Supplementary Information : Real-world clinical validation of brainstem-based ocular biomarkers for ADHD classification in children and adults.

Chamith Ranathunga[1], August Romeo[2], Elizabeth Kilbey[3], and Hans Supèr[1,2,4,*]

[1]Braingaze Ltd, London, UK
[2]Dept of Cognition, Development and Educational Psychology, University of Barcelona, Spain
[3]The Gatehouse, Beckenham, UK
[4]Catalan Institution for Research and Advanced Studies, Barcelona, Spain
[*]*Correspondence:* hans.super@icrea.cat

This document contains supplementary methods, results, figures, and tables to support the main manuscript.

Contents

List of Supplementary Figures

S1 Preliminary analysis of Synchrony distribution

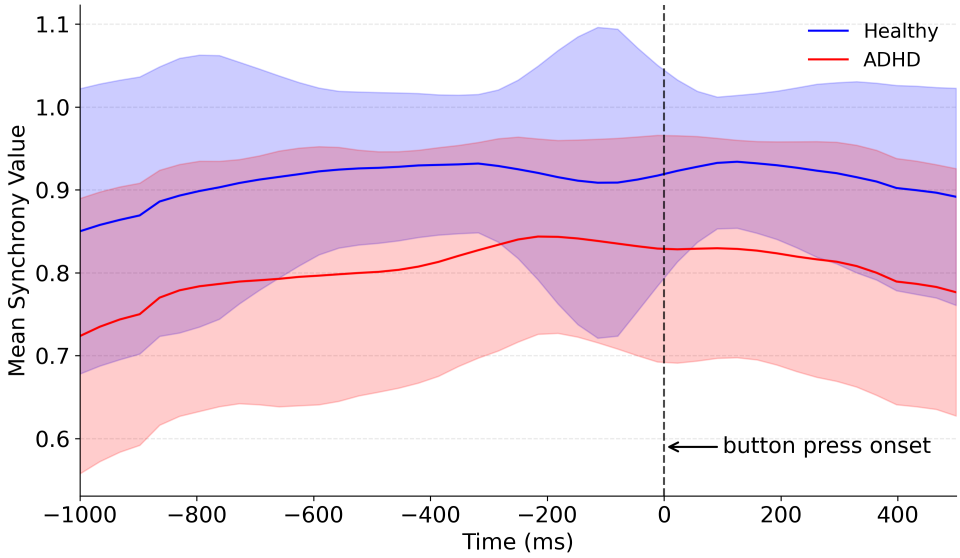

Figure S1: Population-level distributions of temporal synchrony features showing mean and standard deviation. The mean synchrony values are plotted for the Healthy control group (blue line) and the ADHD patient group (red line). Shaded regions denote ± 1 standard deviation for each respective group. The vertical dashed line at 0 ms indicates the button press onset.

S2 Distribution of individual model scores in Logit space

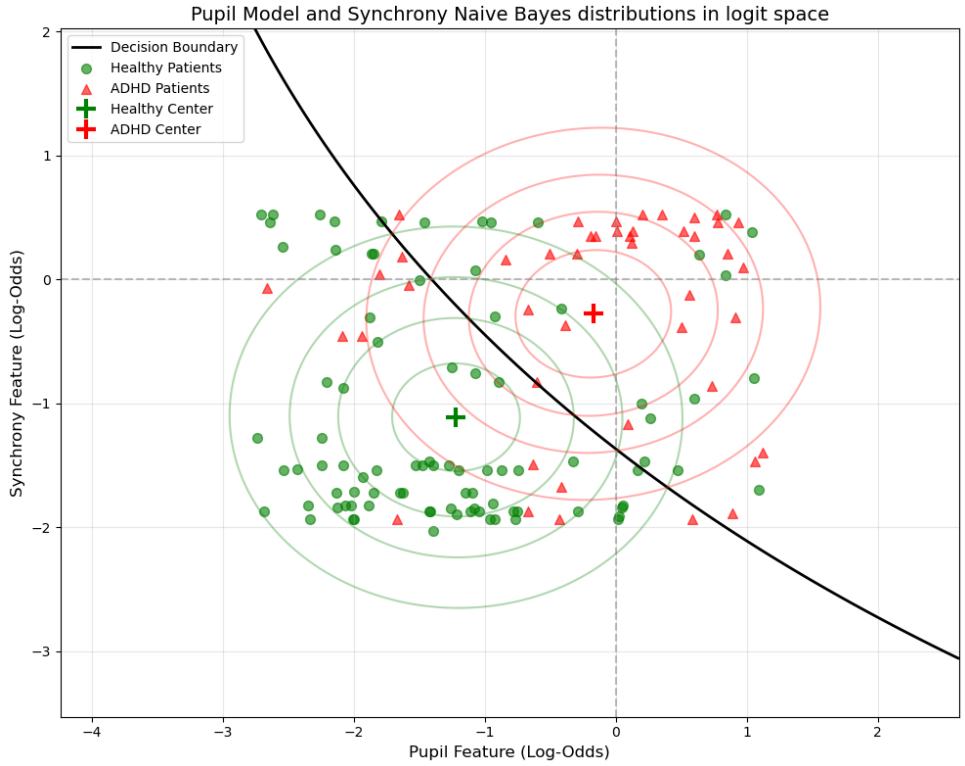

Figure S2: Joint distribution of logit-transformed pupil and synchrony scores. The scatter plot displays the distribution of diagnostic scores mapped into a joint logit space for the pupil model (x-axis) and the Naïve Bayes synchrony model (y-axis). Green circles represent Healthy patients, while red triangles represent ADHD patients. The prominent green and red crosses indicate the estimated centers (means, μ_c) for the Healthy and ADHD distributions, respectively. Concentric ellipses illustrate the covariance matrices (Σ_c) and probability densities for each group. The solid black line represents the linear decision boundary separating the two classifications.

The centres of the distributions $\boldsymbol{\mu}_c$ and covariance matrices $\boldsymbol{\Sigma}_c$ were estimated exclusively from the training data. The fitted Gaussian parameters reveal a separation between groups in the joint logit space. The healthy group is characterised by a mean of $\boldsymbol{\mu}_0 = (-1.22, -1.11)^\top$, indicating consistently lower logit-transformed scores for both the pupil and synchrony markers. In contrast, the ADHD group has a mean of $\boldsymbol{\mu}_1 = (-0.17, -0.28)^\top$, shifted markedly towards zero, reflecting higher and less suppressed marker responses.

Both covariance matrices are near-diagonal, indicating that the two features are largely uncorrelated within each group. For healthy participants, $\boldsymbol{\Sigma}_0 = \begin{bmatrix} 0.991 & -0.009 \\ -0.009 & 0.785 \end{bmatrix}$, and for ADHD participants, $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.955 & 0.030 \\ 0.030 & 0.717 \end{bmatrix}$. The variances are comparable across groups and dimensions, implying that the Gaussian classifier operates primarily through a shift in means rather than through differences in spread or feature correlation.

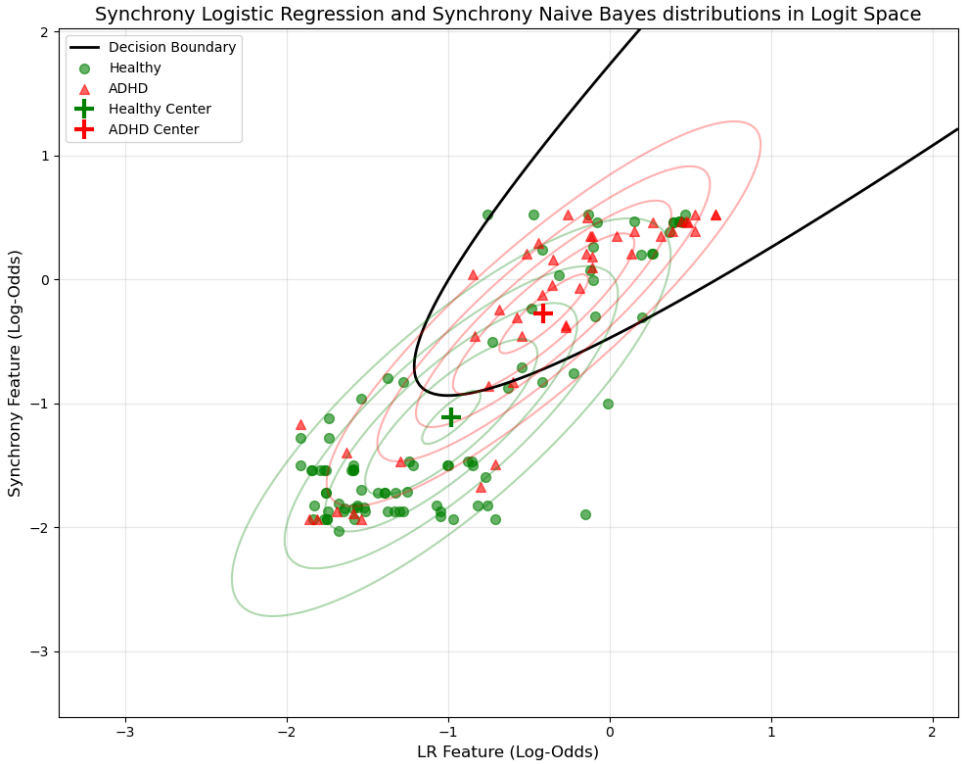

Figure S3: Joint distribution of logit-transformed scores of two synchrony models. The scatter plot displays the distribution of diagnostic scores mapped into a joint logit space for the pupil model (x-axis) and the Naïve Bayes synchrony model (y-axis). Green circles represent Healthy patients, while red triangles represent ADHD patients. The prominent green and red crosses indicate the estimated centers (means, μ_c) for the Healthy and ADHD distributions, respectively. Concentric ellipses illustrate the covariance matrices (Σ_c) and probability densities for each group. The solid black line represents the linear decision boundary separating the two classifications.

The centres of the distributions $\boldsymbol{\mu}_c$ and covariance matrices $\boldsymbol{\Sigma}_c$ for the impulsivity marker were estimated from the training data, mapping the outputs of the Logistic Regression (LR) and Naïve Bayes (NB) synchrony models into a joint logit space. The fitted Gaussian parameters demonstrate a distinct shift between the diagnostic groups. The healthy group is characterised by a mean of $\boldsymbol{\mu}_0 = (-0.98, -1.11)^\top$, indicating lower logit-transformed scores across both synchrony models. In contrast, the ADHD group exhibits a mean of $\boldsymbol{\mu}_1 = (-0.41, -0.28)^\top$, showing a clear shift towards higher log-odds responses.

The covariance matrices for the healthy and ADHD participants are estimated as follows: $\boldsymbol{\Sigma}_0 = \begin{bmatrix} 0.56 & 0.54 \\ 0.54 & 0.79 \end{bmatrix}$ and $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.54 & 0.55 \\ 0.55 & 0.72 \end{bmatrix}$. Unlike the pupil-synchrony fusion (Figure S2), both covariance matrices exhibit substantial off-diagonal elements. This indicates a strong positive correlation between the LR and NB features within both groups, which is visually reflected in the diagonally elongated elliptical contours (Figure S3). This high correlation is expected, as both conventional classifiers model the same underlying ensemble-averaged synchrony trajectories derived from commission errors. Despite this correlation, the variances and covariances remain highly consistent between the two groups. This structural similarity indicates that the classification boundary primarily leverages the distinct shift in the distribution means $\boldsymbol{\mu}_c$ along this correlated axis to discriminate between ADHD and healthy patients.

S3 Derivation of Decision Thresholds via Chow's Rule

To manage diagnostic uncertainty and minimize the expected risk of misclassification in a clinical setting, we employed a selective classification policy based on Chow's Rule. In this framework, the model can either assign a sample to a diagnostic class or defer the decision (placing it in the "uncertain" zone) if the

confidence is insufficient.

Let x be an observation and $P(c|x)$ be the posterior probability of class $c \in \{0,1\}$, where $c = 0$ represents the Healthy class and $c = 1$ represents the ADHD class. We define the costs associated with the possible decisions as follows: the cost of a correct classification is 0, the cost of misclassification is C_e, and the cost of rejection (deferral) is C_r.

According to Bayesian decision theory, a decision should be deferred if the expected risk of rejection is strictly less than the expected risk of classification. Therefore, the condition for rejection is:

$$C_r < C_e \cdot (1 - \max(P(c = 0|x), P(c = 1|x))) \tag{S1}$$

By rearranging this inequality, we define the threshold t for the maximum posterior probability below which a rejection occurs:

$$t = 1 - \frac{C_r}{C_e} \tag{S2}$$

In our clinical context, a misclassification (false positive or false negative) carries a significantly higher burden than a deferral. We conservatively assigned the cost of misclassification to be three times the cost of rejection ($C_e = 3C_r$). Substituting this ratio yields a fixed operating threshold of $t \approx 0.66$.

For a binary classification, the condition for rejection, $\max(p, 1-p) < t$, requires both $p < t$ and $p > 1 - t$. Substituting our calculated threshold $t = 0.66$ defines the upper and lower boundaries of our uncertain zone:

$$0.33 < P(c = 1|x) < 0.66 \tag{S3}$$

This mathematically grounds our fixed operating points of $t_{low} = 0.33$ and $t_{high} = 0.66$.