**Analytic Approach**

**Aim 1: Characterize the longitudinal growth trajectories of early behavioral motor indicators in relation to cortical activation patterns in 2–6-month-olds.**

To characterize the longitudinal growth trajectories of early behavioral motor indicators in relation to cortical activation patterns in 2–6-month-olds, we will use correlational analyses to examine associations between these two domains. Specifically, canonical correlation analysis (CCA) will be used to identify and measure the multivariate relationships between repeated behavioral measures and simultaneously acquired cortical activation data. CCA will allow us to find the best linear combinations across time between the two datasets.

To further explore and validate these associations, we will use partial correlation analysis to examine specific pairwise relationships between behavioral and cortical variables while controlling for potential confounding factors. This step will help isolate direct associations and provide greater insight into the developmental links between behavior and neural activity.

**Power analysis:**

A power analysis was conducted to determine the optimal sample size required for a 0.8 statistical power.

To select the effect size, we conducted a systematic literature review. We chose a study from Heller Murray et al., 2021 (Non-nutritive suck and voice onset time: Examining infant oromotor coordination) to base the inputs for the power calculation. The paper was chosen because this study included non-nutritive sucking as a variable and also used vocalization rate as a response variable, which are very similar to our experiment design. The study employed a regression model with an R² value that we can use to estimate the effect size to calculate the required sample size. R² represents the proportion of variance in the outcome variable that is explained by the predictor variables included in the model.

The effect size index chosen will be Cohen’s f, a typical effect size measurement method for samples with only one group, and it measures the strength of the relationship between a predictor variable and a dependent variable in a regression model. The formula is defined as follows:

Substituting , cohen’s f can be estimated as

The significance level (α) will be set at 0.05, following conventional standards across many fields. The desired level of statistical power will be set at 0.80, indicating an 80% probability of detecting a true effect if it exists.

G\*power was used to calculate the required sample size based on the estimated effect size,  and 0.8 power. Assuming there are 6 predictors on the linear mixed effect model, the sample size required is 24.

\*\* It is important to note that this is a rough approximation and should not be interpreted as a definitive sample size requirement. Another accurate power analysis for linear mixed effects models can be performed using simulation-based methods. However, both these approaches require specifying the number and structure of predictors included in the model—details which are not yet fully determined in this study. For the purposes of this preliminary estimate, we have assumed six predictors.

A graph of a person with a blue line

AI-generated content may be incorrect.

The above graph shows the relationship between power and sample size when there are 6 predictors in the model.

**Aim 2 Determine the predictive value of infants’ 2-6 month behavioral and neural motor indicators on characteristics of 12-month infant vocalizations.** *Longitudinal analysis of NNS and speech/language development*

To examine how early motor indicators predict characteristics on infant vocalizations at 12 months, we will use linear mixed-effects modeling (LMM). This approach is appropriate for handling longitudinal or repeated measures data, allowing for both fixed effects (e.g., specific early motos indicators/NNS variables/Age/group status) and random effects (individual variability across infants). The interaction between NNS and group status will be included in the models to assess whether associations between NNS and speech/language outcomes differ by group.

The primary outcome variable will be characteristics of infant vocalizations at 12 months, such as vocalization rates or consonant inventories. The main predictors will be characteristics that will be chosen from Aim 1 measured between 2-6 months of age. These may include suck burst duration, suck frequency, average cycles per burst etc., simultaneously collected fNIRs data, age and group status. These indicators can be included as continuous measures or categorical variables, depending on how they are coded.

We will fit a linear mixed-effects model of the form:

where is the vocalization characteristic of infant at time . are the fixed effects for motor indicators and covariates, is the random intercept for infant capturing individual variability, and is the residual error term. When repeated measurements are available for predictors or outcome, time can also be included as a fixed and/or random effect.

Relevant covariates will be included in the model to control for potential confounding.