Classification was conducted with a linear support vector machine – [information about the margin here] – in two-way classifications (e.g. speech vs. sign). Data were z-scored by channel with missing values for unreliable channels excluded in the calculation of the mean and standard deviation. The patterns submitted to the analysis were the averaged epoched data subtracted from the baseline for each participant and each condition, such that each participant contributed a single pattern. We used a leave one participant approach, such that the classifier was trained on a balanced training set of neural responses from all of the participants excluding the to-be-classified participant. The training pattern was the averaged pattern from the included participants. The mean was calculated excluding absent channels. This was tested against the held out pattern from each of the two conditions derived from the averaged fNIRS epoch data. Where data was unavailable for a participant/channel in the test pattern – the associated channels were dropped out of the analysis. This meant that classification was achieved using different numbers of channels depending on whether the channel contributed reliable information. [NOTE HERE HOW DOES THE SYSTEM DECIDE WHICH VALUES ARE NaNs PRIOR TO THE TOOLBOX?]. Each fold, e.g. participant, contributed two trials, one from each condition to-be-classified and could be classified as correctly guessed or otherwise by the algorithm. Hence in a data set of 60 participants yielded an accuracy based on successful classifier guesses from 120 trials, and within group analyses of 20 participants derived 40 classification trials on which accuracy was based. Often channels were missing for the babies due to x reason. There is a list of used channels in supplementary materials X used for each baby.

Classification reported is the proportion of correctly guessed trials. To ensure that classification was not biased, permutation testing was conducted. Permutation was conducted by randomly permuting condition labels for each participant, such that the labels were either maintained or swopped randomly for each participant, prior to training and testing the classifier. This ensured that the participant structure was preserved in the label shuffling and that the full set of patterns were permuted, e.g. those both used to train and test. For within group binary classifications, X number of permutations was conducted and a probability value was ascertained by generating a null distribution and identifying the number of observed values that was equal to or exceeded the accuracy derived from the non-shuffled data (Pereira et al., 2009). The observed value was included in the dominator for calculating the p value (Ruxton et al. 2013).

Because different channels contributed to classification of the test pattern for different babies, we calculated the number of channels used in common by X % number of babies. We then trained an SVM using these channels on the average data across all babies and extracted the classifier weight vector. Classifier weights were then multiplied by the average pattern to understand how classification was achieved. A + weight after multiplication was indicative of condition 1 and a – weight for condition 2. We then took the X % largest weights for visualisation as these weights contributed most to classification.

To test lateralisation classifications the observed difference between classifications attained from the left and right hemisphere were subtracted from one another – and the absolute value taken - as the observed difference in lateralisation. Values for the null (permutated) distribution were attained by subtracting classifications from the left and right hemisphere from the permuted distributions in the within hemisphere analyses and then taking the absolute value. The probability of attaining the observed difference value by chance, was estimated by counting the number of difference values equal to or larger than the observed value (Pereira et al., 2009) in the permuted difference distribution. The observed value was included in the denominator (Ruxton et al. 2013).