-value	*p*-value
7 months (vs 4 months)	-0.0006	0.006	-0.108	0.914
11 months (vs 4 months)	-0.002	0.006	-0.312	0.755
Delta (vs Theta)	**0.02**	**0.006**	**3.492**	**0.0005**
Real (vs Rand)	**0.012**	**0.006**	**1.996**	**0.047**
7 months*Delta	-0.002	0.008	-0.261	0.794
11 months*Delta	-0.003	0.008	-0.421	0.674
7 months*Real	0.0009	0.008	0.107	0.915
11 months*Real	0.014	0.008	1.672	0.095
Delta*Real	**0.003**	**0.006**	**2.275**	**0.023**
7 months*Delta*Real	0.0002	0.012	0.015	0.988
11 months*Delta*Real	-0.002	0.011	-0.156	0.876
Intercept	0.101	0.004	23.126	<0.0001

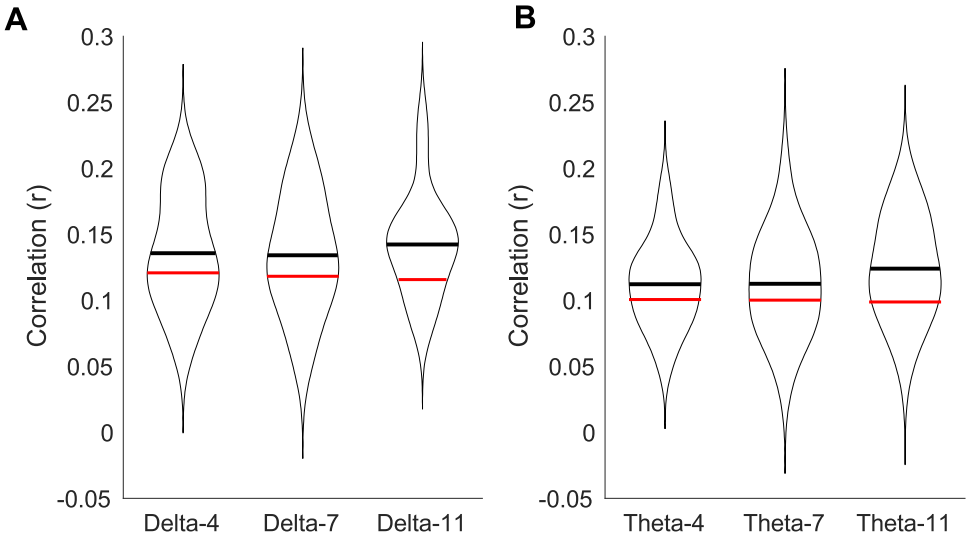

Fig. 3. The CNN accuracy scores for the three ages 4-, 7- and 11-month infants and for both delta and theta bands. The violin plots show the mean accuracy scores (Pearson correlation values). The horizontal black lines denote the average values and horizontal red lines denote the mean chance level values.

Table 4
The results of the mixed effect linear regression for the MI model.

Variable	Beta coefficient	Standard error	*t*-value	*p*-value
7 months (vs 4 months)	-0.0005	0.0006	-0.78	0.437
11 months (vs 4 months)	-0.0009	0.0007	-1.305	0.294
Delta (vs Theta)	**0.071**	**0.0005**	**145.354**	**<0.0001**
Real (vs Rand)	**0.002**	**0.0005**	**4.145**	**<0.0001**
7 months*Delta	0.0008	0.0007	1.12	0.264
11 months*Delta	**0.002**	**0.0007**	**3.635**	**0.0003**
7 months*Real	0.00007	0.0007	0.108	0.915
11 months*Real	0.0003	0.0007	0.379	0.705
Delta*Real	**0.005**	**0.007**	**7.243**	**<0.0001**
7 months*Delta*Real	0.00004	0.001	0.036	0.971
11 months*Delta*Real	0.0004	0.001	0.421	0.674
Intercept	0.148	0.0005	290.656	<0.0001

accurately. The model showed both an overall effect of delta versus theta band ($p < 0.0001$), as well as an interaction between delta band and real data, which shows that significantly higher decoding accuracy was achieved for delta band EEG data. The significant interaction between the delta band and the eleven-month age group also suggests higher decoding values for delta band information for the older infants but, without an interaction with real (versus random) data, we cannot say that this equates to greater accuracy of decoding at age 11 months. Nonetheless, the MI results overall indicate that this modelling approach could decode both delta and theta band information at an above-chance level, and that accuracy was better for the delta band than for the theta band EEG. Taken together with the group analyses (Fig. S3 in Supplementary Information), which all exceeded chance levels, the MI modelling suggests that infant EEG recorded in response to sung speech yields decoding estimates that are reliably above chance for both the delta and theta frequency bands at all ages.

4.4. Relations between the models

In order to establish whether infant data sets that ranked highly in terms of decoding accuracy in one model would also rank highly in the other models, Spearman rank order correlations for real delta band data were computed and are reported in Table 5 for all models, irrespective of whether the group data were significantly above chance (as these correlations are based on individual participant pairings). The correlations show greater similarity for the rankings for the linear and CNN models compared to the other model pairings. Note that the two linear models rank the 4-month delta band data similarly, even though group decoding values for the CNN model were not significantly above chance. Further, the rankings of the infants with the highest to lowest decoding values produced by the MI model were not significantly related to those of the other models. This suggests that the features used by the MI decoding models were different from the other models. This is interesting, as only the MI model consistently yielded above-chance decoding of delta-band information at all ages studied.

5. Discussion

Here we investigated whether different modelling approaches drawn from the adult speech reconstruction literature would converge on similar results when representing the acoustic stimulus (sung speech) from neural activity measured in infants. The main finding was that although the different computational approaches broadly converged in most respects regarding the representation of delta-band speech information in the infant brain, there was less convergence regarding the representation of theta-band speech information. In accord with our primary research question, all models suggested that infants' neural activity is representing speech information, as real data was significantly different from random data in all three models according to LMEM analyses. Regarding whether infants' neural activity is representing speech information in at least two frequency bands that may encode different kinds of linguistic information (delta and theta), both backward linear and CNN models differed from the MI model. Regarding the two linear models, only the CNN model showed above-chance decoding of theta-band information at all three ages studied. Both models showed above-chance decoding of delta band information, but at different ages (4 and 11 months for the backward linear model, 7 and 11 months for

Table 5
Spearman rank order correlations of real data delta band decoding values by age group and decoding approach.

Age	Linear vs CNN	Linear vs MI	CNN vs MI
4 months	$\rho(41) = 0.4^{**}$	$\rho(41) = 0.212$	$\rho(41) = 0.305^{*}$
7 months	$\rho(40) = 0.151$	$\rho(40) = 0.065$	$\rho(40) = 0.236$
11 months	$\rho(43) = 0.398^{**}$	$\rho(43) = 0.287$	$\rho(43) = 0.109$

*** $p < 0.001$.
* $p < 0.05$.
** $p < 0.01$.

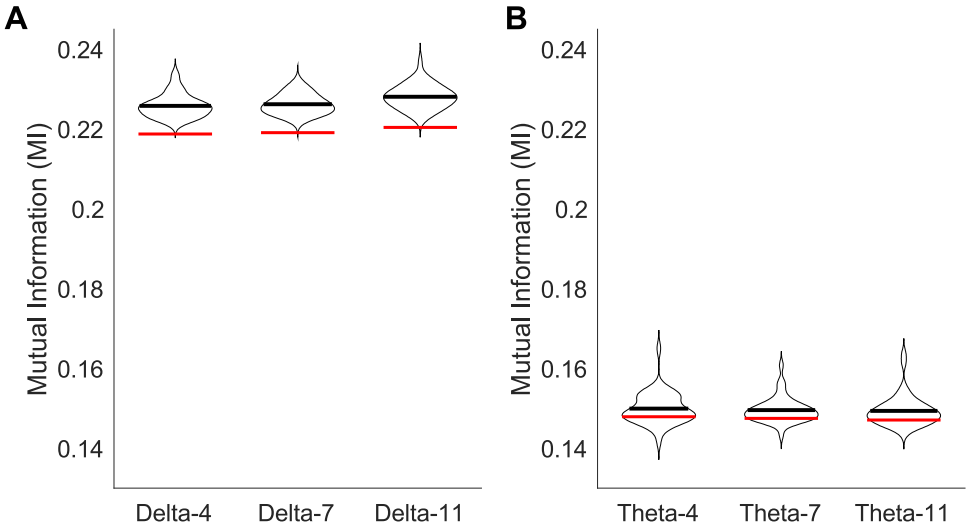

Fig. 4. The MI accuracy scores for the three ages 4-, 7- and 11-month infants and for both delta and theta bands. The violin plots show the mean accuracy scores (Pearson correlation values). The horizontal black lines denote average values and red lines denote the mean chance level values.

the CNN model). In the adult neural speech literature, activity in these two frequency bands is thought to be related respectively to discourse-level information and auditory grouping (delta band), and syllable-level information and speech intelligibility (theta band). In the MI modelling, both delta band and theta band speech information was estimated at above-chance levels at all ages studied. As our infants were pre-verbal, this finding provides confidence to the nascent field of neural studies of language acquisition. It appears that current speech modelling techniques can begin to reveal how the human brain begins to build a language system, particularly when more than one modelling technique is applied to the same data.

Our second research question was whether delta band model outputs would show higher values than theta band model outputs. This question was motivated by earlier computational modelling of infant-directed speech, which found significantly greater modulation energy in a band of amplitude modulation corresponding to the EEG delta band (Leong and Goswami, 2015). Again, there was reasonable convergence between models. All models indicated significantly higher decoding values for the delta band compared to the theta band, however for the backward linear model an interaction effect between band and real versus random data suggested that only the delta band decoding estimates were reliable. This finding differs from Attaheri et al. (2022a), who also applied a backward linear model to the EEG collected from these and some additional participants, including more infants (55 participants). Attaheri et al. (2022a) reported significant decoding values for both delta band and theta band estimates at 4, 7 and 11 months. As noted earlier, in Attaheri et al. (2022a) the delta band was defined as 0.5 – 4 Hz, rather than 1– 4 Hz as here, which could potentially explain this discrepancy. For the current 50 infants, the CNN and MI models clearly showed higher decoding values for the delta band than for the theta band. In both models, both of these estimates were reliably greater for real data compared to chance data, indicating that decoding estimates were significant for both bands. Finally, both the CNN and the MI models yielded model estimates that were significantly greater for the delta band data than for the theta band data. In summary, delta band estimates appear to be most reliable in infant speech EEG studies, replicating Attaheri et al. (2022a).

The tentative predictions concerning developmental effects were not supported by any of the models. Participating infants contributed neural data at three measurement points during their first year of life (4, 7 and 11 months). We had tentatively predicted (following Attaheri et al., 2022a) that over the first year of life, delta band speech information would become less important while theta band information would become more important, possibly because infants were developing better-specified speech-based representations with more detailed encoding of phonology. However, none of the 3 models compared here showed significant effects of age (see Table 1), in contrast to Attaheri et al. (2022a).

Finally, we had proposed that if the three models were picking up on similar features in the EEG related to speech-based encoding, then infant datasets that ranked highly in terms of decoding accuracy in one model should also rank highly in the other models. The rankings (Table 5) suggested that the linear and CNN models were the most similar regarding which EEG features they were selecting. However, for both of these modelling approaches, some of the age-band pairings did not significantly exceed chance values at the group level. For the MI model all age-band pairings exceeded chance, but only one correlation was significant, with the CNN model for the youngest infants ($\rho(41) = 0.305$, 4-month-olds). This is surprising, as the CNN values for the 4-month-olds did not exceed chance performance in the group comparisons (Fig. S2 in Supplementary Information). This contrast between the MI model and the CNN and linear models may be due to the fundamental differences between the approaches, as MI relies on information theory to quantify decoding accuracy and is not involved in stimulus-reconstruction while the backward linear and CNN models are reconstructive approaches. Finally, although all age groups investigated here were pre-verbal, in that they did not yet produce much speech, other studies have shown that by 6 months of age, infants already *comprehend* a surprising number of words (Bergelson and Swingley, 2012). Accordingly, top-down linguistic processes are also likely to be coming online for the 7- and 11-month-olds, which would be reflected in their EEG, creating more features for the models to quantify. Nevertheless, the EEG data collected at 7 months do not show any significant relations across models. This may be explained by particularly noisy data for this age group, which may also explain the non-significant effect for this age group in the mTRF modelling at the group level (Fig. S1 in Supplementary Information).

In summary, all three models converged in demonstrating significant accuracy for decoding speech information in the neural delta band. Both the MI and CNN models decoded both delta and theta information above chance, but surprisingly, no age-related effects were found. Overall, the modelling supports the theoretical view that the neural representation of delta band speech information plays a primary role in developing a language system during the first year of life. This finding is consistent with prior computational modelling of infant-directed speech (IDS) based on a spectral-amplitude modulation phase hierarchy approach (Leong and Goswami, 2015). For English IDS, the modelling demonstrated significantly more modulation energy in the delta band compared to ADS, and this greater energy was consistent across IDS directed to infants of 7, 9 and 11 months of age (Leong et al., 2017). The modelling data presented here suggests that the importance of delta band speech information in the amplitude envelope of IDS is reflected in infant neural encoding of speech.

These findings are also consistent with cognitive behavioural research, which suggests that infants rely on speech rhythm and prosody to begin to build a mental lexicon of word forms, for example using the onsets of stressed syllables (acoustic landmarks which occur on average every 2 Hz across languages, Dauer, 1983) as a clue to word beginnings (Leong et al., 2014; Mehler et al., 1988). The potentially primary role of phrasal-level information is also consistent with current fNIRS and ERP data with infants. Infants can detect prosodic information from birth (Abboub et al., 2016; Fló et al., 2019), and can differentiate native versus non-native prosodic templates from as young as 4 months (Weber et al., 2004). They can also parse words like their own name from connected speech by 4 months of age (Mandel et al., 1995).

The current study has a number of limitations. Firstly, the speech heard by infants was sung or chanted, and this could explain the greater reconstruction accuracy that was found for delta band speech information compared to theta band speech information. Indeed, comparable results regarding the delta band were found for adults who listened to the same sung IDS input (Attaheri et al., 2022b). Sung speech was used because in principle it provides an optimally-structured stimulus for the infant brain, since all the amplitude modulations at different frequencies are temporally aligned with an external beat. However, the speech was thus highly rhythmic, and thereby potentially activated acoustic mechanisms for processing musical or non-speech rhythm in addition to speech rhythm; the former is known to be related to delta band acoustic information (Cirelli et al., 2016). Secondly, a number of modelling assumptions were made during data analysis, and any changes in parameter choice could in principle give different results. This was demonstrated here for the linear model. A third limitation is the relatively small sample of infants ($N = 50$). Although this is a relatively large sample for the infant EEG literature, it would be preferable to apply the same models used here with even larger samples. Nevertheless, although the number of participants was limited, the models were still fitting a substantial amount of neural data. Finally, the data reported here are from an ongoing longitudinal study of neural speech processing in infants, and language outcome data are currently being prepared for analysis. Accordingly, the functional significance of the representation of the acoustic stimulus offered here by the CNN, MI and backward linear models can be assessed in future work.

In conclusion, we show here that the application of sophisticated

computational methods for approximating the acoustic speech signal from evoked neural activity can be successful with infant data. Our analyses suggest accurate representation of speech envelope information in the delta (and likely theta) band by the infant brain from 4 months of age. There were no age-related changes in model estimates, suggesting that the speech processing mechanisms used by the infant brain may be relatively hard-wired (Doelling et al., 2022). Further work is needed to assess whether the models were using similar features earlier in life and diverged as infants developed. Correlating the neural measures obtained here with future language outcome measures could enable calibration of which modelling approach or approaches are best suited to developmental studies of neural speech processing in pre-verbal populations.

CRediT authorship contribution statement

Mahmoud Keshavarzi: Conceptualization, Methodology, Formal Analysis, Visualization, Writing – Original Draft. **Áine Ní Choisdealbha:** Conceptualization, EEG Preprocessing, Investigation, Data Curation, Formal Analysis, Writing – Review & Editing. **Adam Attaheri:** EEG Paradigm Development, EEG Preprocessing, Investigation, Data Curation, Writing – Review & Editing. **Sinead Rocha:** Investigation, Data Curation, Writing – Review & Editing. **Perrine Brusini:** EEG Paradigm Development, Investigation. **Samuel Gibbon:** Investigation, Data Curation. **Panagiotis Boutris:** Investigation, Data Curation. **Natasha Mead:** Investigation, Data Curation. **Helen Olawale-Scott:** Investigation, Data Curation. **Henna Ahmed**: Investigation, Data Curation. **Sheila Flanagan:** Investigation. **Kanad Mandke:** Writing – Review & Editing. **Usha Goswami:** Conceptualization, Methodology, Funding Acquisition, Supervision, Project Administration, Writing – Original Draft.

Declaration of Competing Interest

The authors declare no competing financial interests.

Data availability

Data will be made available on request.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 694786). We would like to thank all the participating infants and their families.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jneumeth.2023.110036.

References

Abboub, N., Nazzi, T., Gervain, J., 2016. Prosodic grouping at birth. Brain Lang. 162, 46–59.
Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2009. Abnormal cortical processing of the syllable rate of speech in poor readers. J. Neurosci. 29 (24), 7686–7693.
Akbari, H., Khalighinejad, B., Herrero, J.L., Mehta, A.D., Mesgarani, N., 2019. Towards reconstructing intelligible speech from the human auditory cortex. Sci. Rep. 9 (1), 874.
Anumanchipalli, G.K., Chartier, J., Chang, E.F., 2019. Speech synthesis from neural decoding of spoken sentences. Nature 568 (7753), 493–498.
Attaheri, A., Ní Choisdealbha, Á., Di Liberto, G.M., Rocha, S., Brusini, P., Mead, N., et al., 2022a. Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. Neuroimage 247, 118698.
Attaheri, A., Panayiotou, D., Phillips, A., Ní Choisdealbha, Á., Di Liberto, G.M., Rocha, S., Brusini, P., Mead, N., Flanagan, S., Olawole-Scott, H., Goswami, U., 2022b. Cortical tracking of sung speech in adults vs infants: a developmental analysis. Front. Neurosci. 16.
Bergelson, E., Swingley, D., 2012. At 6–9 months, human infants know the meanings of many common nouns. Proc. Natl. Acad. Sci. USA 109 (9), 3253–3258.
Broderick, M.P., Anderson, A.J., Di Liberto, G.M., Crosse, M.J., Lalor, E.C., 2018. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Curr. Biol. 28 (5), 803–809.
Cirelli, L.K., Spinelli, C., Nozaradan, S., Trainor, L.J., 2016. Measuring neural entrainment to beat and meter in infants: effects of music background. Front. Neurosci. 10, 229.
Cogan, G.B., Poeppel, D., 2011. A mutual information analysis of neural coding of speech by low-frequency MEG phase information. J. Neurophysiol. 106 (2), 554–563.
Cover, T.M., 1999. Elements of Information theory. John Wiley & Sons.
Crosse, M.J., Butler, J.S., Lalor, E.C., 2015. Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. J. Neurosci. 35 (42), 14195–14204.
Crosse, M.J., Di Liberto, G.M., Bednar, A., Lalor, E.C., 2016. The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. Front. Hum. Neurosci. 10, 604.
Dauer, R.M., 1983. Stress-timing and syllable-timing reanalyzed. J. Phon. 11 (1), 51–62.
Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21.
Di Liberto, G.M., Attaheri, A., Cantisani, G., Reilly, R.B., Ní Choisdealbha, Á., Rocha, S., Brusini, P., Goswami, U., 2023. Emergence of the cortical encoding of phonetic features in the first year of life. Nat. Commun. 14, 7789.
Di Liberto, G.M., Crosse, M.J., Lalor, E.C., 2018a. Cortical measures of phoneme-level speech encoding correlate with the perceived clarity of natural speech. eNeuro 5 (2).
Di Liberto, G.M., Peter, V., Kalashnikova, M., Goswami, U., Burnham, D., Lalor, E.C., 2018b. Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. Neuroimage 175, 70–79.
Di Liberto, G.M., O'Sullivan, J.A., Lalor, E.C., 2015. Low-frequency cortical entrainment to speech reflects phoneme-level processing. Curr. Biol. 25 (19), 2457–2465.
Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19 (1), 158–164.
Ding, N., Simon, J.Z., 2014. Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum. Neurosci. 8, 311.
Doelling K.B., Arnal L.H., Assaneo M.F., 2022. Adaptive oscillators provide a hard-coded Bayesian mechanism for rhythmic inference. bioRxiv.
Doelling, K.B., Arnal, L.H., Ghitza, O., Poeppel, D., 2014. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85, 761–768.
Fló, A., Brusini, P., Macagno, F., Nespor, M., Mehler, J., Ferry, A.L., 2019. Newborns are sensitive to multiple cues for word segmentation in continuous speech. Dev. Sci. 22 (4), e12802.
Ghitza, O., 2012. On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. Front Psychol. 3, 238.
Ghitza, O., Greenberg, S., 2009. On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. Phonetica 66 (1-2), 113–126.
Gibbon, S., Attaheri, A., Ní Choisdealbha, Á., Rocha, S., Brusini, P., Mead, N., et al., 2021. Machine learning accurately classifies neural responses to rhythmic speech vs. non-speech from 8-week-old infant EEG. Brain Lang. 220, 104968.
Giraud, A.L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15 (4), 511–517.
Goswami, U., 2022. Language acquisition and speech rhythm patterns: an auditory neuroscience perspective. R. Soc. Open Sci. 9 (7), 211855.
Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., et al., 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol. 11 (12), e1001752.
Haufe, S., Meinecke, F., Gorgen, K., Dahne, S., Haynes, J.D., Blankertz, B., et al., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87, 96–110.
Jessen, S., Fiedler, L., Munte, T.F., Obleser, J., 2019. Quantifying the individual auditory and visual brain response in 7-month-old infants watching a brief cartoon movie. Neuroimage 202, 116060.
Jessen, S., Obleser, J., Tune, S., 2021. Neural tracking in infants–An analytical tool for multisensory social processing in development. Dev. Cogn. Neurosci. 52, 101034.
Kalashnikova, M., Peter, V., Di Liberto, G.M., Lalor, E.C., Burnham, D., 2018. Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. Sci. Rep. 8 (1), 1–8.
Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L., 2014. Large-scale video classification with convolutional neural networks. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 1725–1732.
Keshavarzi, M., Goehring, T., Zakis, J., Turner, R.E., Moore, B.C.J., 2018. Use of a deep recurrent neural network to reduce wind noise: effects on judged speech intelligibility and sound quality. Trends Hear 22, 2331216518770964.
Keshavarzi, M., Kegler, M., Kadir, S., Reichenbach, T., 2020. Transcranial alternating current stimulation in the theta band but not in the delta band modulates the comprehension of naturalistic speech in noise. Neuroimage 210, 116557.
Keshavarzi, M., Mandke, K., Macfarlane, A., Parvez, L., Gabrielczyk, F., Wilson, A., et al., 2022. Decoding of speech information using EEG in children with dyslexia: less accurate low-frequency representations of speech, Not" Noisy" representations. Brain Lang. 235, 105198.

Keshavarzi, M., Mandke, K., Macfarlane, A., Parvez, L., Gabrielczyk, F., Wilson, A., et al., 2022. Atypical delta-band phase consistency and atypical preferred phase in children with dyslexia during neural entrainment to rhythmic audio-visual speech. NeuroImage Clin. 35, 103054.
Keshavarzi, M., Reichenbach, T., 2020. Transcranial alternating current stimulation with the theta-band portion of the temporally-aligned speech envelope improves speech-in-noise comprehension. Front. Hum. Neurosci. 14, 187.
Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Softw. 82, 1–26.
Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J., 2018. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. J. Neural Eng. 15 (5), 056013.
Leong, V., Goswami, U., 2015. Acoustic-emergent phonology in the amplitude envelope of child-directed speech. PLoS One 10 (12), e0144411.
Leong, V., Kalashnikova, M., Burnham, D., Goswami, U., 2017. The temporal modulation structure of infant-directed speech. Open Mind 1 (2), 78–90.
Leong, V., Stone, M.A., Turner, R.E., Goswami, U., 2014. A role for amplitude modulation phase relationships in speech rhythm perception. J. Acoust. Soc. Am. 136 (1), 366–381.
Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54 (6), 1001–1010.
Mandel, D.R., Jusczyk, P.W., Pisoni, D.B., 1995. Infants' recognition of the sound patterns of their own names. Psychol. Sci. 6 (5), 314–317.
Martin, S., Brunner, P., Holdgraf, C., Heinze, H.J., Crone, N.E., Rieger, J., et al., 2014. Decoding spectrotemporal features of overt and covert speech from the human cortex. Front. Neuroeng. 7, 14.
Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., Amiel-Tison, C., 1988. A precursor of language acquisition in young infants. Cognition 29 (2), 143–178.
Mesgarani, N., Cheung, C., Johnson, K., Chang, E.F., 2014. Phonetic feature encoding in human superior temporal gyrus. Science 343 (6174), 1006–1010.
Mesgarani, N., David, S.V., Fritz, J.B., Shamma, S.A., 2009. Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. J. Neurophysiol. 102 (6), 3329–3339.
Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., Carreiras, M., 2016. Out-of-synchrony speech entrainment in developmental dyslexia. Hum. Brain Mapp. 37 (8), 2767–2783.
Ortiz Barajas, M.C., Guevara, R., Gervain, J., 2021. The origins and development of speech envelope tracking during the first months of life. Dev. Cogn. Neurosci. 48, 100915.
O'Sullivan, J.A., Power, A.J., Mesgarani, N., Rajaram, S., Foxe, J.J., Shinn-Cunningham, B.G., et al., 2015. Attentional selection in a cocktail party environment can be decoded from single-trial EEG. Cereb. Cortex 25 (7), 1697–1706.
Pasley, B.N., David, S.V., Mesgarani, N., Flinker, A., Shamma, S.A., Crone, N.E., et al., 2012. Reconstructing speech from human auditory cortex. PLoS Biol. 10 (1), e1001251.
Pei, X., Barbour, D.L., Leuthardt, E.C., Schalk, G., 2011. Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. J. Neural Eng. 8 (4), 046028.
Power, A.J., Colling, L.J., Mead, N., Barnes, L., Goswami, U., 2016. Neural encoding of the speech envelope by children with developmental dyslexia. Brain Lang. 160, 1–10.
Power, A.J., Mead, N., Barnes, L., Goswami, U., 2012. Neural entrainment to rhythmically presented auditory, visual, and audio-visual speech in children. Front. Psychol. 3, 216.
Power, A.J., Mead, N., Barnes, L., Goswami, U., 2013. Neural entrainment to rhythmic speech in children with developmental dyslexia. Front. Hum. Neurosci. 7, 777.
Qiu, P., Gentles, A.J., Plevritis, S.K., 2009. Fast calculation of pairwise mutual information for gene regulatory network reconstruction. Comput. Methods Prog. Biomed. 94 (2), 177–180.
Riedmiller, M., Braun, H., 1993. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. IEEE Int. Conf. Neural Netw. 586–591.
Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Ball, T., 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. Hum. Brain Mapp. 38 (11), 5391–5420.
Thomas, R.D., Moses, N.C., Semple, E.A., Strang, A.J., 2014. An efficient algorithm for the computation of average mutual information: validation and implementation in Matlab. J. Math. Psychol. 61, 45–59.
Weber, C., Hahne, A., Friedrich, M., Friederici, A.D., 2004. Discrimination of word stress in early infant perception: electrophysiological evidence. Brain Res. Cogn. Brain Res. 18 (2), 149–161.
Weissbart, H., Kandylaki, K.D., Reichenbach, T., 2020. Cortical tracking of surprisal during continuous speech comprehension. J. Cogn. Neurosci. 32 (1), 155–166.