

Modeling Prosodic Development with Prenatal Audio Attenuation

Frank Lihui Tan, Shuang Zheng, Ming Liu, Youngah Do

The University of Hong Kong

1 Introduction

Accumulating evidence from developmental research indicates that speech acquisition commences well before birth, establishing the foundations upon which postnatal language acquisition unfolds (Gervain, 2018). By approximately the 20th week of gestation, the fetal auditory system becomes sufficiently mature to detect and process external sounds (Eggermont & Moore, 2012; Graven & Browne, 2008), raising the possibility that speech in the surrounding environment may shape the earliest phases of speech perception. However, the auditory signals available in utero are markedly different from what infants encounter after birth. The sounds reaching the fetal ear are heavily filtered by maternal tissues, amniotic fluid, and body structures, resulting in the attenuation of higher-frequency components of speech, while primarily retaining low-frequency acoustic energy below approximately 500 Hz (Gélat et al., 2019; Hepper & Shahidullah, 1994). Accordingly, much of the segmental detail of speech, particularly the information carried by frication, bursts, and formants in the high-frequency range, is inaccessible to prenatal fetuses. The remaining signals predominantly convey prosodic information, such as rhythmic and intonational patterns (Querleu et al., 1988). Thus, while the fetus receives sustained exposure to speech, the content of that exposure is highly constrained.

Corresponding to this prenatal biological environment, human neonates demonstrate a remarkable degree of prosody-related competence. Newborns can discriminate between speech in native and non-native languages (Mehler et al., 1988; Moon et al., 1993; Nazzi, Bertocini, et al., 1998) and produce cries with melodic contours mirroring the prosodic patterns of their native language (Mampe et al., 2009). Additionally, neonates can identify word or phrase boundaries (Christophe et al., 1994, 2001; Fló et al., 2019) and distinguish syllables based on intonational and stress patterns (Martinez-Alvarez et al., 2023; Nazzi, Floccia, et al., 1998; Sansavini et al., 1997). These findings suggest that exposure to the prosodic information available in the prenatal environment contributes to early prosodic learning, even before infants have full access to the wider acoustic spectrum postnatally. Studies further indicate that the shortened prenatal exposure among preterm infants is reflected in their later prosodic processing. Preterm infants are found to be less competent in the discrimination of rhythm (Bosch, 2011; Peña et al., 2010), stress (Herold et al., 2008; Ragó et al., 2014), and tone (Gonzalez-Gomez et al., 2021). For example, Gonzalez-Gomez et al. (2021) tracked preterm and full-term infants longitudinally from 7.5 to 12 months using an intermodal preferential looking procedure and discovered a delay in the perceptual narrowing of a non-native Cantonese tone contrast in the preterm group compared to the full-term group.

Developmental and socioeconomic factors have been proposed to explain the learning disadvantages of preterm infants. First, prematurity is often associated with an increased risk of neural injury and atypical brain development (Back & Miller, 2014; Volpe, 2009). Such disruptions may harm the neural systems supporting auditory processing, attention, and learning, even in the absence of overt neurological impairment. Moreover, preterm birth is frequently accompanied by nutritional challenges, including deficits in essential nutrients required for brain growth and myelination, which have been linked to later cognitive and language performance (Coviello et al., 2018). In addition, socioeconomic factors have been shown to covary with both the likelihood of preterm birth and aspects of the early language environment, including parenting style and the extent of parent-child language interaction. These environmental factors have been further demonstrated to affect children's language outcomes (Gonzalez-Gomez et al., 2021; Raviv et al., 2004; Smith et al., 2007).

* Frank Lihui Tan and Shuang Zheng contributed equally to this work and share first authorship. We would like to thank the AMP reviewers, participants at AMP 2025, and members of the HKU Language Development Lab for all the invaluable feedback.

On the other hand, the delayed development of prosodic learning may be straightforwardly related to the quantity and quality of the linguistic input, such as the overall reduction in speech input experienced by preterm infants. Monson et al. (2023) showed that preterm infants receive markedly less nearby human speech than fetuses in utero and are instead exposed to more electronic noise, environmental sounds, and prolonged silence. These cumulative deficits, exceeding 150 hours over the preterm period, may limit the opportunities for preterm infants to consolidate early prosodic regularities. Another potential linguistic explanation for the prosodic processing impairments of preterm infants concerns the concept of starting-small learning effect (Elman, 1993; Newport, 1990; Turkewitz & Kenny, 1982). According to these accounts, early learners benefit from restricted input because reduced complexity allows the developing system to stabilize coarse-grained regularities before being exposed to the full richness of the signals. From this perspective, prenatal exposure predominantly to the low-frequency range of speech may support the consolidation of overarching rhythmic and intonational patterns. Preterm infants, by contrast, enter the postnatal environment prematurely and are immediately confronted with the full spectral complexity of speech, which may disperse attentional resources, causing difficulty in establishing prosodic regularities at the outset of learning.

The present study utilizes computational modeling to investigate how the duration of prenatal auditory exposure may influence later prosodic learning. Specifically, we examine how the total amount of speech input impacts the learning outcomes by altering the training durations, while exploring whether the characteristics of the prenatal input being low-pass filtered play a role in prosodic learning by manipulating the proportion of low-frequency exposure across conditions. Motivated by proposals that early learning benefits from initial restricted input, and by empirical work suggesting that preterm infants with insufficient exposure to low-frequency cues have impaired prosodic development, we posit that longer exposure to low-pass filtered speech facilitates the acquisition of prosodic categories. At the same time, our modeling approach allows us to abstract away from non-linguistic factors, such as medical and environmental factors, and directly test whether the linguistic input in isolation can account for the development of infants' early prosodic processing abilities.

Within the consideration of the linguistic influence on preterm infants' language development, we must address a methodological issue in the empirical literature: the measurement of developmental age. Much of the earlier work documenting disadvantages in preterm infants' prosodic learning relied on birth age or chronological age, implicitly assuming that developmental time begins at birth. However, if prenatal auditory experience contributes meaningfully to prosodic learning as we hypothesized, this practice introduces a systematic bias against preterm infants, as they have less exposure to prenatal sound compared to full-term infants at the same birth age. More recent studies adopted the term age measure, aligning infants by their expected gestations (Bosch, 2011; Herold et al., 2008; Peña et al., 2010). Nevertheless, because preterm infants encounter the full spectral richness of postnatal sound earlier than full-term infants, the difference between groups is contingent on whether early access to full-frequency input facilitates or disrupts learning. In fact, Peña et al. (2010) directly compared the two age metrics and found that the apparent preterm disadvantage emerged only when infants were matched by birth age, but not when they were compared by term age. Given the previous results, we examine learning outcomes under both age metrics.

Convolutional neural networks were trained to perform stress and tone classification, respectively, representing two prosodic domains that largely survive low-pass attenuation in utero. The networks first received low-frequency inputs for varying durations, simulating different durations of prenatal exposure at the time of birth. Training then shifted to full-frequency inputs to approximate the transition to the richer postnatal acoustic environment. We evaluated the models' performance at multiple checkpoints across this developmental trajectory to address two questions: (i) whether longer low-frequency exposure leads to measurable advantages in prosodic learning, and (ii) how long such advantages, if present, persist into later postnatal learning.

2 Method

2.1 Dataset The stress dataset was derived from the *train-clean-100* subset of the LibriSpeech corpus (Panayotov et al., 2015), which contains approximately 100 hours of read speech produced by 251 English speakers at a sampling rate of 16 kHz. Three stress categories were included in this study: *unstressed*, *primary stress*, and *secondary stress*. Syllable-level segmentation and stress labels were taken from the annotations provided by Lugosch et al. (2019). The tone dataset was sourced from AISHELL-3 (Shi et al., 2021), a high-quality Mandarin corpus comprising roughly 85 hours of read utterances

recorded by 218 native speakers. Audio files in the Mandarin corpus were sampled at 44.1 kHz. All four lexical tones of Mandarin were included. Syllable-level segmentations for Mandarin were generated using the Montreal Forced Aligner (McAuliffe et al., 2017) with its pretrained Mandarin acoustic model.

The extracted syllable recordings in both datasets were further standardized through the following preprocessing pipeline. First, we generated a low-frequency version of each dataset by applying a hard-cut 500-Hz low-pass filter implemented with NumPy (Harris et al., 2020), following the methodology outlined in Vogelsang et al. (2023). The corresponding full-frequency version retained the original signal without any filtering. Subsequently, all syllables in both low-frequency and full-frequency versions were padded to a fixed length of 250 milliseconds by adding low-amplitude Gaussian noise to both their onsets and their offsets; the allocation of padding to onsets and offsets was randomized. The duration normalization ensures that all spectrograms have identical temporal dimensions, thereby satisfying the fixed-size input requirement of the convolutional network. After the filtering and the duration normalization, each token was transformed into a Log-Mel spectrogram (in dB) using Librosa (McFee et al., 2015). The Log-Mel spectrograms were computed with a 64-millisecond FFT window and a 2-millisecond hop length, yielding 128 Log-Mel bins and 126 temporal frames per token. A sample Log-Mel spectrogram in low-frequency and full-frequency versions is shown in Figure 1.

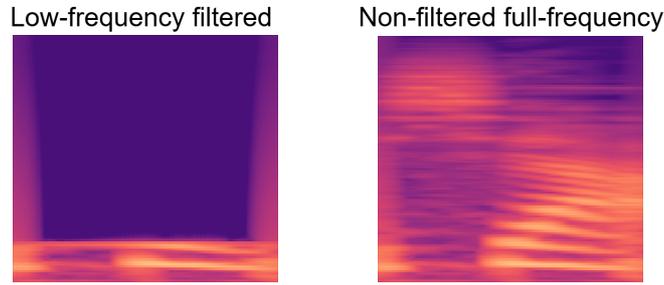


Figure 1: An illustration of the low-pass filtered and full-frequency Log-Mel spectrograms.

2.2 Model architecture A convolutional neural network (CNN), detailed in Figure 2, was trained to classify syllable-level Log-Mel spectrograms into stress or tone categories. The network consists of two convolutional blocks followed by a fully connected classifier, taking a single-channel spectrogram of the size 128×126 (Log-Mel bins \times time frames) as input. Both training and evaluation were performed with a batch size of 128. The first convolutional block employs a two-dimensional convolution with 16 filters (3×3 kernel, 1 pixel stride, 1 pixel padding), followed by batch normalization, a ReLU non-linearity, and 2×2 max pooling. The procedure transforms the input spectrogram into 16 feature maps while halving both the frequency and the temporal dimensions of each feature map by selecting the maximum activation within each non-overlapping 2×2 window. The second block repeats the same convolutional structure with 32 feature maps. The output of the second block is further passed through an adaptive pooling layer that down-samples each feature map to a 4×4 spatial grid. The pooled feature maps, comprising 32 channels of 4×4 spatial grids, are then flattened and fed into a fully connected layer with 64 hidden units, followed by batch normalization, ReLU activation, and dropout with a rate of 0.5 to mitigate overfitting. The final output layer involves a linear projection from the 64-dimensional representation generated by the previous layer to a label space of size 3 or 4, corresponding to the number of classes in the task (3 stress categories or 4 tone categories). During inference, class predictions are obtained by applying a softmax over the output logits and selecting the argmax class.

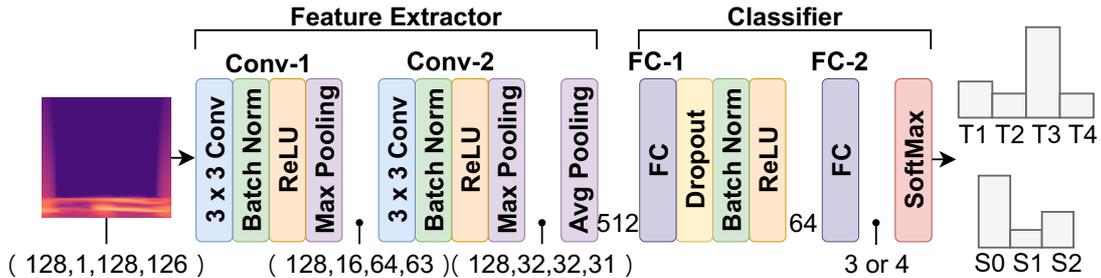


Figure 2: An illustration of model architecture. Data structure is (batch size, channel, height, width).

2.3 Training The training consisted of two stages, designed to approximate prenatal and postnatal auditory experience. In the prenatal stage, the model was trained exclusively on the low-frequency version of the inputs, simulating the acoustically attenuated environment of the womb. In the postnatal stage, training continued with the full-frequency version of the data, reflecting access to the complete acoustic spectrum after birth. We manipulated the durations of the prenatal stage to investigate how they shape subsequent learning. Five conditions were implemented, with prenatal training lasting 20, 40, 60, 80, or 100 epochs, while the total number of training epochs held constant at 200. This allowed us to vary the proportion of early low-frequency exposure while equating the overall training time, or analogously, the term age of human infants. At the same time, we could extract the model performance at the same postnatal checkpoint, parallel to the birth age of human infants.

Each experimental condition was repeated fifteen times with independently initialized models to account for random variability. We applied Kaiming initialization (He et al., 2015) to all convolutional and linear layers at the beginning of each independent run of the model. This procedure produced a distribution of outcomes for each condition, allowing for reliable statistical comparisons across prenatal training durations. Model parameters were optimized using the Adam optimizer (Kingma & Ba, 2014) with a fixed base learning rate of 0.0001. Cross-entropy loss was used as the learning objective.

We implemented an incremental learning regime (Van De Ven et al., 2022) to approximate infants’ gradual and repeatedly sampled exposure to linguistic input. Instead of training on the entire dataset at every epoch, the model was exposed to small subsets that were revisited throughout training. For each run, we randomly sampled 50 subsets from the entire dataset in both low-frequency and full-frequency versions, with each subset containing 1600 tokens for training and 320 tokens for testing. These subsets served as discrete experience units from which the model could draw across epochs. A learning path was then planned over the 200 training epochs. At each epoch t , one subset d_i was selected for training according to a time-dependent probability distribution. This distribution modulated the chance of revisiting the given subset across time, yielding a curriculum in which exposure was both incremental and recurrent. Let \mathcal{N} denote the set of unvisited subsets that have not yet been sampled at or before epoch t , and let \mathcal{O} denote the set of visited subsets that have already been visited. The probability of selecting the subset d_i at epoch t was defined as:

$$P(d_i|t) = \begin{cases} \frac{p_0}{|\mathcal{N}|}, & d_i \in \mathcal{N}, \\ (1 - p_0) \frac{\exp(-\lambda v_i)}{\sum_{d_j \in \mathcal{O}} \exp(-\lambda v_j)}, & d_i \in \mathcal{O}, \end{cases}$$

where $p_0 = 0.5$ controls the probability of selecting a previously unseen subset, $\lambda = 0.3$ determines the steepness of the decay over the visitation count, and v_i denotes the number of times the subset d_i has already been sampled. Intuitively, this scheme encourages early exposure to new subsets while allowing previously encountered subsets to be revisited, with more recently sampled subsets temporarily receiving lower sampling probability.

2.4 Evaluation Model performance was assessed on both the acoustically attenuated low-frequency version of the testing set and the unfiltered full-frequency version of the testing set. We employed classification accuracy as the evaluation metric by calculating the proportion of correctly predicted tokens out of the total number of tokens. The models were evaluated under two age metrics: birth age, which aligns the developmental time relative to the transition point, and term age, which measures learning from the onset of training.

For the birth age evaluation, we extracted accuracy from five evaluation windows, each spanning five epochs and defined with respect to the transition point, which we treated as the analogue of birth. The selected evaluation windows include pre-transition (15 to 11 epochs before transition, marked as -15 in the subsequent results report), immediately preceding transition (5 to 1 epoch before transition, marked as -5), immediately following transition (5 to 9 epochs after transition, marked as +5), and two later checkpoints: post-transition 1 (45 to 49 epochs after transition, marked as +45) and post-transition 2 (95 to 99 epochs after transition, marked as +95). These intervals provide a continuous picture of the models’ development around the transition and well into later learning. Accuracy within each window was averaged across the five epochs and then across the fifteen independently trained runs to yield robust estimates of the models’ performance trajectory.

For the term age evaluation, developmental time was anchored to the onset of training (epoch 0), symbolizing the point at which auditory processing is sufficiently mature for learning to occur. We sampled the models’ behavior at regular 20-epoch intervals starting from epoch 0 (e.g., 0 to 4 epochs, 20

to 24 epochs, 40 to 44 epochs, etc.), averaging performance across the five-epoch windows starting from the checkpoint and similarly across the fifteen runs. Considering that infants have access to different speech signals before and after birth, we evaluated the learning accuracy on the low-frequency testing set during the prenatal learning phase (i.e., before the transition point) and on the full-frequency testing set during the postnatal learning phase. This evaluation metric thus provides a complementary view of the learning progress that is not tied to the transition event itself, capturing how performance evolves as a function of cumulative exposure from the beginning of training.

To understand which acoustic cues the models relied on for their predictions (see Discussion), we further applied Gradient-weighted Class Activation Mapping (Grad-CAM; Selvaraju et al., 2020) to the first convolutional layers in the networks. Grad-CAM generates a class-specific saliency map by weighting the feature maps according to the gradients of the target class, producing a spatial heatmap that identifies the regions that are the most critical for the model’s prediction. Attention patterns were analyzed at two key developmental points: 10 epochs before the transition and 10 epochs after the transition. For each condition, 200 tokens were randomly sampled from the testing data, enabling the inspection of token-specific attention patterns. In addition, aggregated heatmaps were constructed by averaging across the 200 tokens in each condition, providing a more general overview of the dominant frequency regions attended to by the model in the stress and tone classification tasks.

3 Results¹

Figure 3 presents the learning trajectory of the condition with 100 epochs of low-frequency training as an instance of our models’ acquisition process. The progression of stress and tone learning exhibited a similar trend, except for the overall lower accuracy in tone classification compared to stress classification. This was expected, as there were more tone categories (four) compared to stress categories (three). Regardless of the learning task, there was a stable increase in training accuracy as a function of training epochs, confirming the models’ learning. Before the transition, the models’ performance on the low-frequency testing data was generally comparable to their performance on the training data. This proximity suggests that the models could generalize their learning to unseen data within the same frequency spectrum. In comparison, the evaluation accuracy on the full-frequency testing data before the transition appeared to be much lower, especially in the tone model. This discrepancy entails a difficulty for the models to extrapolate knowledge relevant to stress and tone classification from the low-frequency range to the full-frequency range. At the 100th epoch, we observe a drop in low testing accuracy but an increase in full testing accuracy, corresponding to the transition from low-frequency to the full-frequency range in the training data. As the models expanded their focus to the broad-spectrum sound input, they demonstrated a growing capability in full-frequency stress and tone classification. At the same time, the models naturally struggled to perform well on stress and tone classification with sound inputs restricted to the low-frequency range.

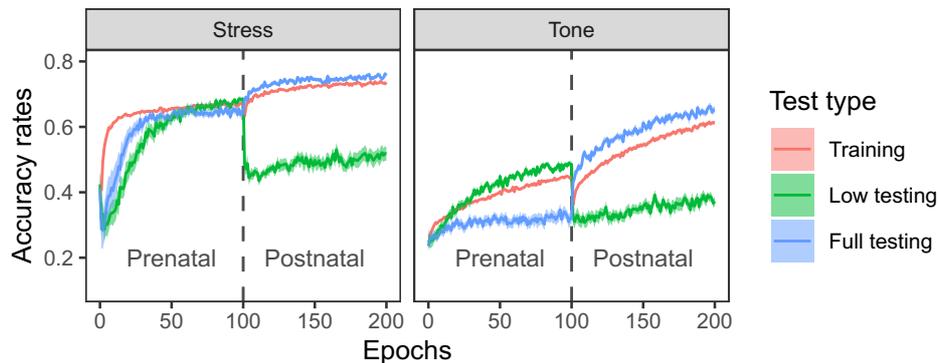


Figure 3: Change in accuracy rates of the condition with 100 epochs of prenatal training along 200 epochs of total training.

¹ All model evaluation results, the analysis script and the full regression results are available at https://osf.io/xybfb/overview?view_only=67b54045fa8646d6be3d3fa3c48b8ddf.

With the understanding of model progression in mind, we inspect how varying the duration of prenatal training impacts models' performance on the two versions of testing data under the two distinct evaluation metrics specified in the Methods section. In Figure 4, testing accuracy in the five conditions was aligned by birth age. The dashed line signifies the transition in training data, while the numbers (-15, -5, +5, +45, +95) signal the five checkpoints relative to the transition. This means that the condition with 20 epochs of prenatal training went through 5 epochs of total training at the first checkpoint (-15), while the condition with 40 epochs of prenatal training went through 25 epochs of total training at that checkpoint. For the accuracy data at each checkpoint, we ran a mixed-effects linear model using the *lme4* package (Bates et al., 2015) developed in R (R Core Team, 2022). The model included the interaction among conditions (Prenatal 20, 40, 60, 80, 100), learning tasks (Stress and Tone), and testing data (Low testing and Full testing) as the fixed effect, and varying intercepts for individual runs as the random effect. Subsequently, we conducted post hoc tests to compare the learning accuracy of various conditions within each version of testing data for each learning task, using the "mvt" method in the *emmeans* package (Lenth, 2025) to correct for multiple testing. The following paragraphs report the differences between conditions on the low testing accuracy in the course of training, followed by an examination of the results on the full testing data.

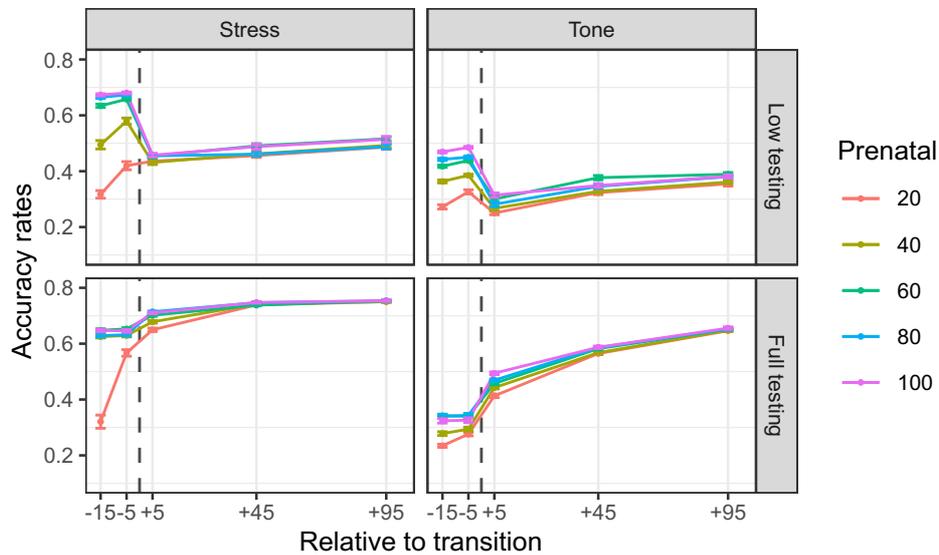


Figure 4: Change in accuracy rates of five conditions along five key checkpoints relative to the transition in training data.

At the pre-transition checkpoint (-15), longer training on low-frequency sounds resulted in higher testing accuracy on low-frequency sounds. The difference between conditions reached statistical significance in 8 out of 10 pairs of comparisons in both stress and tone acquisition (e.g., 20 compared to 40 in stress: $\beta = -0.18$, $z = -14.90$, $p < 0.001$; 20 compared to 40 in tone: $\beta = -0.09$, $z = -7.62$, $p < 0.001$). While the conditions with 60 and 80 prenatal epochs and the conditions with 80 and 100 prenatal epochs failed to show statistically significant differences on both tasks (e.g., 60 compared to 80 in stress: $\beta = -0.03$, $z = -2.54$, $p = 0.082$; 60 compared to 80 in tone: $\beta = -0.02$, $z = -2.06$, $p = 0.236$), the conditions with 60 and 100 prenatal epochs did (e.g., 60 compared to 100 in stress: $\beta = -0.04$, $z = -3.28$, $p = 0.009$; 60 compared to 100 in tone: $\beta = -0.05$, $z = -4.30$, $p < 0.001$). It is possible that a substantial difference in the duration of prenatal training was necessary to elicit a visible difference in the learning accuracy, especially when the learning was nearly at its ceiling. The superior performance on the low testing data from conditions with longer prenatal training persisted until the checkpoint immediately preceding the transition (-5), with statistical significance evident in 7 out of 10 pairs of comparisons in stress learning (e.g., 20 compared to 40: $\beta = -0.16$, $z = -16.52$, $p < 0.001$) and 9 out of 10 pairs in tone learning (e.g., 20 compared to 40: $\beta = -0.06$, $z = -6.11$, $p < 0.001$). The results revealed the benefit of prolonged exposure to the restricted frequency range for enhancing stress and tone learning within that frequency range.

Nevertheless, this benefit of extended low-frequency training for low-frequency testing gradually declined after full-frequency exposure. At the checkpoint immediately following the transition (+5), only one out of the 10 pairs of comparisons in stress learning and 6 out of 10 pairs in tone learning showed statistically significant differences (e.g., 40 compared to 100 in stress: $\beta = -0.03$, $z = -2.96$, $p = 0.026$; 20

compared to 60 in tone: $\beta = -0.05, z = -5.69, p < 0.001$). The effect of prenatal exposure became even more controversial at the two post-transition checkpoints (+45 and +95), where the condition with 60 prenatal epochs outperformed not only the conditions with 20 and 40 prenatal epochs (e.g., compared to 40 in stress at post-transition 1: $\beta = 0.03, z = 3.50, p = 0.004$; compared to 40 in tone at post-transition 1: $\beta = 0.05, z = 5.96, p < 0.001$) but also the condition with 80 prenatal epochs (e.g., in stress at post-transition 1: $\beta = 0.03, z = 3.62, p = 0.003$; in tone at post-transition 1: $\beta = 0.03, z = 3.81, p = 0.001$)². Two factors could have contributed to the unreliability of the low-frequency exposure. First, the shift in models' attention from the low-frequency spectrum to the broad spectrum could have dragged down the performance of all conditions, such that the benefit of prolonged low-frequency exposure was no longer salient. Second, full-frequency training could have facilitated the learning of the conditions with insufficient low-frequency training (see upcoming paragraphs for the power of full-frequency training).

Moving on to the full testing results, the models' performance suggests a partial extrapolation of knowledge from low-frequency acoustic cues to full-frequency stress and tone classification even before the transition. At both checkpoints preceding the transition (-15 and -5), the condition with 20 prenatal epochs yielded worse performance in stress learning compared to the other conditions (e.g., 20 compared to 40 pre-transition: $\beta = -0.30, z = -25.49, p < 0.001$; 20 compared to 40 immediately preceding transition: $\beta = -0.06, z = -6.47, p < 0.001$), although the other conditions had comparable performance (e.g., 40 compared to 60 pre-transition: $\beta = -0.02, z = -1.91, p = 0.310$; 40 compared to 60 immediately preceding transition: $\beta = -0.02, z = -2.44, p = 0.106$). Likewise, in tone learning, we found significantly lower learning accuracy from conditions with 20 and 40 prenatal epochs compared to the other conditions (e.g., 20 compared to 60 pre-transition: $\beta = -0.11, z = -9.01, p < 0.001$; 20 compared to 60 immediately preceding transition: $\beta = -0.06, z = -6.63, p < 0.001$). These results point to a defect in full-frequency stress and tone classification stemming from inadequate exposure to low-frequency sounds.

Furthermore, the assistance of prenatal training on full-frequency stress and tone learning became more prominent immediately following full-frequency exposure (+5). In particular, the conditions with 20 and 40 epochs of prenatal training both displayed inferior performance on stress classification compared to the other conditions (e.g., 20 compared to 60: $\beta = -0.05, z = -5.90, p < 0.001$). In terms of tone classification, regression results indicated significantly lower accuracy from conditions with shorter prenatal exposure in 8 out of 10 pairs of comparisons (e.g., 20 compared to 40: $\beta = -0.03, z = -3.33, p = 0.008$), with the only exceptions being the comparison between the conditions with 40 and 60 prenatal epochs and between the conditions with 60 and 80 prenatal epochs. The difference between conditions at the immediately postnatal checkpoint presents direct evidence that extended prenatal exposure to low-frequency sounds enhances postnatal proficiency in stress and tone classification using full-frequency sounds. However, as training proceeded into the two post-transition checkpoints (+45 and +95), the accuracy on full-frequency stress and tone classification of all conditions gradually converged (e.g., 20 compared to 40 in stress at post-transition 1: $\beta = 0.00, z = -0.37, p = 0.996$; 20 compared to 40 in tone at post-transition 1: $\beta = 0.00, z = -0.37, p = 0.996$). In other words, the disadvantage of inadequate low-frequency exposure weakened over the course of further full-frequency training, highlighting the effectiveness of full-frequency exposure in stress and tone learning.

As an interim summary, the evaluation by birth age reveals that exposure to low-frequency sound input assisted in not only low-frequency stress and tone classification but also full-frequency stress and tone classification. However, the benefit of prolonged low-frequency training decreased with further exposure, whether within the low-frequency range or across the full-frequency spectrum. Ultimately, a trend of convergence emerged across all conditions. To confirm the converging trend, we calculated the correlation between the duration of prenatal training and the mean accuracy rates from both versions of testing data in both learning tasks. At the pre-transition checkpoint (-15), higher accuracy rates were strongly and positively correlated with longer prenatal training ($r = 0.524, p = 0.018$), which substantiates the hypothesis that longer low-frequency training conferred a temporary advantage in stress and tone classification. However, the correlation declined over the course of further training, from $r = 0.325$ ($p = 0.162$) at the checkpoint immediately preceding the transition (-5) to $r = 0.133$ ($p = 0.575$) at the checkpoint immediately following the transition (+5), and to $r = 0.052$ ($p = 0.826$) and $r = 0.038$ ($p = 0.872$) at the post-transition checkpoints (+45 and +95). The correlation test results are in line with the convergence reported in the post hoc test results.

² The superior performance on low-frequency stress and tone classification from the condition with 60 prenatal epochs is not directly related to our research question, as the scenario where the model is trained on full-frequency sounds but tested on low-frequency sounds does not have real-life equivalence.

Figure 5 illustrates the results under the other evaluation metric specified in the Methods section, by aligning the conditions with varying durations of prenatal training from the onset of training. Our primary focus here lies on the low testing results during the prenatal learning phase of various conditions, which could reflect the artificial fetuses' stress and tone knowledge inside the womb, as well as the full testing results during the postnatal learning phase of various conditions, which could represent the infants' stress and tone learning after birth. The irrelevant datapoints are greyed out on the figure. Taking the checkpoint starting from the 40th epoch as an example, the conditions with 60, 80, and 100 prenatal epochs remain in utero during this interval, whereas the conditions with 20 and 40 prenatal epochs have already transitioned to postnatal learning; accordingly, we extracted low-frequency testing accuracy from the former and full-frequency testing accuracy from the latter. Thereafter, the low testing data and the full testing data at each checkpoint were separately fitted to linear models³, with the interaction between conditions (Prenatal 20, 40, 60, 80, 100) and learning tasks (Stress and Tone) as the fixed effect. For each learning task, we conducted pairwise comparisons between the conditions that were in the same learning phase, whether prenatal or postnatal.

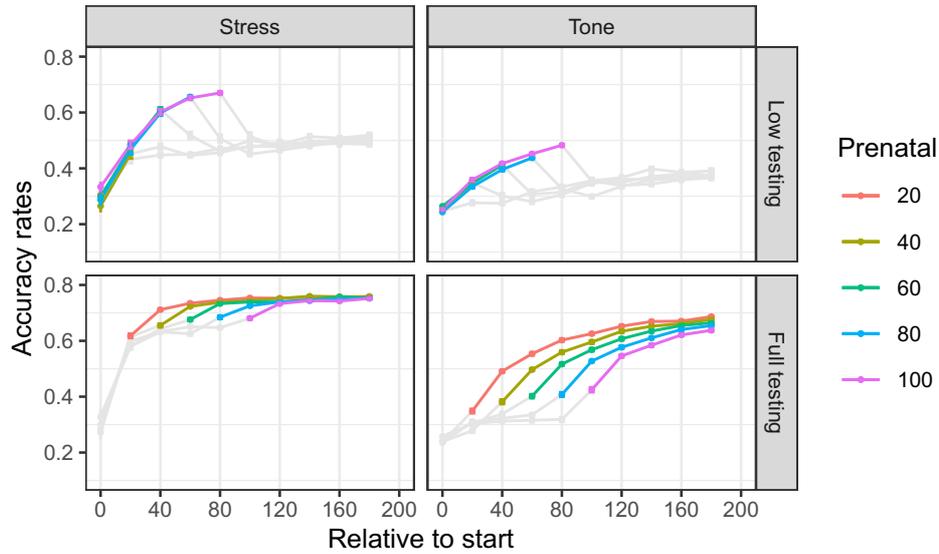


Figure 5: Change in accuracy rates of five conditions along ten key checkpoints relative to the onset of training.

Low testing accuracy at each included checkpoint generally revealed comparable performance across conditions (e.g., 40 compared to 60 in stress at 20 epochs: $\beta = -0.02$, $t = -1.27$, $p = 0.580$)⁴. This verifies an equivalent acquisition among the various conditions prior to the transition to full-frequency exposure. Full testing results for postnatal comparisons, however, contradicted our initial hypothesis regarding preterm deficit. At the checkpoint starting from the 40th epoch, the condition with 20 prenatal epochs demonstrated better development in stress and tone classification compared to the condition with 40 prenatal epochs (stress: $\beta = 0.06$, $t = 6.87$, $p < 0.001$; tone: $\beta = 0.11$, $t = 13.30$, $p < 0.001$), implying that the longer postnatal full-frequency exposure compensated for the lack of low-frequency exposure. At the checkpoint starting from the 60th epoch, both the conditions with 20 and 40 prenatal epochs showed significantly better performance on stress and tone learning compared to the condition with 60 prenatal epochs (e.g., 20 compared to 60 in stress: $\beta = 0.06$, $t = 9.06$, $p < 0.001$; 20 compared to 60 in tone: $\beta = 0.15$, $t = 23.51$, $p < 0.001$), once again corroborating the power of full-frequency training.

That said, there was no significant difference when it came to the stress acquisition between the conditions with 20 and 40 prenatal epochs ($\beta = 0.01$, $t = 1.77$, $p = 0.182$), likely because both had almost reached the ceiling of stress learning. The remaining conditions further corroborate the observation that

³ The random intercept of individual run was not included in these models due to singularity issues.

⁴ However, 2 out of 40 pairs of comparisons yielded statistically significant differences: the stress learning of the conditions with 40 and 100 prenatal epochs at the checkpoint starting from the 0th epoch and the tone learning of the conditions with 80 and 100 prenatal epochs at the checkpoint starting from the 60th epoch. Considering that all conditions went through the exact same training procedure, we hypothesize that the divergence between conditions arose from the random model initializations of the fifteen runs in each condition.

stress learning required approximately 20 epochs of full-frequency exposure to plateau, for example, as evident in the absence of difference between the conditions with 40 and 60 prenatal epochs at the checkpoint starting from the 80th epoch ($\beta = 0.00, t = 0.75, p = 0.878$). Conversely, tone learning appeared to be much slower. For instance, the condition with 40 prenatal epochs did not catch up with the condition with 20 prenatal epochs until the checkpoint at the 160th epoch ($\beta = 0.01, t = 1.81, p = 0.370$). Moreover, the conditions with shorter prenatal exposure were found to display significantly higher accuracy on tone classification compared to conditions with longer prenatal exposure in 7 out of 10 pairs of comparisons at the final checkpoint (e.g., 20 compared to 60: $\beta = 0.02, t = 4.70, p < 0.001$).

4 Discussion

The current study investigated the hypothesis that preterm infants' prosodic impairments can be attributed to insufficient low-pass filtered speech exposure in the prenatal stage. Convolutional neural networks were trained on English stress and Mandarin tone classification using low-frequency speech followed by full-frequency speech, simulating the progression from prenatal to postnatal learning of prosodic patterns. We manipulated the duration of low-frequency speech training (20, 40, 60, 80, and 100 epochs) and examined its effect on stress and tone acquisition. The modeling results can be summarized into three key takeaways. First, longer prenatal low-frequency exposure gave rise to better performance on stress and tone classification, with the effect persisting beyond the transition to full-frequency exposure. Second, the benefit of prolonged low-frequency exposure on stress and tone acquisition gradually faded away as training proceeded further into the postnatal stage. Third, longer prenatal exposure in combination with shorter postnatal exposure resulted in inferior stress and tone learning compared to shorter prenatal exposure combined with longer postnatal exposure.

Our results were partially in line with previous behavioral results on the prosodic development of preterm infants (Bosch, 2011; Gonzalez-Gomez et al., 2021; Herold et al., 2008; Peña et al., 2010; Ragó et al., 2014). For instance, Peña and colleagues (2010) found that preterm infants at 6 months of birth age (equivalent to 3 months of term age) were unable to distinguish their native language from a rhythmically similar language, a skill already present in full-term infants at the same birth age. Correspondingly, in our model simulation, the conditions with shorter prenatal exposure generally demonstrated worse accuracy in full-frequency stress and tone classification immediately following the transition to full-frequency exposure. Under a different age measure, preterm infants at 6 months of term age (equivalent to 9 months of birth age) in Herold et al. (2008) showed a similar deficit in discriminating trochaic and iambic stress patterns compared to matched full-term infants, whereas preterm and full-term infants at 6 months of term age in Peña et al. (2010) did not display any substantially different neural response to the rhythmic pattern in their native speech. Adding to this complication, the current results diverged from both behavioral studies. We found that conditions with shorter prenatal exposure and longer postnatal exposure performed even better on stress and tone classification compared to the conditions with longer prenatal exposure. Our results seem to suggest that low-frequency exposure is valuable for prosodic learning, while full-frequency exposure is even more beneficial.

The primary contribution of this work lies in disentangling the importance of sufficient low-frequency exposure from the influence of the non-linguistic components in shaping preterm infants' prosodic development. Contrary to Peña and colleagues (2010), who ascribed the prosodic impairments of preterm infants to neural maturation, or Gonzalez-Gomez and colleagues (2021), who demonstrated the influence of socioeconomic status on the prosodic development of preterm infants, the current study reveals that even a computationally powerful neural network model faces significant challenges in acquiring prosodic features, such as stress and tone, in the postnatal stage, when it lacks sufficient low-frequency exposure to establish foundational regularities. Furthermore, the model's progression trajectory suggests the potential source of the difficulty in postnatal prosodic learning. At the two checkpoints prior to the transition, we already observed lower accuracy on the low testing data from conditions with shorter low-frequency exposure. As the model extrapolated the stress and tone knowledge acquired from the low-frequency range to the full-frequency range, the conditions with insufficient low-frequency knowledge naturally exhibited poorer performance on the full testing data.

The modeling results did not replicate the behavioral results in terms of the effectiveness of full-frequency training, precisely because we restricted our attention to how the amount and the characteristics of the input impact prosodic learning. First, our models were exclusively instructed to learn the categorization of either stress or tone, while human infants are inevitably in a multitasking environment and need to balance prosodic learning with other facets of speech acquisition, such as the

acquisition of phonemes and phonotactics. This consideration resonates with a discussion point in Herold et al. (2008), suggesting that preterm and full-term infants may have distinct attention assignments to the various tasks in language acquisition. Second, the conditions representing more “preterm” infants and the conditions representing more “full-term” infants in our model had a similar level of computational power, whereas premature human infants may exhibit biological deficiencies or neural immaturities. With that said, our study revealed that full-frequency exposure can compensate for the lack of low-frequency exposure in single-tasking prosodic learning with ample computational power.

Our design considered two distinct aspects of the linguistic input: the total amount of the input data, regulated by the duration of training, and the characteristics of the input data, namely, low-pass filtered or broad-spectrum. Despite an initial advantage of low-frequency exposure under the birth age evaluation, the final testing accuracy across conditions converged by the end of extensive training. In a similar vein, while full-frequency exposure yielded an initial advantage under the term age evaluation, the conditions with shorter postnatal exposure caught up with the conditions with longer postnatal exposure within 20 epochs of stress learning, albeit no less than 100 epochs were required for tone learning in our setup. The weakening of any low-frequency or full-frequency benefit along the course of training implies that the total quantity of speech exposure, rather than the quality of speech exposure, is the key to prosodic acquisition. This idea coincides with the line of research advocating for increasing the amount of speech input for preterm infants, for instance, by placing them in open care units instead of private care units (Pineda et al., 2014). Nevertheless, our results undermine the potential contribution of the starting-small learning effect (Elman, 1993; Newport, 1990; Turkewitz & Kenny, 1982) in infants’ prosodic learning.

In fact, our model surprisingly exhibited better prosodic learning with the full-frequency range of sounds, which encompasses more segmental information, when the primary acoustic cue to stress and tone identity predominantly exists within the low-frequency range. We further inspected the Grad-CAM heatmaps (Selvaraju et al., 2020) to pinpoint the frequency regions that were important to the models’ predictions. These attention heatmaps, as shown in Figure 6, indicate that the models’ learning of stress and tone was more comprehensive than what we expected. Although the models’ attention was restricted to frequencies below 500 Hz before the transition in training data, stress learning incorporated more cues from a larger area, including F0, intensity, and spectral patterns related to phonemes immediately following the transition. Likewise, the tone learning model gradually switched its attention to more spectral components besides F0, such as harmonics and vowel formants, after the transition. Contrary to the customary assumption, the acoustic cues beyond the low-frequency range can be informative for stress and tone categorization.

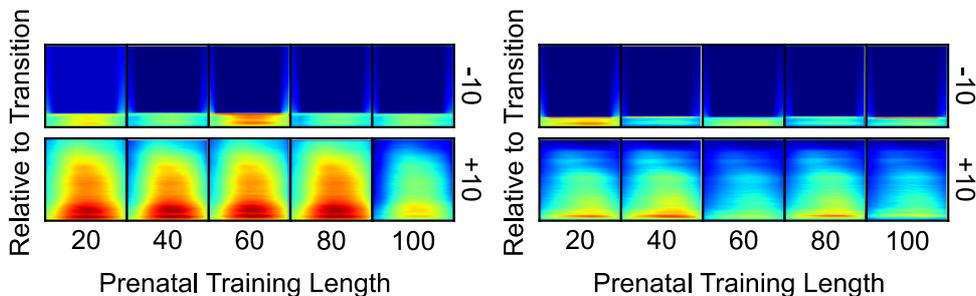


Figure 6: (Left) Accumulated heatmaps of 200 stress tokens. (Right) Accumulated heatmaps of 200 tone tokens. The before transition (-10) heatmaps represent attention to the low-frequency audio, while the after transition (+10) heatmaps represent attention to full-frequency audio.

To conclude, the current study provides evidence for the critical role of prenatal speech exposure to prosodic learning by showing a short-term enhancement in stress and tone classification with increased duration of low-frequency training, which offers a potential linguistic explanation for the prosodic impairments of preterm infants in their first year of life. In addition, our results challenge the conventional belief in the literature that prosodic acquisition primarily relies on acoustic cues in the low-frequency range of speech, such as those encountered in the in-utero environment. We demonstrated the potential advantage of exposure to real-world broadcast speech, which encompasses not only relevant low-frequency cues but also segmental cues, to improve stress and tone classification. Various conditions in the model simulation achieved successful prosodic learning with a sufficient amount of language input, regardless of the frequency range of that language input. These findings underscore the significance of considering both the quantity and the quality of speech sounds in infants’ early prosody development.

References

- Back, Stephen A., & Miller, Steven P. (2014). Brain injury in premature neonates: A primary cerebral dysmaturation disorder? *Annals of Neurology*, *75*(4), 469–486. <https://doi.org/10.1002/ana.24132>
- Bates, Douglas, Mächler, Martin, Bolker, Ben, & Walker, Steve. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.
- Bosch, Laura. (2011). Precursors to language in preterm infants: Speech perception abilities in the first year of life. *Progress in Brain Research*, *189*, 239–257.
- Christophe, Anne, Dupoux, Emmanuel, Bertoncini, Josiane, & Mehler, Jacques. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *The Journal of the Acoustical Society of America*, *95*(3), 1570–1580.
- Christophe, Anne, Mehler, Jacques, & Sebastián-Gallés, Núria. (2001). Perception of Prosodic Boundary Correlates by Newborn Infants. *Infancy: The Official Journal of the International Society on Infant Studies*, *2*(3), 385–394. https://doi.org/10.1207/S15327078IN0203_6
- Coviello, Caterina, Keunen, Kristin, Kersbergen, Karina J., Groenendaal, Floris, Leemans, Alexander, Peels, Barbara, Isgum, Ivana, Viergever, Max A., de Vries, Linda S., Buonocore, Giuseppe, Carnielli, Virgilio P., & Benders, Manon J. N. L. (2018). Effects of early nutrition and growth on brain volumes, white matter microstructure, and neurodevelopmental outcome in preterm newborns. *Pediatric Research*, *83*(1), 102–110. <https://doi.org/10.1038/pr.2017.227>
- Eggermont, Jos J., & Moore, Jean K. (2012). Morphological and Functional Development of the Auditory Nervous System. In Lynne Werner, Richard R. Fay, & Arthur N. Popper (Eds.), *Human Auditory Development* (Vol. 42, pp. 61–105). Springer New York. https://doi.org/10.1007/978-1-4614-1421-6_3
- Elman, Jeffrey L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*(1), 71–99.
- Fló, Ana, Brusini, Perrine, Macagno, Francesco, Nespor, Marina, Mehler, Jacques, & Ferry, Alissa L. (2019). Newborns are sensitive to multiple cues for word segmentation in continuous speech. *Developmental Science*, *22*(4), e12802. <https://doi.org/10.1111/desc.12802>
- Gélat, Pierre, David, Anna L., Haqhenas, Seyyed Reza, Henriques, Julian, Thibaut de Maisieres, Aude, White, Tony, & Jauniaux, Eric. (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. <https://doi.org/10.1016/j.ajog.2019.05.036>
- Gervain, Judit. (2018). The role of prenatal experience in language development. *Current Opinion in Behavioral Sciences*, *21*, 62–67. <https://doi.org/10.1016/j.cobeha.2018.02.004>
- Gonzalez-Gomez, Nayeli, O'Brien, Frances, & Harris, Margaret. (2021). The effects of prematurity and socioeconomic deprivation on early speech perception: A story of two different delays. *Developmental Science*, *24*(2), e13020. <https://doi.org/10.1111/desc.13020>
- Graven, Stanley N., & Browne, Joy V. (2008). Auditory Development in the Fetus and Infant. *Newborn and Infant Nursing Reviews*, *8*(4), 187–193. <https://doi.org/10.1053/j.nainr.2008.10.010>
- Harris, Charles R., Millman, K. Jarrod, Van Der Walt, Stéfan J., Gommers, Ralf, Virtanen, Pauli, Cournapeau, David, Wieser, Eric, Taylor, Julian, Berg, Sebastian, Smith, Nathaniel J., Kern, Robert, Picus, Matti, Hoyer, Stephan, Van Kerkwijk, Marten H., Brett, Matthew, Haldane, Allan, Del Río, Jaime Fernández, Wiebe, Mark, Peterson, Pearu, ... Oliphant, Travis E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, & Sun, Jian. (2015). *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.1502.01852>
- Hepper, Peter G, & Shahidullah, B Sara. (1994). Development of fetal hearing. *Archives of Disease in Childhood Fetal and Neonatal Edition*, *71*(2), F81–F87. <https://doi.org/10.1136/fn.71.2.f81>
- Herold, Birgit, Höhle, Barbara, Walch, Elisabeth, Weber, Tanja, & Obladen, Michael. (2008). Impaired word stress pattern discrimination in very-low-birthweight infants during the first 6 months of life. *Developmental Medicine & Child Neurology*, *50*(9), 678–683. <https://doi.org/10.1111/j.1469-8749.2008.03055.x>
- Kingma, Diederik P., & Ba, Jimmy. (2014). *Adam: A Method for Stochastic Optimization*. <https://doi.org/10.48550/ARXIV.1412.6980>
- Lenth, Russell V. (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. <https://rvlenth.github.io/emmeans/>
- Lugosch, Loren, Ravanelli, Mirco, Ignoto, Patrick, Tomar, Vikrant Singh, & Bengio, Yoshua. (2019). *Speech Model Pre-training for End-to-End Spoken Language Understanding*. <https://doi.org/10.48550/ARXIV.1904.03670>
- Mampe, Birgit, Friederici, Angela D., Christophe, Anne, & Wermke, Kathleen. (2009). Newborns' Cry Melody Is Shaped by Their Native Language. *Current Biology*, *19*(23), 1994–1997. <https://doi.org/10.1016/j.cub.2009.09.064>
- Martinez-Alvarez, Anna, Benavides-Varela, Silvia, Lapillonne, Alexandre, & Gervain, Judit. (2023). Newborns discriminate utterance-level prosodic contours. *Developmental Science*, *26*(2), e13304. <https://doi.org/10.1111/desc.13304>
- McAuliffe, Michael, Socolof, Michaela, Mihuc, Sarah, Wagner, Michael, & Sonderegger, Morgan. (2017). *Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi*. 498–502. <https://doi.org/10.21437/Interspeech.2017-1386>

- McFee, Brian, Raffel, Colin, Liang, Dawen, Ellis, Daniel, McVicar, Matt, Battenberg, Eric, & Nieto, Oriol. (2015). *librosa: Audio and Music Signal Analysis in Python*. 18–24. <https://doi.org/10.25080/Majora-7b98e3ed-003>
- Mehler, Jacques, Jusczyk, Peter, Lambertz, Ghislaine, Halsted, Nilofar, Bertoncini, Josiane, & Amiel-Tison, Claudine. (1988). A precursor of language acquisition in young infants. *Cognition*, 29(2), 143–178. [https://doi.org/10.1016/0010-0277\(88\)90035-2](https://doi.org/10.1016/0010-0277(88)90035-2)
- Monson, Brian B., Ambrose, Sophie E., Gaede, Carey, & Rollo, Derrick. (2023). Language Exposure for Preterm Infants is Reduced Relative to Fetuses. *The Journal of Pediatrics*, 262. <https://doi.org/10.1016/j.jpeds.2022.12.042>
- Moon, Christine, Cooper, Robin Panneton, & Fifer, William P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, 16(4), 495–500. [https://doi.org/10.1016/0163-6383\(93\)80007-U](https://doi.org/10.1016/0163-6383(93)80007-U)
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 756–766. <https://doi.org/10.1037//0096-1523.24.3.756>
- Nazzi, Thierry, Floccia, Caroline, & Bertoncini, Josiane. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development*, 21(4), 779–784. [https://doi.org/10.1016/S0163-6383\(98\)90044-3](https://doi.org/10.1016/S0163-6383(98)90044-3)
- Newport, Elissa L. (1990). Maturation constraints on language learning. *Cognitive Science*, 14(1), 11–28.
- Panayotov, Vassil, Chen, Guoguo, Povey, Daniel, & Khudanpur, Sanjeev. (2015). Librispeech: An ASR corpus based on public domain audio books. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5206–5210. <https://doi.org/10.1109/ICASSP.2015.7178964>
- Peña, Marcela, Pittaluga, Enrica, & Mehler, Jacques. (2010). Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences*, 107(8), 3823–3828. <https://doi.org/10.1073/pnas.0914326107>
- Pineda, Roberta G., Neil, Jeff, Dierker, Donna, Smyser, Christopher D., Wallendorf, Michael, Kidokoro, Hiroyuki, Reynolds, Lauren C., Walker, Stephanie, Rogers, Cynthia, & Mathur, Amit M. (2014). Alterations in brain structure and neurodevelopmental outcome in preterm infants hospitalized in different neonatal intensive care unit environments. *The Journal of Pediatrics*, 164(1), 52–60.
- Querleu, Denis, Renard, Xavier, Versyp, Fabienne, Paris-Delrue, Laurence, & Crèpin, Gilles. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 28(3), 191–212. [https://doi.org/10.1016/0028-2243\(88\)90030-5](https://doi.org/10.1016/0028-2243(88)90030-5)
- R Core Team. (2022). *R* [Computer software].
- Ragó, Anett, Honbolygó, Ferenc, Róna, Zsófia, Beke, Anna, & Csépe, Valéria. (2014). Effect of maturation on suprasegmental speech processing in full- and preterm infants: A mismatch negativity study. *Research in Developmental Disabilities*, 35(1), 192–202.
- Raviv, Tali, Kessenich, Maureen, & Morrison, Frederick J. (2004). A mediational model of the association between socioeconomic status and three-year-old language abilities: The role of parenting factors. *Early Childhood Research Quarterly*, 19(4), 528–547. <https://doi.org/10.1016/j.ecresq.2004.10.007>
- Sansavini, Alessandra, Bertoncini, Josiane, & Giovanelli, Giuliana. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, 33(1), 3.
- Selvaraju, Ramprasaath R., Cogswell, Michael, Das, Abhishek, Vedantam, Ramakrishna, Parikh, Devi, & Batra, Dhruv. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, 128(2), 336–359. <https://doi.org/10.1007/s11263-019-01228-7>
- Shi, Yao, Bu, Hui, Xu, Xin, Zhang, Shaoji, & Li, Ming. (2021). *AISHELL-3: A Multi-speaker Mandarin TTS Corpus and the Baselines* (No. arXiv:2010.11567). arXiv. <http://arxiv.org/abs/2010.11567>
- Smith, Lucy K., Draper, Elizabeth D., Manktelow, Bradley N., Dorling, Jon S., & Field, David John. (2007). Socioeconomic inequalities in very preterm birth rates. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 92(1), F11–F14.
- Turkewitz, Gerald, & Kenny, Patricia A. (1982). Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. *Developmental Psychobiology*, 15(4), 357–368. <https://doi.org/10.1002/dev.420150408>
- Van De Ven, Gido M., Tuytelaars, Tinne, & Tolias, Andreas S. (2022). Three types of incremental learning. *Nature Machine Intelligence*, 4(12), 1185–1197. <https://doi.org/10.1038/s42256-022-00568-3>
- Vogelsang, Marin, Vogelsang, Lukas, Diamond, Sidney, & Sinha, Pawan. (2023). Prenatal auditory experience and its sequelae. *Developmental Science*, 26(1), e13278. <https://doi.org/10.1111/desc.13278>
- Volpe, Joseph J. (2009). Brain injury in premature infants: A complex amalgam of destructive and developmental disturbances. *The Lancet Neurology*, 8(1), 110–124. [https://doi.org/10.1016/S1474-4422\(08\)70294-1](https://doi.org/10.1016/S1474-4422(08)70294-1)