# 1 Introduction

Transcranial magnetic stimulation (TMS) is a common, non-invasive experimental technique used to evoke action potentials in cortical regions of the brain. In particular, researchers often target the motor cortex and measure the motor evoked potential (MEP) aroused by the stimulation. When the purpose of the experiment is to inquire on neuroplasticity, such stimulations are performed under a *paired-pulse paradigm.*

The *paired-pulse paradigm* (or *double pulse paradigm*) consists in eliciting a series of two temporally proximate pulses (in the order of miliseconds). The evoked potentials of the double-stimulations are compared to those of single test stimulations, and their relative amplitude is taken as a proxy of neuroplasticity in the brain. The time separating each of the paired-pulses is termed *interstimulous interval* (ISI). It is the general case that low intervals (4 or 5 miliseconds) produce intracortical inhibition, with the evoked potentials of paired stimulations being generally lower than those of single pulses. Greater intervals, on the other hand, tend to produce facilitation.

In the context of this paper, we shall term any coefficient that serves to represent the proportional relationship between the potentials evoked by paired and test pulses a *measure of relative amplitude.* Measures of relative amplitude in neuroscience are generally computed at the subject level. This is reasonable, since hypotheses generally deal with differences across subject groups. However, TMS experimental results are pulse-specific evoked potentials, and transforming them to subject- or group-specific measures implies down-scaling the data resolution.

The goal of this paper is to provide pulse-specific measures of relative amplitude. This is, coefficients that represent the relative amplitude of each individual paired-pulse with respect to the set of test pulses in an experimental session. The purpose of this endeavor is to keep data resolution at its highest, which on its turn allows for the implementation of data-driven artificial intelligence in the analysis of the experimental results. Thus, from a computational perspective, our objective is constrained to the sphere of feature engineering.

We shall show how pulse-specific measures of relative amplitude allow for otherwise unfeasible computational analyses of TMS data, such as the use of machine learning models for the detection of different pulse-response patterns among different groups of clinical subjects. In particular, we will show they allow for a machine learning classifier to correctly determine whether a subject belongs to one of four clinical categories based only on its evoked potentials across different inter-stimulous intervals with an accuracy of up to 90%.

# 2 Relative amplitude features

For simplicity, we will deal with the hypothetical situation in which a single ISI was used for paired stimulations. All of our results generalize to different ISI.

## 2.1 Definitions

Let $k$ be the number of experimental subjects in some subject group, to each of whom $n$ paired stimulations and $m$ test stimulations were elicited.

**Definition 1** *Let $\boldsymbol{P}^{n\times k}$, $\boldsymbol{T}^{m\times k}$ be matrices representing the paired and test potentials evoked across each of the $k$ subjects, such that*

$$\boldsymbol{P} := \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ & & \cdots & \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad \boldsymbol{T} := \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1k} \\ t_{21} & t_{22} & \cdots & t_{2k} \\ & & \cdots & \\ t_{m1} & t_{m2} & \cdots & t_{mk} \end{bmatrix} \tag{1}$$

**Definition 2** *Let $\mu_t := \begin{bmatrix} \mu_1 & \mu_2 & \dots & \mu_k \end{bmatrix}^\top$ be such that $\mu_i$ is the average test potential for the $i$th subject. In other words,*

$$\mu_i = \frac{1}{m}\sum_{j=1}^{m} t_{ji} \tag{2}$$

**Remark 1** *Evoked potentials of both paired and test pulses conform to a Gamma distribution, as discussed in the appendix (?). This means it is always an implicit assumption that $\forall(x \in \boldsymbol{P}, t \in \boldsymbol{T}) | x, t \in \mathbb{R}^+$.*

## 2.2 The $\rho$ and $\delta$ features

**Definition 3** *Let $x \in \boldsymbol{P}_{\star i}$ be a single paired-stimulation MEP in the $i$th experimental subject, and $\boldsymbol{t} = \boldsymbol{T}_{\star i}$ the vector containing all test MEPs of that subject. Then we define two pulse-specific relative amplitude measures,*

$$\rho(x) := \frac{mx}{\sum_{i=1}^{m} t_j} \quad (3)$$

$$\delta(x) := \frac{x}{m} \sum_{j=1}^{m} \frac{1}{t_j} \quad (4)$$

**Remark 2** $\forall x \in \mathbb{R}^+ | \delta(x) \geq \rho(x)$. *(For a proof of this property, consult the appendix.)*

Notice that $\rho$ is the proportion between the potential $x$, evoked by a paired stimulation, with respect to the average potential of single test stimulations. On the other hand, $\delta$ is the average proportion of $x$ with respect to each single test pulse.

The $\rho$ function is not an entirely new contribution. The traditional group level relative amplitude measure is conceived as the average, across all subjects in a group, of the average paired response divided by the average test response of each subject. If we let $S_i$ be the average relative amplitude of the $i$th subject, then $S_i$ has ben traditionally defined as

$$S_i = \frac{\frac{1}{n}\sum_{j=1}^{n} x_{ji}}{\frac{1}{m}\sum_{j=1}^{m} t_{ji}} = \frac{\rho(x_{1i}) + \cdots + \rho(x_{ni})}{n} \quad (5)$$

as it is easy to see from decomposing the sum in the numerator. In other words, the traditionally used measure of relative amplitude, at the subject level, has always been the average $\rho$ in a subject.

### 2.3 The weighted variants $\rho_w, \delta_w$

As stated earlier, action potentials evoked by trasncranial magnetic stimulation follow a Gamma distribution; one that closely resembles an exponential distribution (see **Statistics** section). A valid form of data augmentation that also attempts to deal with the presence of outliers is to weight the averages involved in $\rho, \delta$ using inverse-variance weights. The attempt is to allow for potentials that lay at the tail of the distribution to contribute in a way proportional to their probability.

If we let $\rho_w, \delta_w$ be the weighted versions of $\rho, \delta$, then we have

$$\rho_w(x) := \frac{xm\sum_{j=1}^{m} w_j}{\sum_{j=1}^{m} t_j w_j} \quad (6)$$

$$\delta_w(x) := \frac{\frac{x}{m}\sum_{j=1}^{m} \frac{w_j}{t_j}}{\sum_{j=1}^{m} w_j} \quad (7)$$

where $\mathbf{w}$ is some appropriate weight vector. Since the purpose of $\mathbf{w}$ is to reduce the contribution of outliers to the overall measure of relative amplitude, it is to be expected that the distribution of $\mathbf{w}$ closely resembles that of evoked potentials.

## 3 Empirical results

We used data collected across $N$ =? subjects at the *Laboratory for the Study of Slow-wave sleep activity*, University of Pennsylvania, with $H$ =? healthy controls and $D$ =? diagnosed with major depressive disorder (MDD). Transcranial magnetic stimulation of the motor cortex was elicited to them after a night of baseline sleep and after a night of slow-wave disruption (SWD) sleep. Motor evoked potentials were measured via an electrode (?) placed on the subjects' thumb. In the slow-wave disruption session, an auditory stimulus with sufficient strength to interrupt the normal occurrence of slow-wave sleep, yet not strong enough to wake the subjects, was elicited. This experimental setting produces four distinct categories, two depending on the subject group and two on the type of sleep session underwent.

We trained the exact same random forests model first on the data without pulse-specific measures of relative amplitude, then on the data including some or all the pulse-specific measures of relative amplitude. The model's task was to classifiy each TMS stimulation into one of the four experimental categories. In other words, it faced the challenge of establishing to which type of subject, and under which type of session, each stimulation was elicited, based only on the evoked potential, the ISI and —later— the relative amplitude measures of the stimulation. Our objective was to evaluate whether the inclusion of pulse-specific measures of relative amplitude improved the model's accuracy, and if so in what degree.

### 3.1 Raw data

When trained on the raw data, without the inclusion of pulse-specific relative amplitude measures, the model's accuracy was $A \approx 34.2\%$.

# 4 Appendix

## Proofs

**Theorem:** $\forall x \in \mathbb{R}^+ | \delta(x) \geq \rho(x)$

Recall that

$$\delta(x) = \frac{x}{m}\sum^m \frac{1}{t_j} \tag{8}$$

$$\rho(x) = \frac{xm}{\sum^m t_j} \tag{9}$$

Let $S_1^m = \sum^m \frac{1}{t_j}, S_2^m = \sum^m t_j$. We operate under the assumption that $t_i \in \mathbb{R}^+$. It is the case that

$$\frac{x}{m}\sum^m \frac{1}{t_j} \geq \frac{xm}{\sum^m t_j} \tag{10}$$

$$\iff S_2^m S_1^m \geq m^2 \tag{11}$$

This holds for $m = 1$, since $\frac{1}{t_1} + t_1 \geq 1 \iff 1 + t_1^2 \geq t_1$. So we may assume $S_1^k S_2^k \geq k^2$ . We may only show now that

$$S_1^{k+1} S_2^{k+1} \geq (k+1)^2 \tag{12}$$

$$S_1^{k+1} S_2^{k+1} \geq (k+1)^2 \tag{13}$$

$$(S_1^k + \frac{1}{t_{t+1}})(S_2^k + t_{k+1}) \geq k^2 + 2k + 1 \tag{14}$$

$$S_1^k S_2^k + t_{k+1} S_1^k + \frac{1}{t_{k+1}} S_2^k + 1 \geq k^2 + 2k + 1 \tag{15}$$

$$S_1^k S_2^k + t_{k+1} S_1^k + \frac{1}{t_{k+1}} S_2^k \geq k^2 + 2k \tag{16}$$

We know $S_1^k S_2^k \geq k^2$ and then it suffices to show $t_{k+1} S_1^k + \frac{S_2^k}{t_{k+1}} \geq 2k$. To prove this, simply observe that

$$\frac{1}{t_{k+1}}\sum_{j=1}^m t_j + t_{k+1}\sum_{j=1}^m \frac{1}{t_j} \geq 2k \tag{17}$$

$$\iff \frac{1}{t_{k+1}}(t_1 + t_2 + ... + t_k) + t_{k+1}(\frac{1}{t_1} + \frac{1}{t_2} + ... + \frac{1}{t_k}) \geq 2k \tag{18}$$

$$\iff \left(\frac{t_1}{t_{k+1}} + \frac{t_2}{t_{k+1}} + ... + \frac{t_k}{t_{k+1}}\right) + \left(\frac{t_{k+1}}{t_1} + \frac{t_{k+1}}{t_2} + ... + \frac{t_{k+1}}{t_k}\right) \geq 2k \tag{19}$$

$$\iff \overbrace{a + \frac{1}{a} + b + \frac{1}{b} + ... + n + \frac{1}{n}}^{2k \text{ terms}} \geq 2k \tag{20}$$

We have $\min f = 2$ for $f(x) = x + \frac{1}{x}$ for $x \in \mathbb{R}^+$. Then $\min(a + \frac{1}{a} + ... + n + \frac{1}{n}) = 2k$ for $a, ..., n \in \mathbb{R}^+$, which concludes the demonstration.