Cognition 256 (2025) 106044

Contents lists available at ScienceDirect

Cognition

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

Full length article

Simulating prenatal language exposure in computational models: An exploration study

María Andrea Cruz Blandón [a], Nayeli Gonzalez-Gomez [b], Marvin Lavechin [c], Okko Räsänen [a,*]

[a] Unit of Computing Sciences, Faculty of Information Technology and Communication Sciences, Tampere University, Finland
[b] Centre for Psychological Research, Oxford Brookes University, United Kingdom
[c] Université Grenoble Alpes, Grenoble INP, GIPSA-lab, France

ARTICLE INFO

Dataset link: https://github.com/SPEECHCOG/ple_exploration_study

Keywords:
Computational modeling
Child language development
Prenatal language exposure
Language acquisition

ABSTRACT

Researchers have hypothesized that infant language learning starts from the third trimester of pregnancy. This is supported by studies with fetuses and newborns showing discrimination/preference for their native language. Jointly with empirical research, initial computational modeling studies have investigated whether learning language patterns from speech input benefits from auditory prenatal language exposure (PLE), showing some advantages for prior adaptation to speech-like patterns. However, these modeling studies have not modeled prenatal speech input in an ecologically representative manner regarding quality or quantity. This study describes an ecologically representative framework for modeling PLE for full-term and preterm infants. The approach is based on empirical estimates of the amount of prenatal speech input together with a model of speech signal attenuation from the external air to the fetus' auditory system. Using this framework, we conduct language learning simulations with computational models that learn from acoustic speech input in an unsupervised manner. We compare the effects of PLE to standard learning from only postnatal input on various early language phenomena. The results show how incorporating PLE can affect models' learning outcomes, including differences between full-term and preterm conditions. Moreover, PLE duration might influence model behavior, depending on the linguistic capability being tested. While the inclusion of PLE did not improve the compatibility of the tested models with empirical infant data, our study highlights the relevance of PLE as a factor in modeling studies. Moreover, it provides a basic framework for modeling the prenatal period in future computational studies.

1. Introduction

Human fetuses undergo, on average, from 38 to 42 weeks of development in the womb (Eggermont & Moore, 2012), constituting the prenatal period. Evidence from a large body of studies indicates that fetuses gain access to speech during the last trimester of pregnancy (Birnholz & Benacerraf, 1983; Hepper & Shahidullah, 1994; Holst et al., 2005). For instance, research in other mammals has demonstrated that external acoustic sounds can reach the womb. However, this signal undergoes a complex modification as it passes through the mother's body tissues, which act as a low-pass filter on the signal's frequency content (Querleu, Renard, Versyp, Paris-Delrue, & Crèpin, 1988). In parallel, the auditory system of the fetus undergoes gradual development (Graven & Browne, 2008; Pujol, Lavigne-rebillard, & Uziel, 1991). For example, the cochlea becomes anatomically functional by the 20th week of gestational age (Graven & Browne, 2008), but observable auditory evoked responses have been recorded from about 21 weeks gestation (Hepper & Shahidullah, 1994). These results collectively demonstrate that the auditory system is operational well before birth, thereby suggesting that language acquisition starts in utero.

Indeed, the prosodic bootstrapping hypothesis (Gervain, 2018) proposes that fetuses not only have access to speech but that their brain also starts adapting to it. Gervain (2018) argues that fetuses can start learning fundamental aspects of speech prosody in the womb, which then act as anchors for later language learning after birth. Support for this hypothesis arises from behavioral studies that show evidence of learning in the womb by assessing linguistic capabilities in fetuses or newborns, such as vowel discrimination (e.g., Cheour-Luhtanen et al., 1995; Moon, Lagercrantz, & Kuhl, 2013; Partanen et al., 2013)

* Corresponding author.
E-mail addresses: maria.cruzblandon@tuni.fi (M.A. Cruz Blandón), ngonzalez-gomez@brookes.ac.uk (N. Gonzalez-Gomez), marvinlavechin@gmail.com (M. Lavechin), okko.rasanen@tuni.fi (O. Räsänen).

https://doi.org/10.1016/j.cognition.2024.106044
Received 4 June 2024; Received in revised form 20 November 2024; Accepted 9 December 2024
Available online 18 December 2024

or preference for the mother's voice (e.g., DeCasper & Fifer, 1980; Hepper, Scott, & Shahidullah, 1993; Lee & Kisilevsky, 2014). More recently, Gonzalez-Gomez, O'Brien, and Harris (2021) contributed new evidence with a longitudinal study of full-term and preterm infants' language development across three linguistic aspects: prosody, phonetic perception, and phonotactics. They observed a developmental delay for preterm infants in prosodic perception but not in the phonetic and phonotactic tasks, possibly due to the duration of prenatal speech exposure (Gonzalez-Gomez et al., 2021). Under this hypothesis, full-term infants have more experience with prosodic information (already available in the womb), but both full-term and preterm infants have the same experience with phonotactic and consonant phonetic information (only available after birth). Together, this suggests that prenatal language exposure (PLE) in general, and the amount of exposure in particular, could impact later language development.

One approach to studying the effects of PLE in language development is using computational modeling. Computational modeling plays a key role in linguistics research. It allows us to propose mechanisms behind language behaviors and demonstrate whether these behaviors can be learned (e.g., Gelderloos, Kamelabad, & Alishahi, 2020; Huebner, Sulem, Fisher, & Roth, 2021; Nikolaus, Alishahi, & Chrupała, 2022). Recently, various studies have successfully used self-supervised artificial neural network models to simulate infant statistical learning,[1] demonstrating autonomous bootstrapping of phonemic and lexical discrimination (Lavechin, de Seyssel, Titeux et al., 2022), syllable and word segmentation (Khorrami & Räsänen, 2021), and learning of referential word meanings (Khorrami & Räsänen, 2021; Merkx, Scholten, Frank, Ernestus, & Scharenborg, 2023) from auditory or audiovisual language exposure without a need for strong linguistic priors or other innate biases. In terms of modeling studies, PLE has been previously examined in the context of speech emotion recognition (Vogelsang, Vogelsang, Diamond, & Sinha, 2023) and phonetic learning (Poli, Schatz, Dupoux, & Lavechin, 2024).

In Vogelsang et al. (2023), the simulation involved a neural network model of supervised learning from 1.6 h of emotion-labeled speech signals using either the original or filtered signal or a combination of both. They found that mimicking the order of speech exposure of infants (i.e., starting the training with the filtered signal and continuing it with the original signal) led to comparable emotion recognition performance to training with an equal amount of hours with the full signal. Their experiments showed that emotion recognition can achieve high performance from low-pass filtered speech, but, most importantly, the model that learned from low-frequency speech generalizes well to full-band speech, suggesting that emotion-related information – mostly carried out through prosody – is available in speech audible in the womb. The improvements in robustness of internal representations observed by Vogelsang et al. (2023) also suggest that access to a simplified stimulus at first might be more advantageous for auditory development than facing the full complexity of the stimulus already from the start. This supports the hypothesis that "early limitations" in input may be advantageous for later development (Turkewitz & Kenny, 1982; Vogelsang, Vogelsang, Pipa, Diamond, & Sinha, 2024). For example, infants' limited visual acuity is thought to prevent information overload by allowing them to focus on nearby objects, which helps synchronize visual and tactile experiences (Turkewitz & Kenny, 1982). Similarly, Newport and Elman have proposed that early limits on memory and other cognitive capabilities help infants process language in smaller, manageable units, aiding in their understanding of language structure (Elman, 1993; Newport, 1988, 1990).

More recently, a study by Poli et al. (2024) investigated integrating innate factors for modeling infants' attunement to speech sounds using a self-supervised statistical learning model. In this work, the authors simulate an evolutionary period during which they train the model on 500 h of various ambient sounds and animal vocalizations, followed by full-band speech data to simulate postnatal learning. Their results showed that the initial predisposition to perceive natural auditory patterns, followed by statistical learning of speech, does not lead to major differences in native vs. non-native phonetic category discrimination after 500 h of learning. However, pre-training[2] enables the separation of phonetic contrasts (in their study [ɹ]-[l], as in 'right' vs. 'light' and [w]-[j], as in 'well' vs. 'yell') already at birth (similarly to infants), whereas this is not the case without the predisposition.

Overall, the results from Poli et al. (2024) and Vogelsang et al. (2023) demonstrate that the inclusion of prenatal exposure in modeling studies can result in distinct learning outcomes and potentially improve compatibility with human data. This suggests that models of infant language learning should somehow incorporate a correct initial state for the learning—a factor that the previous modeling research has not considered except for these two studies. However, even the studies by Poli et al. (2024) and Vogelsang et al. (2023) did not model prenatal language exposure in detail regarding the quantity of speech available to the fetus or the signal attenuation from external air to the auditory system of the fetus. By incorporating a more realistic simulation approach to prenatal language exposure, we could further explore the effects and implications of the prosodic bootstrapping hypothesis using computational means.

The present study aims to investigate whether integrating realistic aspects of prenatal language exposure into language learning simulations results in different learning outcomes compared to those without it, thus aligning with previous studies. More specifically, our study focuses on examining the effects of prenatal exposure through the lens of the prosodic bootstrapping hypothesis, also considering full-term and preterm infants in terms of prenatal exposure duration. Finally, we investigate if prenatal language exposure simulation leads to models more compatible with infant data. Our approach consists first of simulating preterm and full-term learners using statistical learning while accounting for a realistic amount of speech exposure, a developing hearing system in the womb, and a plausible amount of exposure. Second, we assess the models' language learning outcomes through a computational replication of the prosodic and phonotactic tasks of Gonzalez-Gomez et al. (2021) and infant-directed speech (IDS) preference and vowel discrimination tasks from a recent evaluation framework for early language acquisition models (Cruz Blandón, Cristia, & Räsänen, 2023).

2. Methodology

Our methodological aim is to enable plausible simulations of learning from pre- and postnatal language exposure. Our methodology thereby consists of two phases: (1) defining the amount and type of speech exposure before and after birth, and (2) a description of the computational simulation setup to test the prenatal bootstrapping hypothesis with the proposed PLE framework. These are explained in the following sections. Fig. 1 shows an overview of the employed experimental setup.

2.1. Simulation of the Prenatal Language Exposure (PLE)

We focused on three factors in our simulations of PLE: (i) as accurate modeling of hearing in the womb as the existing data on the phenomenon permits, (ii) a realistic amount of language exposure before

[1] See, e.g., de Seyssel, Lavechin, and Dupoux (2023) and Dupoux (2018) for motivation on using neural networks as models for infant statistical learning.

[2] *Pre-training* is a commonly utilized procedure in supervised machine learning, where a model's parameters are first estimated using a potentially large unlabeled dataset and an auxiliary "self-supervised" optimization function before training the final model using the target data (see Erhan, Courville, Bengio, & Vincent, 2010).

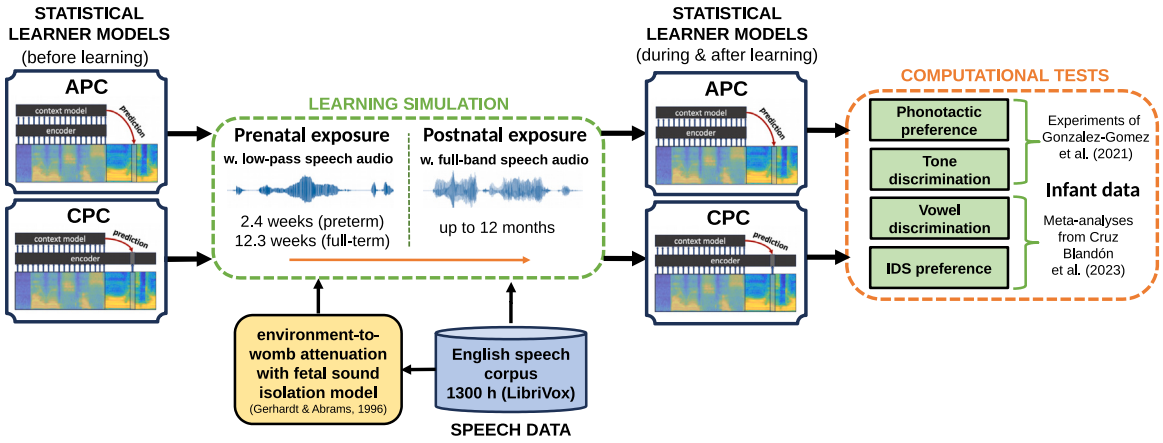

Fig. 1. An overview of the experimental pipeline to simulate prenatal language exposure in case of pre- and full-term infants.

and after birth, and (iii) using purely incremental statistical learner models, i.e., models that can "hear" all speech inputs only once, thereby contrasting with the standard approach to training neural networks in machine learning while being central to realistic simulations of learning from finite input.

Simulation of auditory experience in the womb. Multiple studies have described how the womb acts as a low-pass filter on the speech signal, where frequencies below 400 Hz are well-preserved while increasing levels of attenuation take place at higher frequencies (Lecanuet et al., 1998; Querleu et al., 1988; Richards, Frentzen, Gerhardt, McCann, & Abrams, 1992; Vince, Armitage, Baldwin, Toner, & Moore, 1982; Vince, Billing, Baldwin, Toner, & Weller, 1985). The most recent study on the subject described an attenuation of less than 5 dB for low frequencies (<500 Hz) and 20–30 dB for higher frequencies (Gélat et al., 2019). In turn, studies focused on measuring fetuses' responses to different auditory stimuli have revealed a gradual development of hearing in the womb (e.g., Birnholz & Benacerraf, 1983; Hepper & Shahidullah, 1994). In their longitudinal study, Hepper and Shahidullah (1994) found that fetuses started to exhibit consistent motor responses to tone frequencies lower than 500 Hz at 27 weeks of gestational age and to 1000-Hz frequency by 31 weeks of gestational age. Hence, hearing in the womb is not only affected by the filtering effects of the body tissues but also by the gradual development of the auditory system. A study by Gerhardt and Abrams (1996) incorporated these two aspects by using a hydrophone to measure sound levels in the uterus of an ewe and an electrode to capture cochlear microphonic responses from the round window of the inner ear of the ewe's fetus. The cochlear microphonic measurements characterized what they called *the fetal sound isolation*, which is the level of external sounds that would stimulate the fetal cochlea when the response is compared against an external reference microphone. Their results on fetal sound isolation outline a relatively strong attenuation of approximately 10–35 dB already for low frequencies (<500 Hz) and 35–45 dB for higher frequencies.

For our purposes, we employed a fetal sound isolation model to simulate hearing in the womb according to the combined filtering effects of the womb and responsiveness of fetal hearing. More precisely, we designed a digital filter approximating the transfer function described in Gerhardt and Abrams (1996) using MATLAB Filter Design Toolkit. The resulting infinite impulse response filter had a transfer function in the z-domain of:

$$H(z) = \frac{0.2373x10^{-3} + 0.4746x10^{-3}z^{-1} + 0.2373x10^{-3}z^{-2}}{1 - 1.956z^{-1} + 0.9569z^{-2}} \quad (1)$$

There are multiple challenges involved in accurately measuring the inner ear response to speech stimuli in human fetuses. It is important to note that the used model derives from an animal study and corresponds to a snapshot of a certain developmental stage based on measurements from a single study. Yet, the resulting filter (Fig. 2), according to our knowledge, has been designed based on the best available information on fetal hearing.

Amount of language exposure. To estimate the amount of exposure to speech during the prenatal period, Monson, Ambrose, Gaede, and Rollo (2023) analyzed more than 23,000 h of long-form LENA recordings[3] from 27 developing fetuses (recorder placed near the abdomen of the pregnant participant) and 24 preterm infants (recorder inside the crib) at a neonatal intensive care unit (NICU). Typically, preterm infants spend time in the NICU before being discharged. According to Monson et al. (2023), the average stay duration in the NICU was five weeks, during which the infants were only exposed to 0.5 h of speech per day. In contrast, the average prenatal in-uterus speech exposure was found to be 2.6 h per day.

As for the amount of postnatal speech experience, we used data from Bunce et al. (2021) who investigated the amount of speech exposure in the everyday life of infants across five socio-cultural contexts. They used data from long-form audio recordings recorded with microphones placed in the infant's clothes. We calculated the cross-linguistic average hourly infant speech exposure from their data, resulting in 14.6 min of speech per hour. Assuming 12 daily waking hours for an infant, the estimate transforms into 2.9 h of speech per day, which we then extrapolated to the age range of interest in our experiments (up to 12 months).

To determine the duration of the PLE for our simulations, we combined the age of the operational hearing onset and the age at birth. Various studies have identified the operational hearing onset to be around 27 weeks of gestational age (Chelli & Chanoufi, 2008; Eggermont & Moore, 2012; Querleu et al., 1988; Saffran, Werker, & Werner, 2006). We multiplied the hourly input estimates from Bunce et al. (2021) and Monson et al. (2023) with the average number of days in the womb, in the NICU (for preterms), and after birth in order to create our experimental conditions, as explained in the following section (see also Fig. 1).

2.2. Experimental conditions

To tackle our aim of simulating the effects of PLE in computational models of infant learners, we created three distinct experimental conditions: baseline simulation (BS) using only the original full-band speech signal for learning, full-term learner simulation (FT) using a typical

[3] One common approach to long-form recordings is to use the so-called LENA device, which is a non-intrusive, non-invasive recording device that is placed inside a vest pocket of the child and that can perform 16 h of continuous audio recording.

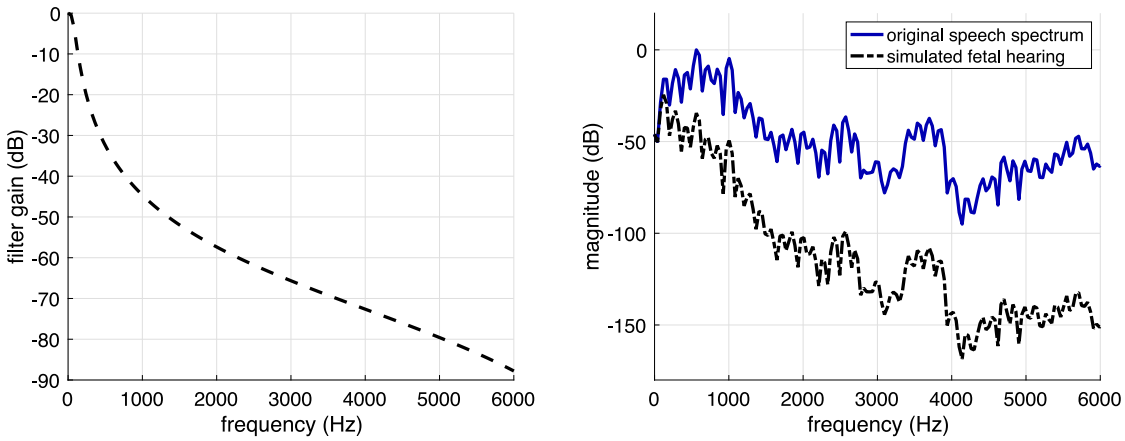

Fig. 2. Left: frequency response of the fetal sound isolation filter used to simulate attenuation of speech sounds in fetal hearing. Right: an example of a typical spectrum of voiced speech (blue solid line) and spectrum of the signal after the fetal sound isolation filter. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

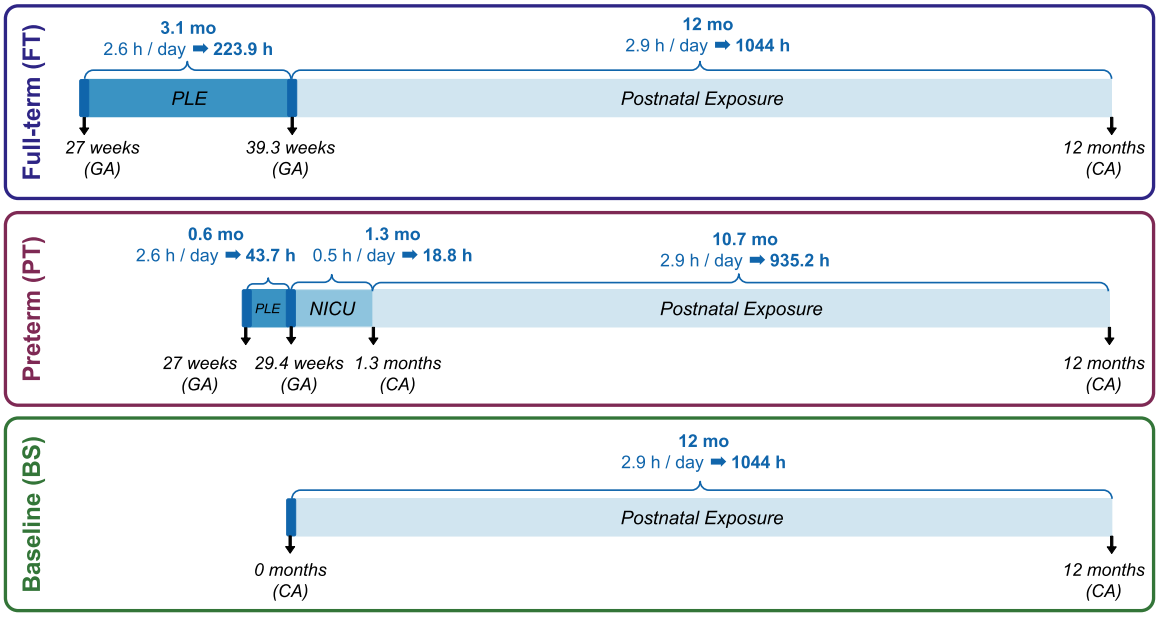

Fig. 3. Simulations included in this study with their corresponding specifications for simulated speech exposure and the duration of the prenatal period. (GA) stands for gestational age, (CA) for chronological age, (PLE) for prenatal language exposure, and (NICU) for neonatal intensive care unit. The prenatal period is enclosed by the operational hearing onset (first dark blue rectangle) and birth (second dark blue rectangle). For the preterm learner, the simulation also includes the time spent at the NICU (after PLE). For each period simulated (PLE, NICU, and postnatal period), the duration in months is shown first in bold font, followed by the daily speech exposure (normal font) and total speech hours of the period (bold font). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

duration of prenatal speech exposure, and a preterm learner simulation (PT) with shorter prenatal exposure followed by speech-impoverished period in the NICU. The latter two included simulation of PLE by employing the fetal sound isolation filter and the estimates of prenatal speech input depending on whether a full-term or preterm learner was being modeled. All simulations also include standard training with the normal full-band speech signal, simulating postnatal learning. Fig. 3 illustrates the three conditions.

Full-term learner simulation (FT). To simulate full-term learners, we set the gestational age at birth to 39.3 weeks, which was the average gestational week for full-term infants in Gonzalez-Gomez et al. (2021) (personal communication). Consequently, the FT condition included 3.1 months of PLE, followed by 12 months of postnatal speech exposure. This resulted in 223.9 h of PLE and 1044 h of full-band postnatal input.

Preterm learner simulation (PT). For the preterm condition, we used the average gestational age of 29.4 weeks reported by Gonzalez-Gomez et al. (2021), thereby resulting in mere 0.6 weeks of PLE. Therefore, for the preterm learners, we have 43.7 h of PLE (filtered signal), followed by 18.8 h of speech in the NICU, and finally, 935.2 h of full-band postnatal input to simulate input up to the 12 months of chronological age.

Baseline simulation (BS). In the baseline condition, only learning from a full-band signal was employed, thereby aligning with the previous work on the present type of computational models (Khorrami & Räsänen, 2021; Lavechin et al., 2024). Similarly to the postnatal exposure in the FT and PT conditions, this corresponded to 1044 h of postnatal full-band input to simulate 12 months of age.

Speech data. As speech data, we used a subset of the LibriVox corpus (Kearns, 2014) as preprocessed and formatted by Lavechin, de

Seyssel, Titeux et al. (2022)[4] that comprises 1300 h of read-aloud English books. The speech data were divided into disjoint sections for the PLE and postnatal language exposure in the different conditions. Note that although using day-long recordings of infants' natural learning environments could provide more realistic data for simulating infant learning (Lavechin, de Seyssel, Gautheron, Dupoux and Cristia, 2022), the amount of noise in these typically mono-channel recordings is actually more than what infants with binaural hearing are exposed to, and causes substantial problems for the present-type of computational models designed for clean speech audio (see discussion in Lavechin et al., 2024).

2.3. Computational learner models

For the simulations, we used two self-supervised models—autoregressive predictive coding (APC) (Chung, Hsu, Tang, & Glass, 2019) and contrasting predictive coding (CPC) (van den Oord, Li, & Vinyals, 2018). These models employ a form of learning that operates without labels, relying solely on input data; hence, their name "self-supervised". In essence, they learn by predicting the speech signal ahead in time, which forces them to leverage speech patterns that enable accurate temporal predictions. Previous studies have employed these algorithms as models of infant statistical learning from acoustic speech, demonstrating learning of phonemic discrimination skills (e.g., Liu, Tang, & Goldwater, 2023; Poli et al., 2024) and lexical word-form acquisition (Khorrami, Cruz Blandón, & Räsänen, 2023; Lavechin, de Seyssel, Titeux et al., 2022; Lavechin et al., 2023) in a manner compatible with the hypothesis that infants' language acquisition might be largely driven by the acquisition of statistical regularities from the sensory input (Aslin & Newport, 2014; Saffran, Aslin, & Newport, 1996; Saffran & Kirkham, 2018). Moreover, Cruz Blandón et al. (2023) recently applied these models to study developmental trajectories of several linguistic capabilities, comparing trajectories of models to those estimated from infants. As these models can learn directly from the speech signal in a self-supervised manner, they impose no prior assumptions on the type of input representations available to learners (see discussion in Dupoux, 2018).

In the present study, we simply view these models as two different implementations of predictive processing-based statistical learning that is hypothesized to be a central mechanism to language learning in infants (Aslin & Newport, 2014; Saffran et al., 1996; Saffran & Kirkham, 2018). We simultaneously acknowledge that they are unlikely to be complete mechanistic descriptions of real infant learners due to their simplistic nature (see also (de Seyssel et al., 2023), for a discussion on the CPC model). In essence, these models are a way to study to what extent statistical learning of acoustic patterns might explain the early stages of infant language learning, serving as a minimal basis on which additional mechanisms, biases and constraints could be incorporated as needed to explain developmental processes (see, e.g., Lavechin, de Seyssel, Gautheron et al., 2022). At the same time, they are one of the few existing models capable of tackling several aspects of language development concurrently without imposing linguistic priors to the model (e.g., Khorrami, 2024; Lavechin, de Seyssel, Titeux et al., 2022) and while learning from raw speech signals. Any observed similarities between the two models may reflect the learning outcomes achievable with statistical learning, while any differences likely arise from the differences in how predictive processes are implemented in them.

To simulate realistic language exposure and learning, we trained the APC and CPC models using an incremental learning curriculum, where each utterance was presented only once to the model during the learning process. Input to both models consisted of two-second speech clips, represented by 40 log-Mel features (auditorily-weighted spectral features) that were extracted using 25-ms windows with 10-ms steps (frames). Utterances longer than two seconds were split into multiple two-second clips and fed in the original order to the models.

[4] The corpus was preprocessed to remove silences.

Autoregressive Predicting Coding (APC). The APC is a neural network model that attempts to predict the future evolution of a speech signal, given access to current and past observations of the signal. The learning process aims to minimize the mean absolute error between predicted and actual future signal observations, as represented by frames of feature vectors that encode the spectral envelope of the speech signal (Chung et al., 2019). The model architecture used in our simulations aligns with that employed by Liu et al. (2023) and Yang, Yeh, Chung, Glass, and Tang (2022). The model architecture consists of a context modeling block that uses recurrent layers to capture the dependencies between encoder output across time, implemented with three Long Short-Term Memory (LSTM) layers with 512 units each. Following this, there is a fully connected linear layer that converts the output of the context module into a prediction of the future speech observation k time-steps ahead. In our implementation, the model predicts three frames in the future, equivalent to 30 ms. We used Adam optimizer (Kingma & Ba, 2014) with a learning rate of 10^{-4} and trained the model until total speech hours were exhausted.

Contrastive Predicting Coding (CPC). The CPC model is conceptually similar to APC. However, instead of predicting future speech observations (speech feature frames), the learning criterion of CPC focuses on predicting the model's own internal representations of future speech observations (van den Oord et al., 2018). By mapping the input speech features into latent representations using a learnable encoder module while simultaneously trying to predict these representations over time, CPC has to discover a set of internal representations that best serve the prediction task. We follow a similar implementation to that used in (Chung et al., 2019; Nguyen et al., 2020). The model architecture corresponds to a multilayer perceptron (MLP) encoder block of three fully connected layers with 512 units each. Then, two LSTM layers with 256 units form the context modeling block, which is used to produce predictions of the internal encoder representations $1–k$ time steps ahead. The model is optimized by using a contrastive loss function (see Gutmann & Hyvärinen, 2010; van den Oord et al., 2018) that evaluates the accuracy of distinguishing between actual future internal representations in contrast with the predictions generated by the context block and negative samples drawn from the same utterance. We used 128 negative samples and predicted a total of 12 steps ahead. We used Adam optimizer with a learning rate of 10^{-5} and trained the model until total speech hours were exhausted.

2.4. Assessment of learning outcomes

For the evaluation of models' learning outcomes, we employed psycholinguistically inspired tests by implementing computational versions of the tasks described in Gonzalez-Gomez et al. (2021), complemented by previously reported tasks from Cruz Blandón et al. (2023). In total, we tested the models with four different linguistic tasks (see summary in Table 1).

In their study, Gonzalez-Gomez et al. (2021) examined developmental changes at the ages of 7, 9.5, 10, and 12 months of chronological age for three different linguistic aspects: prosody, phonotactics, and phonetic. To investigate the effects of the duration of prenatal exposure in our study, we evaluated the models using *tone discrimination* and *phonotactics preference* tasks. The tone discrimination task related to prosodic development revealed developmental differences between full-term and preterm infants. Meanwhile, the phonotactics preference task showed no differences between the infant groups.

Moreover, we complement our analysis with *infant-directed speech (IDS) preference* and *vowel discrimination* tasks from a recent evaluation framework for computational models of language learning known as MetaEval, as derived from "Meta-analytic evaluation of computational models" (see Cruz Blandón et al., 2023). MetaEval focuses on comparing an age-dependent developmental trajectory of a language capability, as obtained from a meta-analysis pooling over several empirical studies on infants, to a corresponding developmental trajectory

Table 1
Linguistic tasks for assessing the models. The phonetic transcriptions used in the examples are in IPA, except for the tone discrimination task that uses Jyutping phonetic alphabet.

Task	Linguistic aspect	Stimulus example
Tone discrimination	Prosody	/baa2/ vs /baa3/
Phonotactics preference	Phonotactics	/dʌs/ vs /tʃʌʃ/
IDS preference	Prosody	IDS: ‘‘Do you wanna hold the handle? Look at that’’ ADS: ‘‘So it’s good for yeah for mixing’’
Vowel discrimination	Phonetic	/hɛd/ vs /hɪd/

of the model under examination (Cruz Blandón et al., 2023). This is achieved through a computational test of the same linguistic capability tested in the meta-analysis, which calculates the effect size (attentional preference or discrimination competence) obtained by the model trained on an increasing amount of data, thereby simulating different ages of an infant. A model is deemed compatible with the human reference data if it obtains an effect size equal to or greater than the lower bound of the effect size observed for infants in the meta-analysis at the given age checkpoints of interest. This is since the infant data is inherently noisy, and thereby, the measured effect sizes form a lower bound for the real effect magnitudes (see Cruz Blandón et al., 2023, for details). Therefore, effect sizes derived from models should be interpreted cautiously, reflecting not an exact magnitude that infants might exhibit but rather the relative strength of the tested capability over time. It is important to note that, for the tasks from MetaEval, based on the available meta-analytic data, information is solely available for full-term babies, for which compatible comparisons can only be made for full-term and baseline simulations.

For all the tests, tone discrimination and phonotactics preference included, we adopted the MetaEval framework, in which the models are expected to produce two types of outputs for a given speech stimulus: internal representations and attentional preference scores, both as a function of stimulus time. The internal representations are vectors of the network activations for a given stimulus. In the case of the current models, these correspond to the internal speech representations the models have learned. On the other hand, the attentional preference scores represent attentional saliency of a stimulus according to the models’ encodings of the stimulus.

In our experiments, we implement attentional preference as novelty preference often observed in empirical infant studies, where higher surprisal (lower familiarity) of the stimulus results in stronger attentional attraction. This approach assumes that infants are efficient information seekers and prioritize inputs with higher expected information value (lower predictability) over those that carry less information due to being predictable.[5] This approach aligns with the hypothesis presented in Räsänen et al. (2018) (see also MacDonald et al., 2020) that IDS may be attentionally attractive due to its prosody being less predictable than that of ADS, and allows us to test if IDS preference can be learned from speech experience instead of being innate. In practice, the novelty preference was implemented using the models’ loss functions as the attentional preference scores, as the losses directly reflect the models’ ability to predict the evolving signal in time based on what they have learned about speech so far. Hence, higher loss values indicate lower predictability of the input from the models’ perspective, thereby indicating higher surprisal (see Cruz Blandón et al., 2023 for more details).

Tone discrimination task. This task consisted of a central fixation procedure where infants, English learners, were presented with two Cantonese lexical tones (tone 25: high rising and tone 33: mid-level. See Gonzalez-Gomez et al., 2021; Yeung, Chen, & Werker, 2013) using two minimal pairs of the form /CV/. Gonzalez-Gomez et al. (2021) reported that preterm infants consistently displayed the ability to discriminate the two tones until 10.5 months of chronological age, while full-term infants lose this ability after 9 months of chronological age.

To make the task suitable for computational models, we synthesized 234 Cantonese minimal pair syllables with the two tones 25 and 33[6] using the text-to-speech (TTS) system of Google Cloud. We employed all four voices (in the `Standard` voice model) available for Cantonese: two female voices and two male voices. We opted for a test setup similar to an ABX format, given that this is a discrimination task (see Cruz Blandón et al., 2023; Schatz, 2016; Schatz et al., 2014). In the test, we calculate the distance between the internal representations of two given syllables. If the syllables are the same, i.e., same /CV/ context and tone, the distance is expected to be smaller than if they were different, i.e., same /CV/ context but different tone. Hence, the test determines the degree of tone discrimination by comparing within-tone perceptual distances to across-tone distances. The test’s outcome is the effect size as the quantified strength of discrimination capability.

Phonotactics preference task. In Gonzalez-Gomez et al. (2021), the phonotactics task consisted of exposing infants to two types of /CVC/ triphones in English: highly probable and less probable triphones, and then comparing infant responses between these conditions (see Gonzalez-Gomez et al., 2021; Jusczyk, Luce, & Charles-Luce, 1994). As a result, Gonzalez-Gomez et al. (2021) found that both preterm and full-term infants did not exhibit a preference for either group of triphones until the age of 10.5 months, when they started to prefer highly probable triphones.

For the computational version, we used the same 12 highly and 12 less probable triphones of English used in the behavioral study. We synthesized the triphones into speech using the Google Cloud TTS system, generating a total of 324 samples per condition (27 tokens per triphone, 13 female and 14 male voices available for English neural TTS models). Since this is a preference task, the test used the models’ attentional preference scores as their behavioral responses for the highly and less probable triphones. We calculated the average per sample preference within each triphone group and estimated the effect size between the two groups as a measure of preference, a positive effect standing for preference towards the high probability triphones (see also the infant-directed speech preference task in Cruz Blandón et al. (2023) for more details on this type of testing).

Infant-directed speech (IDS) preference. This test consists of 120 IDS and 120 adult-directed speech (ADS) utterances from the large-scale ManyBabies study on infant IDS preference (Frank et al., 2020). The stimuli are presented to the model to measure the attentional scores and, thus, the preference towards the IDS style speech. North-American English infant data from Bergmann et al. (2023) is used as the data for this task to compute reference developmental trajectories of infants (see also Cruz Blandón et al., 2023). In this task, full-term infants displayed an increased preference for IDS with age (Frank et al., 2020).

[5] Alternative ways to implement attentional preference mechanism would have been to consider familiarity preference. However, learning-based novelty preference seemed the most justified choice since we know that IDS preference increases with infant age (Frank et al., 2020) and MacDonald, Räsänen, Casillas, and Warlaumont (2020) and Räsänen, Kakouros, and Soderstrom (2018) show that IDS prosody is less predictable than that of ADS.

[6] To extract the minimal pairs, we used the syllable combinations reported in https://humanum.arts.cuhk.edu.hk/Lexis/lexi-mf/syllables.php.

Vowel discrimination. The vowel discrimination test consists of native (English) vowel contrasts embedded in a /hVd/ context from the Hillenbrand corpus[7] (Hillenbrand, Getty, Clark, & Wheeler, 1995). The test covers five different vowel contrasts from 139 speakers (see Cruz Blandón et al., 2023 for more details). The contrasts are presented to the model using a similar ABX format to determine the strength of the discrimination between same vowel and different vowel conditions. The test uses as reference data the vowel discrimination meta-analysis available on MetaLAB database[8] (Bergmann et al., 2018; Lewis et al., 2016). According to the reference data, full-term infants exhibited a positive effect for the native vowel discrimination capability, with this ability remaining consistent across ages.

Analysis of learning outcomes in different exposure conditions. First, we focused on whether the two simulations with PLE (FT and PT) display different behaviors compared to the baseline learning strategy without PLE (BS condition). To address this, we derived and compared FT, PT, and BS developmental trajectories of all four tested capabilities (native vowel discrimination, IDS preference, phonotactics preference, and tone discrimination) using model checkpoints at chronological ages of $\{0, 7.5, 9, 10.5, 12\}$ months. These checkpoints include those studied by Gonzalez-Gomez et al. (2021) and one new checkpoint, 0 months, corresponding to the time of birth for FT and PT simulations. Results at 0 months is not included for the BS simulation, as it corresponds to a randomly initialized model that results in outlier measurements. Instead, the BS simulations included the checkpoint of 0.5 months, corresponding to the same amount of speech (43.7 h) as the total amount of prenatal exposure at birth for the PT simulation. We also included a 2.6-month checkpoint for BS, which corresponds to the same amount of speech (223.9 h) as prenatal speech input at birth for the FT simulation.

Training and testing of both models were conducted three times, each run using a different set of initial parameters and different order of the training data. This was done to average out variability from these task- and condition-independent random factors. The mean effect size and its standard error were then calculated for each test and age checkpoint across the three runs. This allows us to assess whether different checkpoints and simulations exhibit consistently different behavior for the various linguistic capabilities tested, and how the duration of the PLE changes the models' behavior in the simulations.

Apart from analyzing potential differences in the model behavior, we were also interested in whether the current experiments align with the qualitative results from the previous studies. Poli et al. (2024) and Vogelsang et al. (2023) suggested that models with prenatal auditory exposure were potentially more compatible with human data than those without PLE. For this purpose, we followed the methodology in MetaEval for the vowel discrimination and IDS preference tests to obtain the total number of age checkpoints that were compatible with the effect sizes derived from the meta-analytic data. We calculated the mean effect size across the three runs and evaluated if it was larger or equal to the lower bound of the confidence interval from the meta-analytic reference data at each age checkpoint. Then each model's compatibility was derived as the total of compatible age checkpoints. Since Gonzalez-Gomez et al. (2021) is not a meta-analysis, we approached the compatibility analysis differently. Thus, for the phonotactics preference and tone discrimination tests, we used the range given by the standard error of the mean effect size (SE) across the three runs. If the range given by mean effect±SE included zero, then the checkpoint was marked as having no effect (no preference or discrimination). Otherwise, it was marked as having an effect. To be consistent with what was reported in Gonzalez-Gomez et al. (2021), we coded the effects of the phonotactics preference as familiar effects. Since the models are simulating novelty preference (larger loss for novel stimuli than familiar stimuli; see Section 2.3), we multiplied by −1 the effects obtained, thus matching the familiarity effect (i.e., lower loss error for highly probable than lower probable triphones). The effects were then compared directly to the results in Gonzalez-Gomez et al. (2021) to determine the number of infant-compatible age checkpoints.

[7] http://homepages.wmich.edu/~hillenbr/, accessed on 16.10.2023.
[8] https://langcog.github.io/metalab/.

3. Results

Figs. 4–7 show the results from the experiments. There are three main findings. First, the results show consistent differences between the baseline and prenatal language exposure simulations, thus confirming that such differences are also present in realistic setups (see Fig. 4). Second, the duration of the prenatal exposure simulation can lead to differences between the full-term learner and preterm learner simulations' behaviors (see Fig. 5). Third, despite previous studies suggesting a potential increase of the qualitative compatibility with infant data by simulating PLE, our results show no quantitative advantage against the baseline simulation in the current setup (see Figs. 6 and 7) (see discussion in Section 4). We analyze these three main findings in the following subsections.

3.1. Finding 1: Inclusion of PLE results in different model behaviors

Fig. 4 shows the developmental trajectories of APC and CPC models for the four linguistic capabilities tested: phonotactics preference, tone discrimination, IDS preference, and vowel discrimination. The mean effect sizes and their standard errors across the three runs are shown for each age checkpoint of interest for the three simulations: full-term learner (purple), preterm learner (dark pink), and baseline (green).

The results show that, for both models, there are differences in the developmental trajectories of the baseline simulations and the PLE simulations. Even in the case of overlapping trajectories, the trajectory direction can differ between BS and PLE conditions. For example, in the case of the APC model's IDS preference, the BS shows an increasing preference from 10.5 months to 12 months, while FT and PT simulations show a decreasing tendency (see Fig. 4 bottom-right). These results indicate that regardless of the model employed to simulate a statistical learner (either the APC or CPC model), the simulation of prenatal language exposure results in different model behaviors compared to ignoring the PLE completely. Given that these patterns are a result of multiple independent runs of the experiment, these differences are not merely due to random factors in the models or data. This corroborates the findings from Poli et al. (2024) and Vogelsang et al. (2023), where the studied models had different learning outcomes when their training involved prenatal auditory exposure.

When compared directly, the APC and CPC models show that they can capture relevant information from the input speech by reflecting non-random linguistic behaviors in all four tests to a certain degree. In general, both models have non-zero effects for all the simulations (BS, FT, and PT). However, the observed patterns vary between the models. In the case of the APC model, the differences between PLE and BS conditions are clearer at early ages. Also, the pattern is consistent across the type of test, where IDS and triphone preference tests showed an increasing preference for the BS condition and a decreasing preference for the PLE conditions from 0 to 7.5–9 months of age (see Fig. 4 left-top and left-bottom). Furthermore, for the APC discrimination tests, the BS simulation shows a decreasing discrimination capability while the PLE simulation shows an increasing capability from 0 to 7.5 months (see Fig. 4 right-top and right-bottom). This is different from the CPC model, which does not exhibit the same patterns for the same tests. For instance, the CPC phonotactics preference test shows an increasing preference with age, while for the IDS preference, the BS simulation shows a decreasing preference while the PLE simulations show an increasing preference (see Fig. 4 left-top and left-bottom). This behavior is qualitatively similar to infants who show an increasing IDS

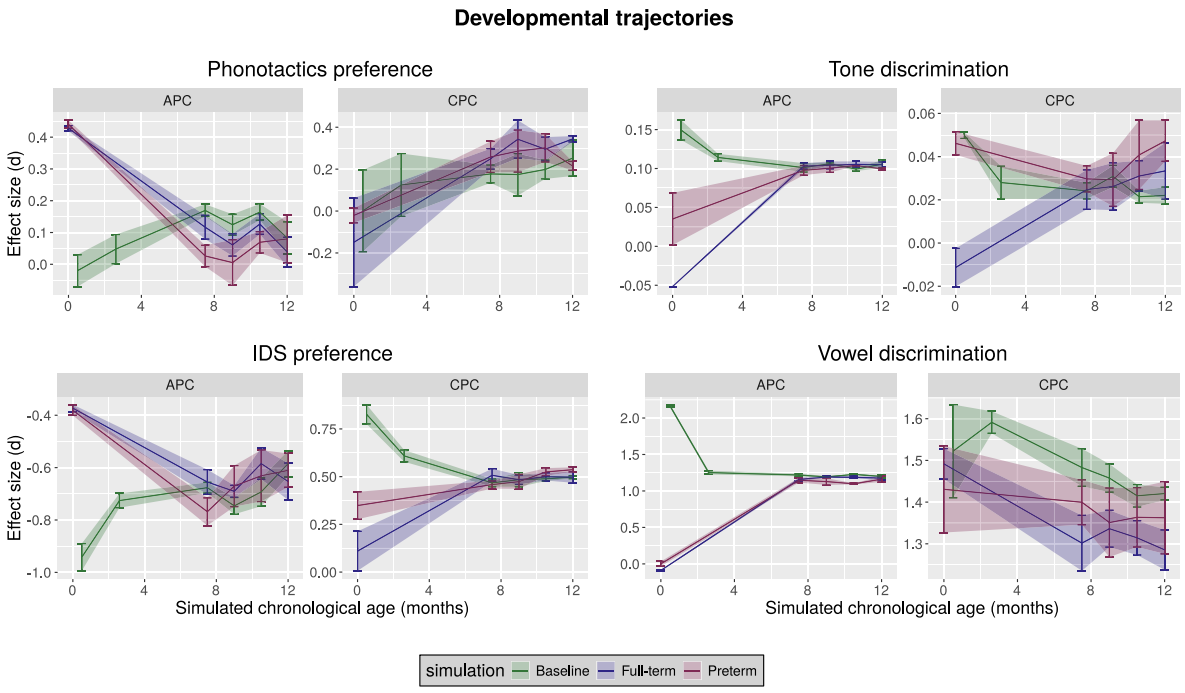

Fig. 4. Developmental trajectories for the three conditions: FT (Full-term learner simulation; in purple), PT (Preterm learner simulation; in dark pink), and BS (Baseline simulation; in green). Results are shown for the four linguistic capabilities (Phonotactics preference top-left, Tone discrimination top-right, IDS preference bottom-left, and Vowel discrimination bottom-right) and for the two models (APC and CPC). Each plot shows the effect sizes obtained for the given task and model for the simulated ages at zero months (at birth), 7.5 months, 9 months, 10.5 months, and 12 months. For BS, 0.5 and 2.6 months are shown, representing the same total amounts of speech perceived during PLE by PT and FT, respectively. Error bars and shaded areas represent ±1 standard error of the mean effect size, as calculated across the three independent runs of the experiment. For the IDS preference, a positive effect indicates a preference towards IDS utterances. For the phonotactics preference, a positive effect indicates a preference towards highly probable triphones. Note the differences in the scale of the effect size for all capabilities (axis y). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

preference (Frank et al., 2020) and start exhibiting preference towards high-probability triphones after 10.5 months of chronological age, that is, after some experience with the full-band speech signal (Gonzalez-Gomez et al., 2021). Note that for the IDS preference, given the current implementation, these results suggest that a novelty preference could replicate the qualitative trajectories exhibited by infants in the CPC model.

The CPC model seems more robust than APC to the qualitative change in the audio that takes place during the transition from filtered speech to full-band signals. This is illustrated by the effect sizes at the 0 age checkpoint of the PLE simulations compared to the corresponding checkpoints of BS simulations, where a matching amount of speech hours have been used for training the PLE for PT (0.6 months) and FT (2.6 months) models. For three out of the four linguistic tests, the effect sizes obtained by the BS condition are close to those obtained by the PT and FT conditions. Notice that for the FT and PT simulations, by 0 months, the model has only seen filtered signals during training, but the tests are all performed with the full-band signals.

3.2. Finding 2: Duration of PLE can impact model's behavior

One prediction arising from the prosodic bootstrapping hypothesis is that the duration of PLE can affect later language learning outcomes. In order to analyze if the PLE duration impacts the learning outcomes of the present models, we plotted the developmental trajectories in three views (see Fig. 5): (1) Simulated chronological age, same as previous visualization. (2) simulated corrected age, which corresponds to aligning the PT simulation to the FT simulation, and (3) total amount of speech hours observed. We chose the phonotactics preference and tone discrimination task trajectories of the CPC model, as they are representative of the behaviors found in the experiments. The plots for the other tests and the APC model can be found in Appendix A.

For the phonotactics preference, the CPC model shows that behaviors of the FT and PT simulations are quite similar, except for a decrease in the preference from 10.5 to 12 months of chronological simulated age for the PT simulation, as compared to a continued increase by the FT simulation. On the other hand, for tone discrimination, the simulations are very different at birth, but they start to be more similar from 7.5 months with a constant increase in the tone discrimination capability. By aligning the data in terms of corrected age (Fig. 5 middle), we observe that the differences still remain. Moreover, the overlap regions of the effect sizes in the chronological age view are less prominent in the corrected age view. Similarly, in the total speech view (see Fig. 5 right), there are remaining differences between the two simulations. Hence, these results indicate that the differences are not due to an alignment of the simulated age or the speech used to train the models. Since the main difference between the two simulations is the duration of the PLE period, these findings suggest that the duration of the PLE simulation might affect the model behavior.

3.3. Finding 3: PLE does not improve quantitative compatibility with infant data with the present models

Fig. 6 displays compatibility of the models against empirical reference data from infants on the IDS preference and vowel discrimination tasks, and Fig. 7 shows the corresponding results for the tone discrimination and phonotactics preference tasks from Gonzalez-Gomez et al. (2021). Notice that there are no reference data for preterm infants in the meta-analyses of vowel discrimination and IDS preference; therefore, PT condition was not included in the analysis for those two tasks.

As seen from Fig. 6, the CPC model is more in line with infant data than the APC model. Nevertheless, both the APC and CPC models exhibit no difference in terms of the quantitative compatibility between FT and BS simulations. The APC model obtained compatibility with 4

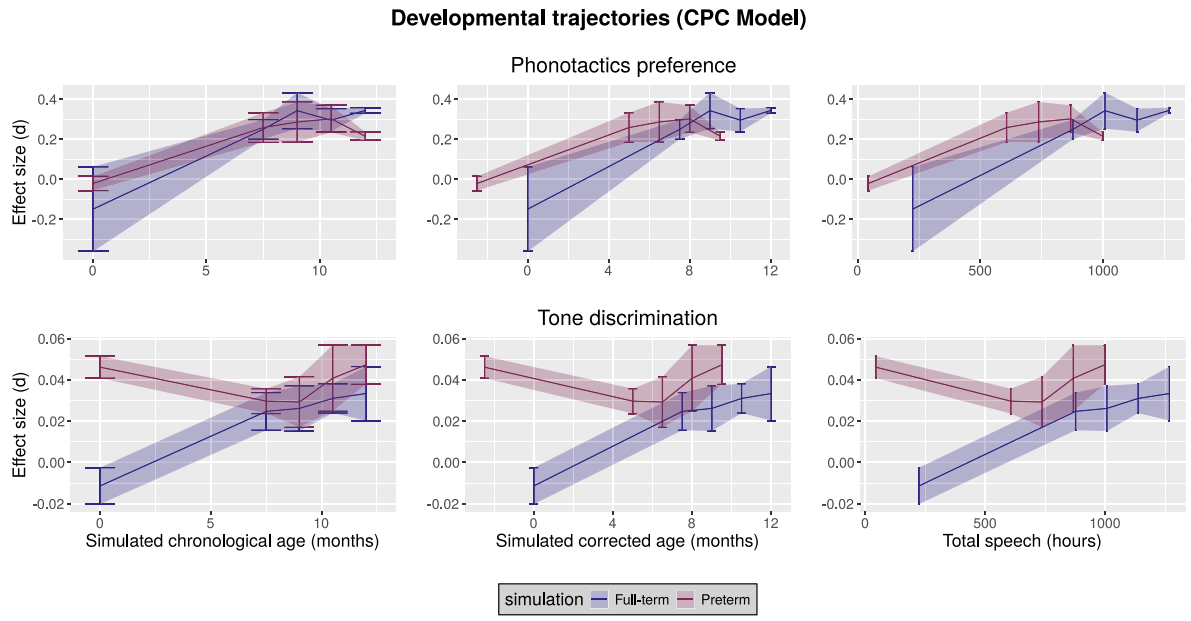

Fig. 5. Developmental trajectories for the CPC model of the phonotactics preference (top) and tone discrimination (bottom) tests. The trajectories are either represented using the chronological simulated age (left), corrected simulated age (middle), or the total amount of speech overall (PLE included) used for training (right). Each plot shows the trajectories of the FT learner (purple) and PT learner (dark pink). The error bars and shaded area represent ±1 standard error of the mean effect size across three runs. Note the differences in the scale of the effect size for both capabilities (y axis). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

	Infants (months)				APC (months)								CPC (months)							
	Full-term infants				Full-term simulation (FT)				Baseline simulation (BS)				Full-term simulation (FT)				Baseline simulation (BS)			
	7.5	9	10.5	12	7.5	9	10.5	12	7.5	9	10.5	12	7.5	9	10.5	12	7.5	9	10.5	12
Vowel discrimination	0.33	0.33	0.33	0.33	1.16	1.19	1.19	1.18	1.22	1.19	1.23	1.20	1.30	1.34	1.31	1.29	1.48	1.46	1.42	1.42
IDS preference	0.20	0.30	0.39	0.45	-0.66	-0.69	-0.58	-0.65	-0.68	-0.75	-0.69	-0.58	0.51	0.48	0.50	0.50	0.47	0.48	0.49	0.50
					Overall compatibility FT: 4/8 BS: 4/8								*Overall compatibility FT: 8/8 BS: 8/8*							

Fig. 6. Quantitative compatibility of the APC and CPC models and simulations with respect to the infant data on vowel discrimination and IDS preference. Compatibility is shown for the four age checkpoints of interest. Purple color marks the model effect sizes that are not compatible with the human data, whereas green marks the compatible ones. The overall compatibilities of the models and simulations are listed at the bottom. Note that the infant data set the lower bound of the confidence interval of the effect sizes, and for the models to be considered compatible with the reference data, they need to obtain an effect size larger or equal to the indicated lower bound (see Cruz Blandón et al., 2023). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

	Infants (months)				APC (months)								CPC (months)							
					PLE simulations				Baseline simulation (BS)				PLE simulations				Baseline simulation (BS)			
	7.5	9	10.5	12	7.5	9	10.5	12	7.5	9	10.5	12	7.5	9	10.5	12	7.5	9	10.5	12
Phonotactics preference FT	✗	✗	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Phonotactics preference PT	✗	✗	✓	✓	✗	✗	✓	✓	NA	NA	NA	NA	✓	✓	✓	✓	NA	NA	NA	NA
Tone discrimination FT	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tone discrimination PT	✓	✓	✓	✗	✓	✓	✓	✓	NA	NA	NA	NA	✓	✓	✓	✓	NA	NA	NA	NA
					Overall compatibility FT: 3/8 PT: 7/8 BS: 4/8								*Overall compatibility FT: 4/8 PT: 5/8 BS: 4/8*							

Fig. 7. Compatibility of the APC (middle) and CPC (right) models with infant data on phonotactics preference and tone discrimination with respect to infant reference data (left) from Gonzalez-Gomez et al. (2021). The comparison results are shown separately for PLE and BS conditions, comparing the full-term (FT) and preterm (PT) infants and conditions with each other. × indicates non-preference or discrimination skill and ✓ indicates a positive effect. Cells denoted with *NA* mark conditions without available data, as the BS simulation is how full-term learners have been simulated previously. The overall compatibilities are listed at the bottom of the model blocks.

out of 8 total age checkpoints, with full compatibility in vowel discrimination but failing on IDS preference. The CPC model is compatible with all the 8 checkpoints.

Fig. 7 shows no difference in compatibility of CPC between the FT and BS simulations for the phonotactics preference and tone discrimination tasks, but an increase of one checkpoint is observed for the APC model in the BS simulation against the FT simulation in the last age checkpoint (12 mo) of the phonotactic preference.

Both models exhibit positive tone discrimination capability for all age checkpoints. This is unexpected, given that the duration of PLE can affect models' behavior, and this task is related to prosodic features available in the filtered signal. However, these results suggest that,

unlike infants, both models continue exploiting lexical tone information for up to 12 simulated months regardless of the duration of PLE and despite the tones being non-phonemic in English (see Section 4). Thus, the differences between FT and BS are marginal or non-existent for both models. Interestingly, the PT condition obtained the highest compatibility among the three simulations for both models (APC: 7 out of 8; CPC: 5 out of 8).

Overall, there is no clear quantitative advantage for simulating the learners with PLE in terms of compatibility with infant data. This is in contrast to previous studies that suggested that PLE simulations are potentially more compatible models than models without PLE. Nevertheless, findings 1 and 2 showed that from the qualitative perspective, introducing PLE in the simulations led to changes in models' behavior with infant trends. This contrast highlights the challenges of assessing model alignment with infants' developmental timelines. Although visible differences emerged in the developmental trajectories of baseline and PLE simulations, the current simulations fall short of capturing precise linguistic behavior in the simulated age range (see also Discussion below).

Compared to the previous studies, the current work examined the compatibility of models against infant data by analyzing the temporal alignment of developmental trajectories with infant data for four linguistic capabilities, whereas Poli et al. (2024) assessed overall phonetic learning patterns as a continuum, comparing the ultimate learning outcomes in models with and without "before birth" attunement to auditory processing. Hence, the present work provides a broader view of how multiple different language capabilities are affected by PLE, including phonetic discrimination studied by Poli et al. (2024), but also making interpretation more challenging due to the attempt to explicitly connect models' and infants' developmental timelines instead of focusing on qualitative phone discrimination performance advantage per se. Our approach also differs from that of Vogelsang et al. (2023), who focused on the benefits of learning from low-pass filtered (simplified) speech input before full-band (rich) speech when solving a high-level perceptual task of speech emotion recognition. Given the disparity in their and our present tasks and in the general problem setting, direct comparison of the present results with Vogelsang et al. (2023) is not meaningful, but both studies show how prior exposure to simplified speech stimuli can impact later learning outcomes. Also, neither of the earlier studies used a combination of realistic amounts of prenatal speech, low-pass filtering motivated by womb attenuation and fetal hearing, and incremental learning over the finite speech data, further complicating direct comparisons.

4. Discussion

With the current experiments, we showed that plausible modeling of prenatal language exposure can impact the learning outcomes of computational models, thereby also potentially affecting the findings and conclusions of computational studies. Regardless of the variability observed in the simulations, the results consistently reveal substantial differences up to 7.5 months of simulated age for most of the capabilities, as well as changes in the direction of trajectories between PT and FT simulations beyond that point (e.g., an increasing or decreasing effect size trend with increasing simulated age). The results also showed that the duration of prenatal language exposure can affect the behavior of models. However, whether or not and how PLE duration affects models' behavior seems tied to the linguistic capability being tested. This result is aligned with the prosodic bootstrapping hypothesis that predicts delays or temporary disruption in infants' language development with different prenatal experience durations and upon information available during prenatal exposure (Gervain, 2018).

Previous studies suggest that potentially higher model compatibility with infant data could be obtained by including PLE simulation (Poli et al., 2024; Vogelsang et al., 2023), where the latter study also used the same CPC model as the present study. However, the current experimental setup and models show little compatibility with infant data and no difference between baseline and PLE simulations in this regard. Although we used models suitable for simulating statistical learners, as also demonstrated by other prior studies (Cruz Blandón et al., 2023; Lavechin et al., 2024), these models still fail to capture subtleties of infant language learning. The main differences between Poli et al. (2024) and Vogelsang et al. (2023) and our simulations is the more realistic prenatal simulation in terms of speech signal quality and quantity, and the method to explicitly evaluate models' compatibility with empirical infant data.

In our view, the current learning models and evaluation tools make analysis and comparison of temporal trajectories a valuable addition to the toolkit used to model infant developmental processes. Temporal comparison of learning trajectories across multiple concurrent capabilities enables a stringent assessment approach, helping to rule out models that diverge from expected infant behavior and supporting the development of more unified models of learning. In addition, well-fitting models could then be used to create hypotheses for infant age groups for which empirical data does not yet exist on a given phenomenon of interest. Yet, the exact matching of speech input hours between models and infants comes with many complications (including uncertainty and variability in how much speech infants actually hear; see Coffey, Räsänen, Scaff, & Cristia, 2024), and thereby less stringent temporal comparisons could also be applied, e.g., by comparing the shapes of the infants' and models' learning trajectories without requiring exact one-to-one mapping in terms of speech hours observed.

Apart from the differences in PLE simulation and evaluation methodology, our study employed purely incremental learning to simulate a human-like learning experience. This is in contrast to the usual machine learning training procedure, which involves multiple iterative training steps over the same dataset. In fact, we performed post-hoc tests to see whether more standard iterative training might have improved the models' linguistic competence further. By using a phoneme discrimination test from the ZeroSpeech Challenge evaluation toolkit (`abxLS test`, Nguyen et al., 2020), we calculated phoneme discrimination for all simulations and checkpoints. The APC and CPC models reached an ABX error rate between 9% and 13% regardless of the simulations. The obtained ABX error rates are higher than those reported for the same learner models in iterative training experiments with ABX scores of ≈6% (Liu et al., 2023), hence suggesting that the incremental training did limit the competence of the models to some degree. Consequently, further investigation into the effects of PLE in simulations with realistic settings, possibly with alternative learning curricula like regular machine learning training, could help mitigate training effects and more clearly unveil any consistent differences between BS and PLE simulations.

With the current methodological framework and evaluation, it may be possible to predict linguistic behavior for age checkpoints where infant data is currently unavailable. However, for the reasons outlined above, our current simulations – while suitable for examining PLE's influence on linguistic capabilities – did not achieve a level of compatibility sufficient for meaningful forecasts on human language acquisition. Nevertheless, this study establishes several key contributions that can support further exploration and refinement of models to more closely reflect infant language development, and eventually enable validation of models by testing hypotheses arising from modeling experiments.

5. Conclusions

This study investigated the inclusion of prenatal language exposure in computational modeling of infant learning. To this end, we first designed an experimental setup to simulate PLE in computational models of learning from speech audio. Two alternative statistical learner models, APC and CPC, were then used to simulate infants of different ages by training them with increasing amounts of speech audio, also

taking into account prenatal language exposure in the womb, and including both full-term and preterm infant simulations with different PLE speech exposures. The models were then tested on four linguistic capabilities for which reference empirical data from real infants was available.

Our findings revealed discernible differences between simulations that took PLE into account and the "standard" modeling approach, where statistical learning is assumed to start from birth. While the inclusion of PLE in the modeling setup did not improve the models' overall compatibility with the infant data, there were other notable differences between the learning trajectories of models with and without PLE taken into account. Thereby, the present experiments demonstrate that prenatal exposure in general, and duration of the PLE period in particular, influenced the learning trajectories of models in the postnatal part of the simulation. Therefore, PLE is a factor that should be considered in computational modeling studies of infant language learning. This is in agreement with behavioral evidence that language abilities may vary between full-term and preterm infants (Gonzalez-Gomez et al., 2021; Peña, Pittaluga, & Mehler, 2010; Sansavini et al., 2010), and with the hypothesis that prenatal experience may have consequences on language learning (Gervain, 2018).

Our simulation approach, although aiming for a more realistic simulation than the earlier attempts, naturally has its limitations. First, the gradual development of the fetal auditory system and notable uncertainties involved in modeling the pathway from the mother's speech production system to the inner ear of a fetus mean that the present approach is far from perfect. For instance, our setup used the data on fetal auditory stimulation based on a small-sample sheep model, and it completely ignored the ambient noise environment and, thereby, the speech-to-noise ratio in the womb (see also discussion in Poli et al., 2024). Moreover, direct sound conduction through tissues of the mother (e.g., from the vocal tract to the lungs and from there to the uterus) is not included in the model, where the attenuation is measured with respect to a reference level outside the mother. Similarly, our simulation of the NICU exposure focused solely on the amount of speech preterm infants hear, whereas actual NICU auditory environments include additional factors, such as predominant electronic sounds from medical equipment (Monson et al., 2023) or elevated sound levels (see Best, Bogossian, & New, 2018, for a review). Incorporating these elements could amplify the differences between preterm and full-term models, but would also require a much more complex model for the auditory environment to account for the non-speech input at NICU but ideally also in utero and at home.

In addition, training of models involved exposure to an utterance only once. While this reflects natural human language exposure, it poses challenges for the present type of machine learning-based computational models that are typically trained iteratively several times across the same set of data. Despite the several hundreds of hours of speech data available for learning, the models did not reach their full potential in terms of linguistic capabilities, as demonstrated in earlier studies (see the discussion in the previous section). In addition, the current simulation used speech data from read-aloud audiobooks due to their high audio quality and quantity. While often applied to similar modeling studies (Cruz Blandón et al., 2023; Khorrami et al., 2023; Lavechin et al., 2024), this type of speech is not directly comparable with spontaneous speech heard by fetuses during the prenatal period or with the infant-directed speech heard by infants in their everyday environments. Therefore, future modeling studies should tackle these limitations in order to understand better how different factors and assumptions contribute to the modeling outcomes. Finally, the employed statistical learner models, although successful in explaining aspects of phonetic and lexical learning (Khorrami et al., 2023; Lavechin et al., 2024), are still far from being good models of infant learners. As already shown in Cruz Blandón et al. (2023), these simplistic models struggle to fit the developmental changes observed in infants with age, and the inclusion of PLE in the simulations does not change the situation completely. Hence, more work is also needed in terms of model development.

Overall, the main contribution of this study is to provide a framework for simulating prenatal language exposure in contemporary computational models of infant language acquisition. At the same time, the study highlights the challenges inherent in implementing more realistic simulations of language acquisition and the relevance of transitioning towards a naturalistic setting for models of language learners. We believe such a transition will allow the study of research questions that may otherwise be impractical to investigate through behavioral research. Finally, this work also contributes more widely to the developmental field by highlighting the importance of prenatal language exposure in language development.

CRediT authorship contribution statement

María Andrea Cruz Blandón: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nayeli Gonzalez-Gomez:** Writing – review & editing, Resources, Methodology. **Marvin Lavechin:** Writing – review & editing, Resources, Data curation. **Okko Räsänen:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to improve readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgments

Authors MACB and OR were supported by Academy of Finland, Finland grants no. 314602, 320053, and 345365, and by Kone Foundation grant L-SCALE. NGG was funded by The Leverhulme Trust, United Kingdom (ECF-2015-009). ML was partially supported by MIAI at Grenoble Alpes (ANR-19-P3IA-0003). We would like to thank our colleagues for the insightful discussions at LAAC lab and SpeCog Research Group, from which this manuscript benefited. We also thank ENS and LAAC lab for sharing the preprocessed training data we used in the simulations. The code related to this work is available on https://github.com/SPEECHCOG/ple_exploration_study

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cognition.2024.106044.

Data availability

The code is available on a GitHub repository at https://github.com/SPEECHCOG/ple_exploration_study Data used for training the models is openly available and linked in the manuscript.

References

Aslin, R. N., & Newport, E. L. (2014). Distributional Language Learning: Mechanisms and Models of Category Formation. *Language Learning*, *64*(s2), 86–105. http://dx.doi.org/10.1111/lang.12074, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/lang.12074.

Bergmann, C., Frank, M. C., Gonzalez, N., Bergelson, E., Cristia, A., Ferguson, B., et al. (2023). ManyBabies 1: Infant-directed speech preference [dataset]. URL osf.io/re95x.

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., et al. (2018). Promoting Replicability in Developmental Research Through Meta-analyses: Insights From Language Acquisition Research. *Child Development*, *89*(6), 1996–2009. http://dx.doi.org/10.1111/cdev.13079.

Best, K., Bogossian, F., & New, K. (2018). Language Exposure of Preterm Infants in the Neonatal Unit: A Systematic Review. *Neonatology*, *114*(3), 261–276. http://dx.doi.org/10.1159/000489600, arXiv:https://karger.com/neo/article-pdf/114/3/261/3237184/000489600.pdf.

Birnholz, J. C., & Benacerraf, B. R. (1983). The development of human fetal hearing. *Science*, *222*(4623), 516–518. http://dx.doi.org/10.1126/science.6623091, arXiv:https://www.science.org/doi/pdf/10.1126/science.6623091 URL https://www.science.org/doi/abs/10.1126/science.6623091.

Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein, A., Alam, F., et al. (2021). A cross-cultural examination of young children's everyday language experiences. *PsyArXiv*, http://dx.doi.org/10.31234/osf.io/723pr, version 3.

Chelli, D., & Chanoufi, B. (2008). Audition føe tale. Mythe ou réalité? *Journal de Gynécologie Obstétrique et Biologie de la Reproduction*, *37*(6), 554–558. http://dx.doi.org/10.1016/j.jgyn.2008.06.007, URL https://www.sciencedirect.com/science/article/pii/S0368231508002275.

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., et al. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, *82*(1), 53–58. http://dx.doi.org/10.1016/0378-5955(94)00164-L, URL https://www.sciencedirect.com/science/article/pii/037855594000164L.

Chung, Y. A., Hsu, W. N., Tang, H., & Glass, J. (2019). An unsupervised autoregressive model for speech representation learning. In *Proceedings of the annual conference of the international speech communication association (interspeech)* (pp. pp. 146–150). http://dx.doi.org/10.21437/Interspeech.2019-1473.

Coffey, J., Räsänen, O., Scaff, C., & Cristia, A. (2024). The difficulty and importance of estimating the lower and upper bounds of infant speech exposure. In *Interspeech 2024* (pp. 3615–3619). http://dx.doi.org/10.21437/Interspeech.2024-2102.

Cruz Blandón, M. A., Cristia, A., & Räsänen, O. (2023). Introducing meta-analysis in the evaluation of computational models of infant language development. *Cognitive Science*, *47*(7), Article e13307. http://dx.doi.org/10.1111/cogs.13307, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cogs.13307 URL https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.13307.

de Seyssel, M., Lavechin, M., & Dupoux, E. (2023). Realistic and broad-scope learning simulations: first results and challenges. *Journal of Child Language*, *50*(6), 1294–1317. http://dx.doi.org/10.1017/S0305000923000272.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176. http://dx.doi.org/10.1126/science.7375928, arXiv:https://www.science.org/doi/pdf/10.1126/science.7375928 URL https://www.science.org/doi/abs/10.1126/science.7375928.

Dupoux, E. (2018). Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition*, *173*, 43–59. http://dx.doi.org/10.1016/j.cognition.2017.11.008, URL https://www.sciencedirect.com/science/article/pii/S0010027717303013.

Eggermont, J. J., & Moore, J. K. (2012). Morphological and functional development of the auditory nervous system. In L. Werner, R. R. Fay, & A. N. Popper (Eds.), *Human auditory development* (pp. 61–105). New York, NY: Springer New York, http://dx.doi.org/10.1007/978-1-4614-1421-6_3.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*(1), 71–99.

Erhan, D., Courville, A., Bengio, Y., & Vincent, P. (2010). Why does unsupervised pre-training help deep learning? In Y. W. Teh, & M. Titterington (Eds.), *Proceedings of machine learning research*: *vol. 9*, *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 201–208). Chia Laguna Resort, Sardinia, Italy: PMLR, URL https://proceedings.mlr.press/v9/erhan10a.html.

Frank, M. C., Alcock, K. J., Arias-Trejo, N., Aschersleben, G., Baldwin, D., Barbu, S., et al. (2020). Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24–52. http://dx.doi.org/10.1177/2515245919900809.

Gélat, P., David, A. L., Haqhenas, S. R., Henriques, J., Thibaut de Maisieres, A., White, T., et al. (2019). Evaluation of fetal exposure to external loud noise using a sheep model: quantification of in utero acoustic transmission across the human audio range. *American Journal of Obstetrics and Gynecology*, *221*(4), 343.e1–343.e11. http://dx.doi.org/10.1016/j.ajog.2019.05.036, URL https://www.sciencedirect.com/science/article/pii/S0002937819307082.

Gelderloos, L., Kamelabad, A. M., & Alishahi, A. (2020). Active word learning through self-supervision. In *Proceedings of the 42nd annual meeting of the cognitive science society*.

Gerhardt, K. J., & Abrams, R. M. (1996). Fetal hearing: Characterization of the stimulus and response. *Seminars in Perinatology*, *20*(1), 11–20. http://dx.doi.org/10.1016/S0146-0005(96)80053-X, URL https://www.sciencedirect.com/science/article/pii/S014600059680053X, Vibration Exposure in Pregnancy.

Gervain, J. (2018). The role of prenatal experience in language development. *Current Opinion in Behavioral Sciences*, *21*, 62–67. http://dx.doi.org/10.1016/j.cobeha.2018.02.004, URL https://www.sciencedirect.com/science/article/pii/S2352154617301365, The Evolution of Language.

Gonzalez-Gomez, N., O'Brien, F., & Harris, M. (2021). The effects of prematurity and socioeconomic deprivation on early speech perception: A story of two different delays. *Developmental Science*, *24*(2), Article e13020. http://dx.doi.org/10.1111/desc.13020, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/desc.13020 URL https://onlinelibrary.wiley.com/doi/abs/10.1111/desc.13020.

Graven, S. N., & Browne, J. V. (2008). Auditory development in the fetus and infant. *Newborn and Infant Nursing Reviews*, *8*(4), 187–193. http://dx.doi.org/10.1053/j.nainr.2008.10.010, URL https://www.sciencedirect.com/science/article/pii/S1527336908001347, Brain Development of the Neonate.

Gutmann, M., & Hyvärinen, A. (2010). Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. *vol. 9*, In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (pp. pp. 297–304).

Hepper, P. G., Scott, D., & Shahidullah, B. S. (1993). Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, *11*(3), 147–153. http://dx.doi.org/10.1080/02646839308403210, arXiv:https://doi.org/10.1080/02646839308403210.

Hepper, P. G., & Shahidullah, B. S. (1994). Development of fetal hearing. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *71*(2), F81–F87. http://dx.doi.org/10.1136/fn.71.2.F81, arXiv:https://fn.bmj.com/content/71/2/F81.full.pdf URL https://fn.bmj.com/content/71/2/F81.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Acoustic Characteristics of American English Vowels*, *97*(5), 3099–3111. http://dx.doi.org/10.1121/1.411872.

Holst, M., Eswaran, H., Lowery, C., Murphy, P., Norton, J., & Preissl, H. (2005). Development of auditory evoked fields in human fetuses and newborns: A longitudinal MEG study. *Clinical Neurophysiology*, *116*(8), 1949–1955. http://dx.doi.org/10.1016/j.clinph.2005.04.008, URL https://www.sciencedirect.com/science/article/pii/S1388245705001434.

Huebner, P. A., Sulem, E., Fisher, C., & Roth, D. (2021). BabyBERTa: Learning more grammar with small-scale child-directed language. In A. Bisazza, & O. Abend (Eds.), *Proceedings of the 25th conference on computational natural language learning* (pp. 624–646). Online: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.conll-1.49, URL https://aclanthology.org/2021.conll-1.49.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*(5), 630–645. http://dx.doi.org/10.1006/jmla.1994.1030, URL https://www.sciencedirect.com/science/article/pii/S0749596X84710308.

Kearns, J. (2014). Librivox: Free public domain audiobooks [dataset]. *Reference Reviews*, *28*(1), 7–8, URL https://librivox.org/.

Khorrami, K. (2024). *Computational modeling of early language acquisition with multimodal neural networks* (Ph.D. thesis), Tampere University.

Khorrami, K., Cruz Blandón, M. A., & Räsänen, O. (2023). Computational insights to acquisition of phonemes, words, and word meanings in early language: Sequential or parallel acquisition? In *Proceedings of the 45th annual meeting of the cognitive science society*.

Khorrami, K., & Räsänen, O. (2021). Can phones, syllables, and words emerge as side-products of cross-situational audiovisual learning? - A computational investigation. *Language Development Research*, *1*, 123–191. http://dx.doi.org/10.34842/w3vw-s845, URL https://lps.library.cmu.edu/LDR/article/id/434/.

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. URL https://arxiv.org/abs/1412.6980.

Lavechin, M., de Seyssel, M., Gautheron, L., Dupoux, E., & Cristia, A. (2022). Reverse engineering language acquisition with child-centered long-form recordings. *Annual Review of Linguistics*, *8*(Volume 8, 2022), 389–407. http://dx.doi.org/10.1146/annurev-linguistics-031120-122120, URL https://www.annualreviews.org/content/journals/10.1146/annurev-linguistics-031120-122120.

Lavechin, M., de Seyssel, M., Métais, M., Metze, F., Mohamed, A., Bredin, H., et al. (2024). Modeling early phonetic acquisition from child-centered audio data. *Cognition*, *245*, Article 105734. http://dx.doi.org/10.1016/j.cognition.2024.105734, URL https://www.sciencedirect.com/science/article/pii/S0010027724000209.

Lavechin, M., de Seyssel, M., Titeux, H., Bredin, H., Wisniewski, G., Cristia, A., et al. (2022). Can statistical learning bootstrap early language acquisition? A modeling investigation. *PsyArXiv*, http://dx.doi.org/10.31234/osf.io/rx94d, version 4.

Lavechin, M., Sy, Y., Titeux, H., Cruz Blandón, M. A., Räsänen, O., Bredin, H., et al. (2023). BabySLM: language-acquisition-friendly benchmark of self-supervised spoken language models. In *Proc. iNTERSPEECH 2023* (pp. 4588–4592). http://dx.doi.org/10.21437/Interspeech.2023-978.

Lecanuet, J.-P., Gautheron, B., Locatelli, A., Schaal, B., Jacquet, A.-Y., & Busnel, M.-C. (1998). What sounds reach fetuses: Biological and nonbiological modeling of the transmission of pure tones. *Developmental Psychobiology*, *33*(3), 203–219. http://dx.doi.org/10.1002/(SICI)1098-2302(199811)33:3<203::AID-DEV2>3.0.CO;2-V, arXiv:https://onlinelibrary.wiley.com/doi/

pdf/10.1002/(SICI)1098-2302(199811)33:3<203::AID-DEV2>3.0.CO;2-V URL https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1098-2302(199811)33:3<203::AID-DEV2>3.0.CO;2-V.

Lee, G. Y., & Kisilevsky, B. S. (2014). Fetuses respond to father's voice but prefer mother's voice after birth. *Developmental Psychobiology*, *56*(1), 1–11. http://dx.doi.org/10.1002/dev.21084, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/dev.21084 URL https://onlinelibrary.wiley.com/doi/abs/10.1002/dev.21084.

Lewis, M., Braginsky, M., Tsuji, S., Bergmann, C., Piccinini, P. E., Cristia, A., et al. (2016). A quantitative synthesis of early language acquisition using meta-analysis. *PsyArXiv*, http://dx.doi.org/10.31234/osf.io/htsjm, version 3.

Liu, O. D., Tang, H., & Goldwater, S. (2023). Self-supervised Predictive Coding Models Encode Speaker and Phonetic Information in Orthogonal Subspaces. In *Proc. iNTERSPEECH 2023* (pp. 2968–2972). http://dx.doi.org/10.21437/Interspeech.2023-871.

MacDonald, K., Räsänen, O., Casillas, M., & Warlaumont, A. S. (2020). Measuring prosodic predictability in children's home language environments. In *Proceedings of the 42nd annual meeting of the cognitive science society*.

Merkx, D., Scholten, S., Frank, S. L., Ernestus, M., & Scharenborg, O. (2023). Modelling human word learning and recognition using visually grounded speech. *Cognitive Computation*, *15*(1), 272–288. http://dx.doi.org/10.1007/s12559-022-10059-7.

Monson, B. B., Ambrose, S. E., Gaede, C., & Rollo, D. (2023). Language exposure for preterm infants is reduced relative to fetuses. *The Journal of Pediatrics*, http://dx.doi.org/10.1016/j.jpeds.2022.12.042, URL https://www.sciencedirect.com/science/article/pii/S0022347623000550.

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paediatrica*, *102*(2), 156–160. http://dx.doi.org/10.1111/apa.12098, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/apa.12098 URL https://onlinelibrary.wiley.com/doi/abs/10.1111/apa.12098.

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American sign language. *Language Sciences*, *10*(1), 147–172. http://dx.doi.org/10.1016/0388-0001(88)90010-1, URL https://www.sciencedirect.com/science/article/pii/0388000188900101.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*(1), 11–28. http://dx.doi.org/10.1207/s15516709cog1401_2, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog1401_2 URL https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog1401_2.

Nguyen, T. A., de Seyssel, M., Rozé, P., Rivière, M., Kharitonov, E., Baevski, A., et al. (2020). The zero resource speech benchmark 2021: Metrics and baselines for unsupervised spoken language modeling. *arXiv*.

Nikolaus, M., Alishahi, A., & Chrupała, G. (2022). Learning English with Peppa Pig. *Transactions of the Association for Computational Linguistics*, *10*, 922–936. http://dx.doi.org/10.1162/tacl_a_00498, arXiv:https://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00498/2042609/tacl_a_00498.pdf.

van den Oord, A., Li, Y., & Vinyals, O. (2018). Representation Learning with Contrastive Predictive Coding. *Computing Research Repository*, abs/1807.0 URL https://arxiv.org/abs/1807.03748.

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. *Proceedings of the National Academy of Sciences*, *110*(37), 15145–15150. http://dx.doi.org/10.1073/pnas.1302159110, arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1302159110 URL https://www.pnas.org/doi/abs/10.1073/pnas.1302159110.

Peña, M., Pittaluga, E., & Mehler, J. (2010). Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences*, *107*(8), 3823–3828. http://dx.doi.org/10.1073/pnas.0914326107, arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.0914326107 URL https://www.pnas.org/doi/abs/10.1073/pnas.0914326107.

Poli, M., Schatz, T., Dupoux, E., & Lavechin, M. (2024). Modeling the initial state of early phonetic learning in infants. *Language Development Research*, *5*, http://dx.doi.org/10.34842/y89t-6q31, URL https://lps.library.cmu.edu/LDR/article/id/717/.

Pujol, R., Lavigne-rebillard, M., & Uziel, A. (1991). Development of the human cochlea. *ACTA Oto-laryngologica*, *111*(sup482), 7–13. http://dx.doi.org/10.3109/00016489109128023, arXiv:https://doi.org/10.3109/00016489109128023.

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crèpin, G. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, *28*(3), 191–212. http://dx.doi.org/10.1016/0028-2243(88)90030-5, URL https://www.sciencedirect.com/science/article/pii/0028224388900305.

Räsänen, O., Kakouros, S., & Soderstrom, M. (2018). Is infant-directed speech interesting because it is surprising? – Linking properties of IDS to statistical learning and attention at the prosodic level. *Cognition*, *178*, 193–206. http://dx.doi.org/10.1016/j.cognition.2018.05.015, URL https://www.sciencedirect.com/science/article/pii/S0010027718301355.

Richards, D. S., Frentzen, B., Gerhardt, K. J., McCann, M. E., & Abrams, R. M. (1992). Sound levels in the human uterus. *Obstetrics and gynecology*, *80*(2), 186–190, URL http://europepmc.org/abstract/MED/1635729.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. http://dx.doi.org/10.1126/science.274.5294.1926, URL https://science.sciencemag.org/content/274/5294/1926.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. *Annual Review of Psychology*, *69*(1), 181–203. http://dx.doi.org/10.1146/annurev-psych-122216-011805.

Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The Infant's Auditory World: Hearing, Speech, and the Beginnings of Language. *vol. 2*, In *Handbook of child psychology: cognition, perception, and language* (6th ed.). (pp. 58–108). Hoboken, NJ, US: John Wiley & Sons, Inc..

Sansavini, A., Guarini, A., Justice, L. M., Savini, S., Broccoli, S., Alessandroni, R., et al. (2010). Does preterm birth increase a child's risk for language impairment? *Early Human Development*, *86*(12), 765–772. http://dx.doi.org/10.1016/j.earlhumdev.2010.08.014, URL https://www.sciencedirect.com/science/article/pii/S0378378210002161.

Schatz, T. (2016). *ABX-discriminability measures and applications* (Ph.D. thesis), Université Paris 6 (UPMC).

Schatz, T., Peddinti, V., Cao, X. N., Bach, F., Hermansky, H., & Dupoux, E. (2014). Evaluating speech features with the Minimal-Pair ABX task (II): Resistance to noise. In *Proceedings of the annual conference of the international speech communication association (interspeech)* (pp. pp. 915–919).

Turkewitz, G., & Kenny, P. A. (1982). Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. *Developmental Psychobiology*, *15*(4), 357–368. http://dx.doi.org/10.1002/dev.420150408, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/dev.420150408 URL https://onlinelibrary.wiley.com/doi/abs/10.1002/dev.420150408.

Vince, M. A., Armitage, S. E., Baldwin, B. A., Toner, J., & Moore, B. C. J. (1982). The sound environment of the foetal sheep. *Behaviour*, *81*(2/4), 296–315, URL http://www.jstor.org/stable/4534209.

Vince, M. A., Billing, A. E., Baldwin, B. A., Toner, J. N., & Weller, C. (1985). Maternal vocalisations and other sounds in the fetal lamb's sound environment. *Early Human Development*, *11*(2), 179–190. http://dx.doi.org/10.1016/0378-3782(85)90105-7, URL https://www.sciencedirect.com/science/article/pii/0378378285901057.

Vogelsang, M., Vogelsang, L., Diamond, S., & Sinha, P. (2023). Prenatal auditory experience and its sequelae. *Developmental Science*, *26*(1), Article e13278. http://dx.doi.org/10.1111/desc.13278, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/desc.13278 URL https://onlinelibrary.wiley.com/doi/abs/10.1111/desc.13278.

Vogelsang, L., Vogelsang, M., Pipa, G., Diamond, S., & Sinha, P. (2024). Butterfly effects in perceptual development: A review of the 'adaptive initial degradation' hypothesis. *Developmental Review*, *71*, Article 101117. http://dx.doi.org/10.1016/j.dr.2024.101117, URL https://www.sciencedirect.com/science/article/pii/S0273229724000017.

Yang, G.-P., Yeh, S.-L., Chung, Y.-A., Glass, J., & Tang, H. (2022). Autoregressive predictive coding: A comprehensive study. *IEEE Journal of Selected Topics in Signal Processing*, *16*(6), 1380–1390. http://dx.doi.org/10.1109/JSTSP.2022.3203608.

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, *68*(2), 123–139. http://dx.doi.org/10.1016/j.jml.2012.09.004, URL https://www.sciencedirect.com/science/article/pii/S0749596X12001052.