



# Reliability of TMS phosphene threshold estimation: Toward a standardized protocol



Chiara Mazzi <sup>a, b</sup>, Silvia Savazzi <sup>a, b</sup>, Arman Abrahamyan <sup>c</sup>, Manuela Ruzzoli <sup>d, \*</sup>

<sup>a</sup> Perception and Awareness (Panda) Laboratory, Department of Neuroscience, Biomedicine and Movement Sciences, University of Verona, Italy

<sup>b</sup> National Institute of Neuroscience-Verona, Italy

<sup>c</sup> Department of Psychology and Neurosciences Institute, Stanford University, USA

<sup>d</sup> Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain

## ARTICLE INFO

### Article history:

Received 10 June 2016

Received in revised form

30 January 2017

Accepted 31 January 2017

Available online 2 February 2017

### Keywords:

Phosphene threshold

Transcranial magnetic stimulation (TMS)

Psychophysics

Vision

## ABSTRACT

**Background:** Phosphenes induced by transcranial magnetic stimulation (TMS) are a subjectively described visual phenomenon employed in basic and clinical research as index of the excitability of retinotopically organized areas in the brain.

**Objective:** Phosphene threshold estimation is a preliminary step in many TMS experiments in visual cognition for setting the appropriate level of TMS doses; however, the lack of a direct comparison of the available methods for phosphene threshold estimation leaves unsolved the reliability of those methods in setting TMS doses. The present work aims at fulfilling this gap.

**Methods:** We compared the most common methods for phosphene threshold calculation, namely the Method of Constant Stimuli (MOCS), the Modified Binary Search (MOBS) and the Rapid Estimation of Phosphene Threshold (REPT). In two experiments we tested the reliability of PT estimation under each of the three methods, considering the day of administration, participants' expertise in phosphene perception and the sensitivity of each method to the initial values used for the threshold calculation.

**Results:** We found that MOCS and REPT have comparable reliability when estimating phosphene thresholds, while MOBS estimations appear less stable.

**Conclusions:** Based on our results, researchers and clinicians can estimate phosphene threshold according to MOCS or REPT equally reliably, depending on their specific investigation goals. We suggest several important factors for consideration when calculating phosphene thresholds and describe strategies to adopt in experimental procedures.

© 2017 Elsevier Inc. All rights reserved.

## 1. Introduction

Transcranial magnetic stimulation (TMS) is a non-invasive method for temporarily and reversibly interacting with functions of a targeted brain area. The experimental practice involves setting different parameters to define the TMS protocol (i.e., stimulation intensity, frequency duration, and pulse latency) [1]. Among those parameters, the intensity of the magnetic field (the percentage of the maximum stimulator output, MSO), has special relevance since it is used to measure and standardize the stimulus doses among individuals. Recent evidence suggests that choosing high or low TMS doses can either suppress or enhance behavioural response

outcomes, respectively [2–5] thus making TMS intensity a critical parameter for data interpretation in cognitive models and for clinical applications.

There are different approaches for setting TMS intensity and one of them involves choosing a fixed value for all participants (e.g. 65%) [6], thus saving time. However, differences in cortical excitability, cortical structure and skull shape between participants require adjusting the TMS pulse individually, for example calculating psychophysical threshold on motor (motor threshold – MT) or visual (phosphene threshold – PT) cortices, respectively. While there are well-established procedures for computing MT [7–9], PT procedures are scarce. This is an important gap in experimental routine, because MT ceases to provide a reliable index for individually setting TMS intensity moving from motor cortex to other cortical areas [10]. Conversely, PT is a good candidate for establishing the effectiveness of the stimulation of the posterior brain.

\* Corresponding author.

E-mail address: [manuela.ruzzoli@upf.edu](mailto:manuela.ruzzoli@upf.edu) (M. Ruzzoli).

Indeed, reliable phosphenes have been reported after stimulation of retinotopic areas both in occipital and parietal cortex [11–13], making PT a cardinal parameter in investigating complex cognitive processes, like visual attention and awareness [14].

A standardized procedural definition of PT calculation is particularly challenging since phosphenes are subjective reports. Adding to complexity, PT can be affected by several other factors. For example, keeping the eyes closed or open changes the PT [15,16], as do the light level of the environment [17], instructions given to participants, participants' expectancy and expertise in phosphene perception, and the method used to calculate the threshold. Our aim is to compare psychophysical methods to calculate the PT.

Three main procedures are typically used to estimate PT: the Modified Binary Search (MOBS) [18,19], the Method of Constant Stimuli (MOCS) [20], and the Rapid Estimation of Phosphene Threshold (REPT) [21]. We conducted two experiments. In the first one three methods (i.e. MOBS, MOCS, and REPT) were compared within one session on three different days. This way, we aimed at testing the reliability of each method between and within participants in repeated sessions. We also considered participants' expertise with phosphene phenomenon comparing two groups (naïves vs. experts). In Experiment 2, we used data from Experiment 1 as prior to evaluate possible improvements in PT estimation due to the initial intensity range of two of the PT methods (MOCS and REPT) employed in Experiment 1.

## 2. Method

### 2.1. Experiment 1

#### 2.1.1. Participants

A total of 22 healthy participants with normal or corrected to normal visual acuity were recruited for the experiment and reimbursed for their participation. They had no contraindications to TMS, as assessed by the safety screening questionnaire (adapted from Ref. [22]). Three participants were not able to perceive phosphenes and were excluded from the experiment. Moreover, two participants dropped out after the first day of testing and they did not complete the experiment. The remaining 17 participants (4 males), including two authors (C.M. and S.S.), took part in the experiment (19–39 years old; one left-handed). Ten of them were naïve to phosphene. All of them but the two authors were naïve to the goals of the study.

All the participants gave their written informed consent before participating in the study, which was conducted in accordance with the 2013 declaration of Helsinki and approved by the local Ethics Committee.

#### 2.1.2. Apparatus and procedure

The participants were seated in a dark room at a distance of 57 cm from a 17 inch LCD monitor (LG L1753HM), with chin and forehead steadied so that eyes were aligned with the centre of the screen. The participants wore earplugs and a close-fitting cap with the 10–20 international system marked on.

To illustrate what a phosphene is, we showed participants some examples of phosphenes drawings [15]. Participants were instructed to not expect a flash of light, instead to covertly explore all the visual field and to look for a spatially circumscribed visual change (in brightness or texture) on the background.

Participants were adapted to the light conditions of the experimental room for at least 5 min prior to the experiment. They were instructed to keep the fixation steady on the central spot on a completely black screen. To confirm that participants perceived authentic phosphenes, some criteria [23] such as the dependence

on the stimulated hemisphere (i.e. phosphenes in the contralateral visual field [24]), the dependence on gaze direction [24], and the visibility with eyes both open and closed [25] had to be satisfied. We further tested phosphenes reliability asking participants to describe the position, size, texture, colour and shape of perceived phosphenes. A few days before the beginning of the experiments, naïve participants were invited to the laboratory and familiarized with TMS procedure and phosphenes perception. Furthermore, all the participants ran at least one threshold estimation before the beginning of the experiment as training.

At the beginning of each session, the coil was placed over O1 (we only tested the left hemisphere) and it was slightly moved in all directions in a region within a circle of 2 cm in diameter centered on O1 until bright and reliable phosphenes were induced. By means of a mechanical arm ([www.manfrotto.com](http://www.manfrotto.com)) the coil was then fixed over this “hotspot” (i.e. the best location where circumscribed, right hemifield-lateralized phosphenes were perceived) and this position was marked on the cap to be used throughout the experimental session. The coil was placed tangentially to the skull, parallel to the mid-line with the handle pointing upwards to avoid unspecific activation of neck and shoulder muscles. Single-pulse TMS was discharged using a Magstim Rapid<sup>2</sup> (Magstim Company Limited, Whitland, UK) system (MSO = 3.5 T) through a 70 mm figure-of-eight coil. Both TMS pulse trigger and response acquisition were controlled by Matlab (The MathWorks, Natick, MA) and the Psychtoolbox (ver. 3.00 [26–28]).

#### 2.1.3. Phosphene threshold measurement

We collected data from three psychophysical methods (MOBS, MOCS, and REPT) on three days (Monday, Wednesday and Friday of the same week within participant). Each day, the order of methods was counterbalanced among participants. The experiment lasted about half an hour each day.

On each trial, participants were instructed to keep their eyes open, to maintain visual fixation on the central spot and, after each TMS pulse, to report the presence or the absence of a phosphene with a “yes/no” response by pressing two different keys. Based on the response and psychophysical method, the computer automatically adjusted the TMS intensity for the subsequent trial, thus avoiding the experimenter to manually adjust the intensity after each response. Participants had to press the spacebar to continue to the next trial, thus promoting coordination between pulse delivery and attending to phosphenes. The inter-pulse interval was longer than 3 s for all methods employed, in accordance with the safety guidelines [22].

#### 2.1.4. Modified Binary Search (MOBS)

MOBS [18] is an adaptive non-parametric algorithm which uses the mathematical method of bisection combined with additional heuristics to estimate PT. Initially, the first TMS intensity is set to the midpoint of the available range (51% of MSO). The extreme values (1% and 100%) are called the bottom and top elements and 51% is called middle element. After participant's response, the search range for PT is updated by setting new values of each element. For example, if participant responded “no” to seeing phosphene at 51% intensity, the bottom and top elements are set to 51 and 100%, respectively, and the middle element is set to 76%. The middle element (76%) is the next stimulation intensity. If participant responds “yes” at this intensity, the new search range will be set from 51 to 76% while if the response is “no” the search range is moved from 76% to 100%. If two consecutive “yes” (or “no”) are reported, the next stimulation intensity will be the bottom (or top) element of the available range. A process of “regression” is needed if the participant responds inconsistently to her/his previous response. In this case, all elements are moved up by one, losing the

top element and setting the bottom one to the appropriate end of the range so that the range is widened. The procedure ends after a fixed number of reversals (i.e. changes from unseen to seen and vice versa—6 in this experiment). The threshold is calculated as the mean of top elements from each range used.

The number of trials required for estimating a threshold depends on the consistency of participant's reports (in this experiment between 8 and 16 trials, with an average test time of 36.61 s, SD = 10.83).

#### 2.1.5. Method of Constant Stimuli (MOCS)

MOCS is a non-adaptive, psychophysical method to calculate thresholds where a pre-arranged set of stimuli (to cover sub- and supra-threshold values) is tested many times in a randomized order. In this experiment, 13 randomly intermixed TMS intensities (from 51% to 87% in steps of 3%) were employed. Each intensity was tested 10 times per session for a total of 130 pulses. While it is possible to choose less number of intensities and less number of pulses [29,30], we chose such a large range to avoid running a range estimation procedure prior to running the MOCS. However, in order to understand whether MOCS can be reliable with less intensity steps and less repetitions, we conducted a Monte Carlo simulation. The average testing time was about 10 min. For each participant, we computed the proportion of “yes” responses for each TMS intensity. These data were fitted with a cumulative Weibull psychometric function using the maximum likelihood procedure implemented in Palamedes toolbox [31] and with fixed lapse rate of 4%. The TMS intensity at which the participant perceived phosphene 50% of the times was taken as the threshold value.

#### 2.1.6. Rapid Estimation of Phosphene Threshold (REPT)

REPT [21] consists of a Bayesian adaptive staircase procedure where TMS intensities are selected by combining Bayes' rule and information theory [32] to maximize the expected information gain from each trial and, at the same time, to minimize the uncertainty in the probability distribution of the parameters. Being a Bayesian procedure, REPT updates the posterior distribution of a range of threshold and slope values after each successful or unsuccessful response. To select the next TMS intensity, REPT computes the entropy based on the updated posterior distributions. Smaller entropies indicate psychometric functions that better match subjects' performance. Such placement of the TMS intensity is expected to provide an outcome that leads to discovering the psychometric function that best describes the observer. PT was calculated in 30 trials for every participant. The procedure lasted on average 1 min and 31 s (SD = 15.84). The range of stimulation intensities allowed for PT estimation was generally set between 45% and 75% of MSO, however each participant used a different range within the general range 45–75 based on his/her own responses (mean range calculated across all REPTs and participants = 60%; SD = 11% of MSO) (please refer to [21] for additional details about the staircase functioning).

The estimated threshold according to REPT corresponds to the stimulation intensity that evoked phosphene 50% of the times (position parameter of a Weibull function fitted to the proportion of phosphene responses between 0 and 1 with a lapse rate of 4%).

#### 2.1.7. Statistical analysis

Individual PTs (expressed in percentage of MSO) were calculated for each method during each of three days of testing.

We first performed a 2-way repeated measures analysis of variance (ANOVA) with *Method* (MOBS, MOCS and REPT) and *Day* (Monday, Wednesday and Friday) as within-subjects factors. To appreciate differences in PT estimation due to the three methods we performed three one-way ANOVA with *Method* as within-

subjects factor on the three days data separately and three one-way ANOVA with *Day* as within-subjects factor on the data from three methods separately. Furthermore, participants' expertise was taken into account by a 3-way mixed ANOVA with *Method* (MOBS, MOCS and REPT) and *Day* (Monday, Wednesday and Friday) as within-subjects factors and expertise as a between subjects factor (expert versus naïve participants). In order to appreciate the within subjects' variability across different sessions, we calculated the standard deviation for each method and performed a within-subjects ANOVA with *Method* as main factor (MOBS, MOCS, REPT). Post-hoc tests were Bonferroni-corrected t-tests.

#### 2.1.8. Monte Carlo simulation

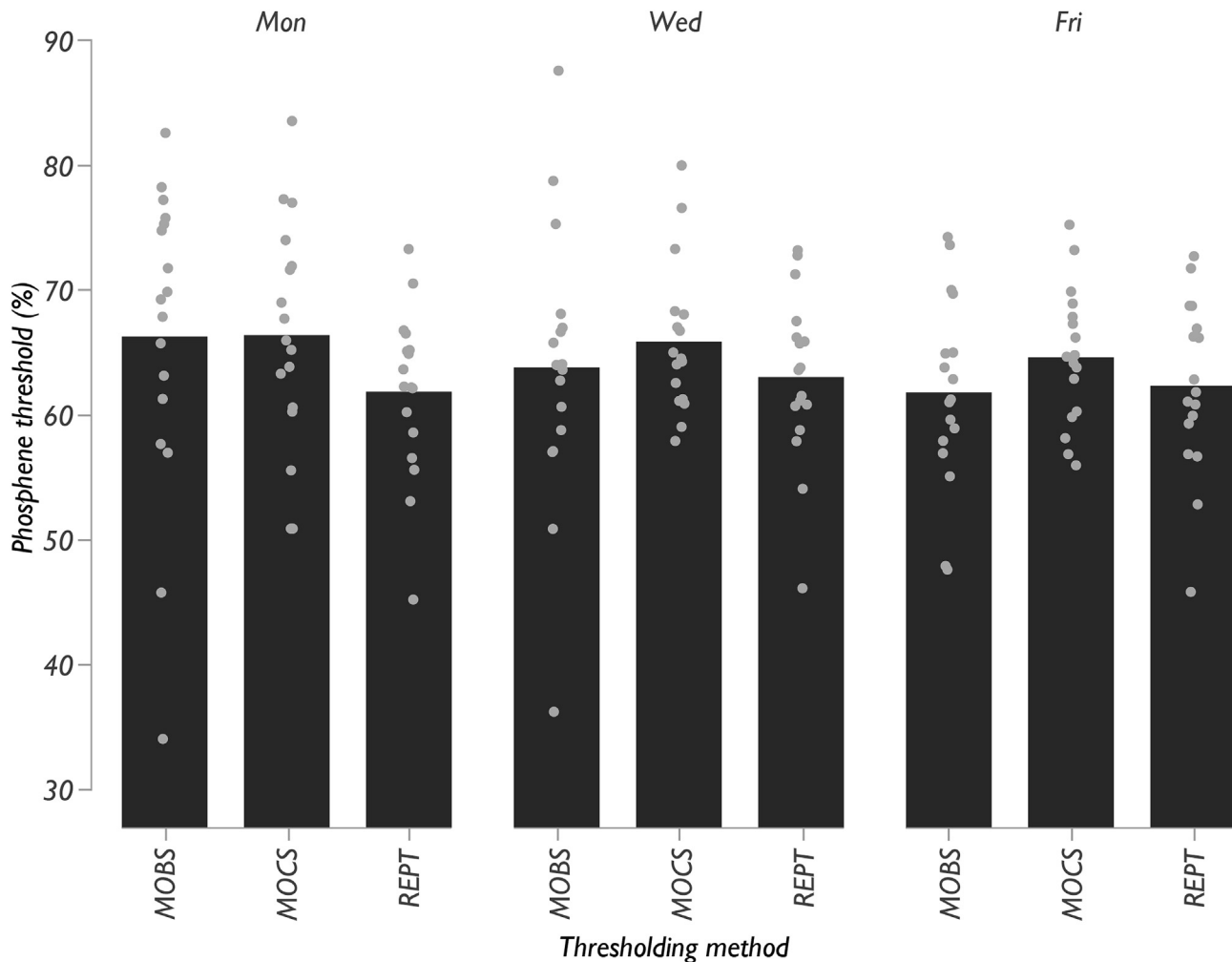
To estimate the minimum number of trials per intensity required to compute reliable thresholds, we ran a Monte Carlo simulation using MOCS data collected on Day 1 in Experiment 1. Simulation probabilities were computed as proportion of correct responses. For each participant and for each number of pulses (2–20), 300 simulations were made. To estimate 50% phosphene threshold a Weibull function was fit after each simulation. The median of 300 simulated threshold values was used as an estimate of individual threshold and standard deviation as a measure of phosphene reliability. We simulated using a range of 13 intensities used in Experiment 1, which were spaced in steps of 3%, but also ran a simulation with a more sparsely distributed range of 7 intensities which covered the same intensity range but spaced in steps of 6%. Data from four participants were excluded because low goodness of fit of the Weibull function. This was due to atypical response patterns, which resulted in “locking” the fit to similar psychometric functions creating an erroneous impression of stable thresholds.

## 2.2. Results

The 2-way repeated measure ANOVA ( $n = 17$ ) revealed a significant main effect of *Method* [ $F(2,32) = 3.636$ ;  $p < 0.05$ ;  $\eta^2 = 0.19$ ]. Post-hoc comparisons showed that REPT thresholds (mean REPT = 62.49, SD = 6.72) were lower than MOCS (mean MOCS = 65.70, SD = 6.95,  $p < 0.05$ ). No significant differences in terms of *Day* of administration were found [ $F(2,32) = 0.963$ ;  $p = 0.392$ ;  $\eta^2 = 0.06$ ]. However, the interaction between *Method* and *Day* [ $F(4,64) = 2.524$ ;  $p < 0.05$ ;  $\eta^2 = 0.14$ ] was significant (Fig. 1). To further investigate this effect, three separate one-way ANOVA with *Method* as within-subjects factor were performed for each day separately. Only in the first day of administration REPT produced lower threshold [ $F(2,32) = 5.422$ ;  $p < 0.01$ ;  $\eta^2 = 0.25$ ; mean REPT = 61.94, SD = 6.74] with respect to MOCS (mean MOCS = 66.47, SD = 9.14). We also performed three separate one-way ANOVA with *Day* as within-subjects factor on each method separately; however no significant results were present.

We suspected that the difference in PTs on the first day could be driven by participants' expertise. That is, experts ( $n = 7$ ), being more accustomed to phosphene perception, could have had lower and/or less variable PT estimations compared to naïves participants ( $n = 10$ ). However, *expertise* was not significant as main effect [ $F(1,15) = 0.723$ ;  $p = 0.41$   $\eta^2 = 0.05$ ], (Fig. 2), nor interacted with other factors.

To assess the reliability of each method across sessions, we computed the standard deviation for each method across the three days of testing, individually, such that small values indicate more reliable measurements within a particular method. Standard deviations were entered into a one-way ANOVA with *Method* as within-subject factor. We found that the three methods were different in reliability as shown on Fig. 3 [ $F(2,32) = 5.304$ ,  $p < 0.01$ ;  $\eta^2 = 0.25$ ]. Specifically, MOBS produced more variable thresholds compared to MOCS ( $p < 0.01$ ). There was no difference in reliability



**Fig. 1.** The effect of psychophysical method and testing day on phosphene thresholds. Bars indicate group mean values (PT in % MSO) for each psychophysical method and day. The dots represent individual PT estimation. Over three days of testing, REPT produced lower thresholds compared to MOCS or MOBS. However, the effect was driven by the first day of testing (Mon), during which REPT measures were lower compared to MOCS or MOBS. On second (Wed) and third (Fri) days of testing the psychophysical method employed did not affect phosphene thresholds.

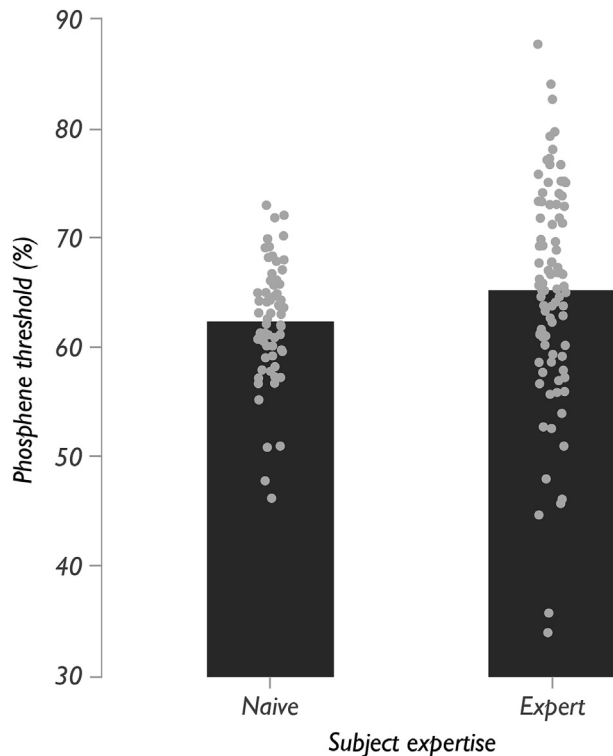
between REPT and MOBS ( $p = 0.179$ ) and between REPT and MOCS ( $p = 1$ ).

In order to obtain a measure of reliability of the data fitting with the Weibull function, we also decided to consider a further parameter: the standard deviation of the goodness of fit (Weibull function) calculated for each MOCS threshold measurement according to the non-parametric bootstrap procedure implemented in Palamedes toolbox. This parameter can only be obtained for MOCS data because for MOBS and REPT it is often the case that one level of intensity is administered only once in a run, therefore making the fitting procedure highly unreliable. We noticed that the standard deviation of the Weibull fitting, performed on MOCS data, exceeded the value of two at least in one session out of three in four participants. We therefore decided to exclude them and we run the same analyses performed before on the remaining thirteen participants: The 2-way repeated measure ANOVA (*Method*  $\times$  *Day* of administration) confirmed the main effect of *Method* [ $F(2,24) = 6.147$ ;  $p < 0.01$ ;  $\eta^2 = 0.34$ ] as for the entire sample of participants ( $n = 17$ ). The interaction between *Method* and *Day* was significant [ $F(4,48) = 2.977$ ;  $p < 0.05$ ;  $\eta^2 = 0.20$ ]. Data from 14 participants highlighted not only a difference between REPT and MOCS thresholds in the first day of administration REPT produced lower threshold with respect to MOCS [ $F(2,24) = 11.657$ ;  $p < 0.01$ ;

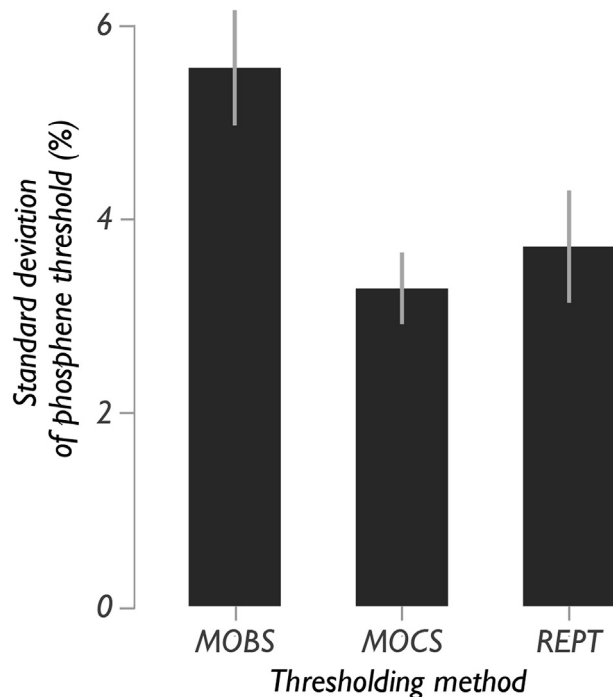
$\eta^2 = 0.49$ ; mean REPT = 62.08, SD = 4.19; mean MOCS = 66.47, SD = 9.14], but also a further difference between REPT and MOBS thresholds (mean MOBS = 68.46, SD = 7.14;  $p < 0.05$ ). Finally, post-hoc comparisons concerning the ANOVA on standard deviation data in the three methods showed a tendency for REPT to be less reliable than MOCS ( $p = 0.055$ ).

Taken together, these results show that (1) MOBS is highly variable; (2) MOCS is the most stable method having the smallest changing across sessions; and (3) REPT provides the lower TMS doses administrable but it tends to be less reliable than MOCS. Based on data from Experiment 1 we can conclude that MOCS and REPT are reliable methods in terms of number of administrations. We aimed at further testing the reliability of these two methods, focusing on their sensitivity for the initial range of power values. Indeed both MOCS and REPT could be susceptible to the range of intensities chosen as default parameters when estimating thresholds [33,34]. In Experiment 2 we enrolled a subset of participants already tested in Experiment 1. The rationale was to capitalize on real data from Experiment 1 in order to tailor the input parameters for subsequent calculations (both using MOCS and REPT). In Experiment 2 we ran MOCS adopting as midpoint value the average of all REPT data collected in Experiment 1 (across sessions and participants). We also performed three repetitions of REPT with





**Fig. 2. The effect of expertise.** Expert participants who were accustomed to perceiving phosphenes, did show phosphene thresholds similar to those of naïve participants, who were introduced to phosphene perception during this study.



**Fig. 3. Reliability.** To assess similarity of each method reproducing phosphene thresholds within participants, we computed the standard deviation of the PT estimation across three days of testing, for each method. By this measure, MOBS was more unreliable compared to MOCS. REPT had a trend to be less reliable compared to MOCS, but the difference was not significant. There was no difference between MOBS and REPT. Error bars are within-subjects SEMs.

three intensity ranges calculated considering as midpoint the individual REPT thresholds collected in the previous experiment.

To assess how reliability of phosphene threshold estimation changes as a function of number of delivered pulses, we used Monte Carlo simulation using participants' responses from Experiment 1, Day 1 ( $n = 13$ ). Data from 4 participants were excluded because of the high variability ( $>2$  SD) in the Weibull fitting. For each participant, we simulated thresholds 300 times and computed the median phosphene threshold (and standard deviation) for each set of pulses. The simulation was completed for step-size intensity of 3% (a total of 13 intensities) and 6% (a total of 7 intensities) in the range from 51% to 87%. To assess how accurately phosphene thresholds change depending on the number of pulses, we first computed the difference between thresholds estimated using different number of pulses. In other words, we computed the difference between  $PT(N+1 \text{ pulse}) - (N \text{ pulse})$ . We found that the number of pulses did not significantly affect threshold estimates: this was true for 13 intensities (Fig. 5A) [ $F(18,216) = 1.05$ ,  $p = 0.4$ ] and 7 intensities (Fig. 5C) [ $F(18,216) = 0.64$ ,  $p = 0.9$ ].

To find out which parameters affect the reliability of threshold estimation, we computed the standard deviation of 300 simulated thresholds for both 13 and 7 intensities (Fig. 5B and D). The analysis of interaction between of the number of pulses (2–20) and range of intensities (3% or 6% step) was significant [ $F(18,216) = 5.60$ ,  $p < 0.001$ ]. This result indicates that when we spaced intensities in steps of 3% the threshold variability was smaller compared to when the intensities were incremented by 6% (compare Fig. 5B and D). Therefore, more reliable thresholds are produced when intensities are more densely distributed. Next, we wanted to find out what is the minimum number of pulses required for reliable estimation of phosphene thresholds. Indeed, there was a main effect of number of pulses in the case of standard deviations both when intensities were spaced at 3% [ $F(18,216) = 58.11$ ,  $p < 0.01$ ] and 6% [ $F(18,216) = 24.26$ ,  $p < 0.01$ ]. Pairwise comparisons, corrected using the Holmes method, revealed that starting from 7 pulses and onwards there was no difference in standard deviations, indicating that reliability did not grow as the number of pulses was increased from 7 on when intensities were spaced in steps of 3%. Similarly, in the case of intensities spaced at 6%, reliability did not improve after delivering 6 or more pulses. Results from the Monte Carlo simulation suggest that it is recommended to deliver at least 7 pulses per intensity to obtain accurate and reliable threshold measurement.

### 3. Experiment 2

#### 3.1. Method

##### 3.1.1. Participants

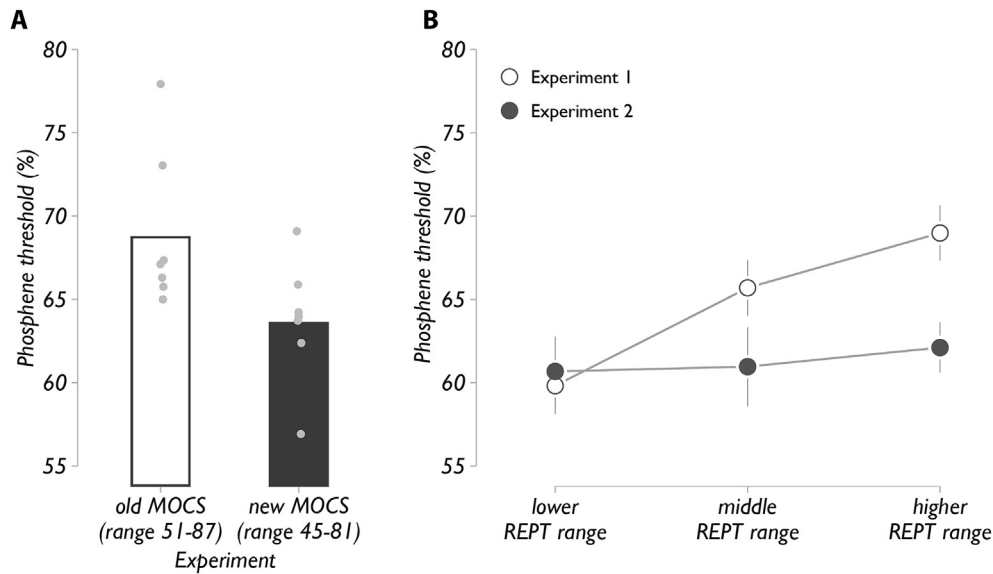
Seven participants (1 male, age range 23–39 years) who participated in Experiment 1 were tested in Experiment 2. Written informed consent before participating in the study was collected from all the participants. The experiment was conducted in accordance with the 2013 declaration of Helsinki and approved by the local Ethics Committee.

##### 3.1.2. Apparatus, procedure and phosphene threshold measurement

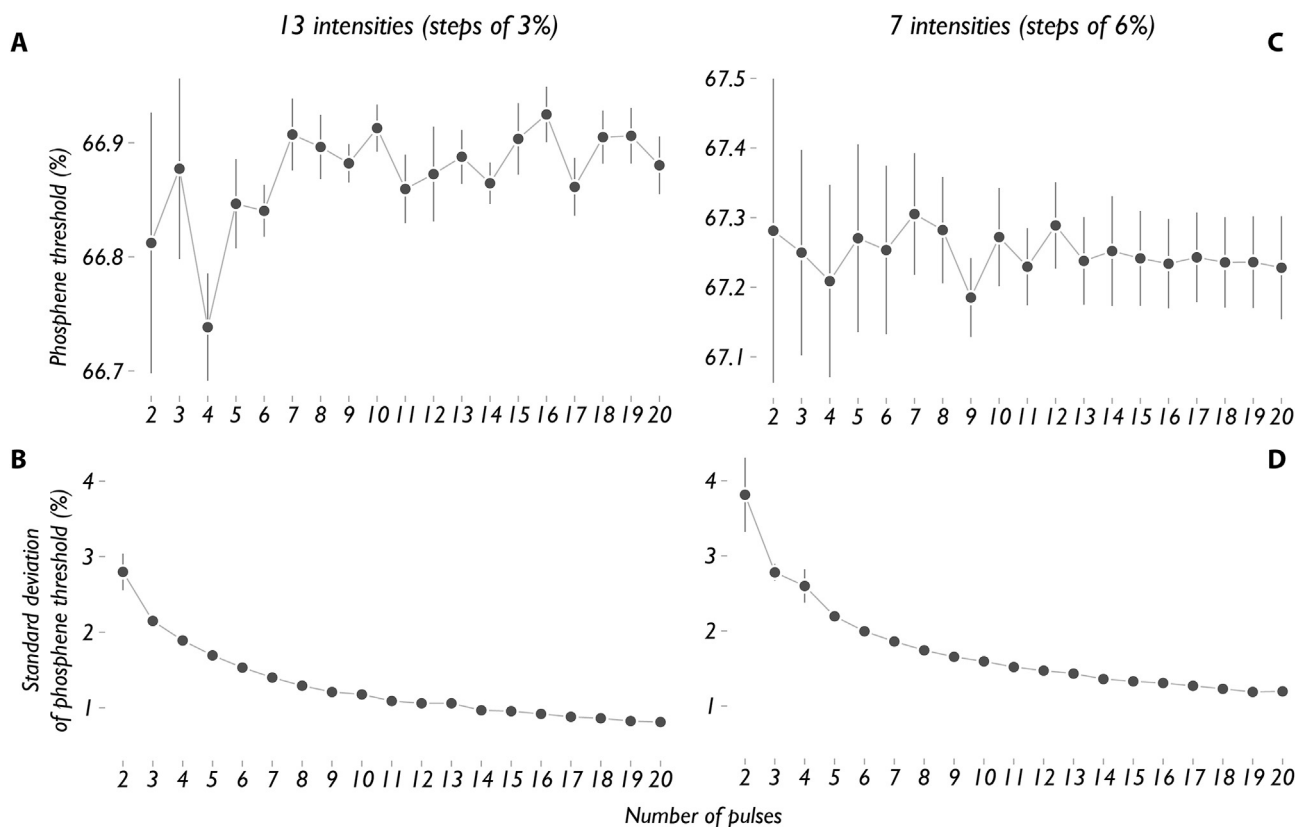
Apparatus and procedure were the same as in Experiment 1.

PT was consecutively assessed in a single day session by MOCS and REPT in about half an hour. Each participant performed:

- one MOCS whose midpoint value (63% of MSO) was the average of all REPT data collected in Experiment 1. The intensities employed ranged from 45% to 81% of MSO (step of 3%). Ten pulses were administered for each intensity for a total amount of 130 pulses;



**Fig. 4.** The effect of stimulation range on phosphene thresholds. **A:** MOCS PT calculated from two different ranges in Experiment 1 and 2, respectively. **B:** REPT PT calculated using 3 different ranges of initial values in Experiment 1 (open circles) and 2 (filled circles).



**Fig. 5.** Monte Carlo simulation. Estimating optimal parameters for MOCS using Monte Carlo simulations. To compute the minimum number of pulses that requires for reliable estimation of phosphene thresholds, we simulated observers based on the data from Experiment 1. Phosphene threshold estimates were generally not affected by the number of pulses (A and C). However, the reliability of phosphene threshold estimates depended on the number of TMS pulses. To reliably estimates thresholds, MOCS required at least 7 pulses when 13 different intensities were used (B) and 6 pulses when a range of seven intensities was selected (D).

- three REPT with different intensity range each calculated considering as midpoint the individual REPT thresholds collected in the three sessions of Experiment 1. As a result, all the participants had three different ranges of intensities

depending on their own previous data. The range consists of 31 points, therefore, if one REPT PT was 54% in Experiment 1, the REPT range used in Experiment 2 ranged from 39% to 69% (or,

another example, from 52% to 82% for a previous REPT PT of 67%).

The order of the administration of MOCS and REPT was counterbalanced among participants.

### 3.1.3. Statistical analysis

Individual PTs (in percentage of MSO) were calculated for MOCS and REPT. We first conducted a one-way repeated measure ANOVA comparing the new MOCS values with the mean of the three MOCS collected in the first experiment. Furthermore, data from the REPT thresholds from Experiments 1 and 2 were re-arranged in ascending order for each participant. Subsequently, a two-way repeated measure ANOVA was carried out using *REPT threshold* (3 levels) and *Experiment* (1 or 2) as within-subjects factors. Please note that the three levels of *REPT threshold* considered in the two-way ANOVA (“lower REPT”, “middle REPT”, “higher REPT”) are labelled according to the data collected in Experiment 1.

### 3.2. Results

In Experiment 2, we found that using a moderately weaker stimulation intensities for MOCS slightly lowered PTs compared to Experiment 1 but the difference was not significant [ $F(1,6) = 4.26$ ,  $p = 0.08$ ;  $\eta^2 = 0.42$ ] (Fig. 4A).

Concerning REPT (Fig. 4B), individually adjusting the range of TMS intensities did not change thresholds (filled circles), whereas in Experiment 1, in which a common range was used for all participants, REPT had greater variability (open circles). Indeed, the factor *REPT threshold* was significant [ $F(2,12) = 17.31$ ,  $p < 0.01$ ;  $\eta^2 = 0.74$ ], however, the effect was driven by the interaction between *REPT threshold* and *Experiment* [ $F(2,12) = 13.86$ ;  $p < 0.01$ ;  $\eta^2 = 0.70$ ], where only in Experiment 1 there was a significant differences among the three REPT values collected (all  $ps < 0.01$ ).

Altogether, results from Experiment 2 showed that tailoring the intensity range used as input in case of REPT decreased the variability.

## 4. Discussion

Phosphene threshold (PT) is widely used in cognitive neuroscience as an index for non-invasive assessment of visual cortex excitability [35], but can also be used as a covariate of subjective variability with other physiological measures (e.g. Ref. [36]). Furthermore, it is measure of TMS doses in basic research and clinical purposes [3–5,35]. Due to its subjective nature, the need for a standardization of PT calculation is essential for data comparison and generalized conclusions. Here, we aimed at assessing the reliability of three psychophysical methods used for PT estimation, comparing MOCS, MOBS and REPT. Results showed that MOBS is the less reliable method across three sessions of administration. MOCS and REPT are equivalent in terms of within participant's variability across sessions, even if REPT estimation was more unstable the first day of application. We further assessed the dependency of the final estimation on the initial values of TMS intensity allowed for the PT calculation in MOCS and REPT only. MOCS appeared to be stable regardless the midpoint initial value. REPT PT estimation tended to linearly increase as the range of initial values increased (Experiment 1), however, if the initial intensity range is individually adjusted, REPT PT estimation appeared to be very stable (Experiment 2).

Here, we carefully maintained other parameters known as affecting PT fixed, such as the instruction given to the participants, the environmental light, and the pre-experimental dark adaptation (see *Apparatus and procedure*). Especially in studies where PT is the

dependent variable used for inferring the excitability of visual areas, limitation of possible sources of variability is essential. To account for reproducibility of the results, we made use of manual self-administration for TMS pulses release, therefore ensuring that participant's expectation to TMS pulses was uniform trial-by-trial. Self-triggering of TMS pulses is an important aspect that might affect perceptual report in terms of expectation [37] and sense of agency [38]. A note of caution should, however, be made about the stimulation site targeting since no neuro-navigation system was used across experimental sessions and this could introduce a possible uncontrolled source of variance.

In the first experiment, we also considered the possibility that experienced participants might be more reliable in reporting phosphenes since “the ability to detect phosphenes must be “learned”, much as a gestalt processes” ([23] pag. 136). Although Fig. 2 seems to show a difference between the two tested populations, the size of the effect was negligible, suggesting that, once a participant understood what a phosphene is and is able to report it, then her/his ability to detect phosphenes in further sessions is stable. We can therefore claim that PT is a reliable index to adopt in clinical setting for example, where participants are tested many times on the same task in order to establish the efficacy of a treatment.

In the second experiment, we considered the possibility of an influence of the estimated PT on the initial range of intensities for the two most reliable methods, namely MOCS and REPT. We reasoned that threshold estimation can be sensitive to the extremes of a given input range [33,34]. Therefore, we capitalize on real data from Experiment 1 for establishing new range input in that particular sub-population of subjects. We found that MOCS can be considered a stable method, given that PT estimates are not sensitive to the range of the intensity used. REPT appeared to be more sensitive to the initial input values (Experiment 1), even if, individual adjustment significantly improved PT reliability across sessions (Experiment 2; Fig. 4B). We think that the rationale we applied in Experiment 2 can be easily applied in any TMS lab context where participants' database is often fixed. Indeed tested screened participants, possibly having individual MRI (or fMRI) scans, has a strong impact on research or clinical costs. Here the idea is to turn constraints into advantages. A long-term strategy for a lab involved in TMS application is to tailor as much as possible the TMS parameters setting based on individual data. Thus, whatever the psychophysical method chosen for PT estimation, the initial input values can be set based on real data, rather than relying on simulations or choosing arbitrarily.

In determining the best method for PT calculation, time consumption is another important factor to consider. Since PT calculation can be the preliminary step for a TMS study, the experimenter would like to limit resources for this procedure. Faster PT estimation has also the potential to prevent discomfort to the subject due to multiple pulses, as well as weakening the possibility to induce carry-over physiological changes that might interact with the target intervention in a study [39,40]. Even if we did not systematically store this information for all the experimental conditions, we can estimate that MOBS (average test time of 36.61 s,  $SD = 10.83$ ) and REPT (average test time of 1 min and 31 s,  $SD = 15.84$ ) are the fastest methods and MOCS is the longer one (about 10 min). Considering a trade-off between time consumption and reliability, we can conclude that REPT might be the best option to consider in PT estimation when this step is the preliminary one to be followed by the actual experiment. Conversely, if the PT estimation is the core of the experiment (e.g. Ref. [12]) MOCS would be a better solution, given its higher reliability. Indeed, MOCS data (considering an appropriate number of intensities and trials per intensities) can be modelled based on psychometric fitting

procedures (here we used a Weibull function). Psychophysical modelling based on TMS intensity can be useful not only for phosphene threshold variable, but also for the parameterization of other physiological variable in a study (e.g. TEP: TMS-evoked potentials [41]). Furthermore, the fitting procedure on MOCS data offers the possibility to establish a descriptive value for assessing the reliability of participants' responses. For example, in this study we considered the standard deviation of the Weibull fitting for pinpointing outliers in the tested sample.

A third suggestion aimed at maintaining the high reliability of MOCS while reducing the time of administration can be to adopt a shorter MOCS procedure by using a minimum of 7 pulses per intensity and having sufficiently sparse distribution of intensities. Indeed, the results of the Monte Carlo simulation and data present in literature coming from Kammer's lab [29,30] are in favour of adopting such a procedure.

The last aspect we would like to highlight is related to the intrinsically variable nature of any threshold estimation (not just PT). This aspect forces the experimenter to limit the duration of the study after PT estimation or rather, consider repeating the estimation several times in long experiments, to ensure that the TMS doses compensate for fatigue effects, habituations or random and spontaneous variation in the excitability of the stimulated cortex. To this respect we suggest to choose a fast method, like REPT for PT calculation. A different elegant possibility is to combine the two methods for PT estimation. So for example, it seems optimal to first assess PT by MOCS, and then apply REPT for getting fastest PT values in between a long experiment, customizing the intensity range based on the previous MOCS value.

In conclusion, PT provides a reliable measure of assessing visual cortical excitability [25,35] if taking care of the aspects that might affect it, such as the instructions given to the participants (e.g., visual examples [15]), a fixation point on the screen, or the self-administration of TMS pulses by the participant and manual responses. Our data suggest that MOCS and REPT (free download: <http://www.psych.usyd.edu.au/tmslab/rapid2andrept.html>) provide reliable PT estimations. We also suggest considering the possibility to combine MOCS and REPT methods to compensate for threshold variability and to consider the possibility to feed the chosen algorithm with individual data whenever possible.

## Funding

M.R. was supported by Juan de la Cierva postdoctoral fellowship (JCI-2012-12335) from the Spanish Government.

## Acknowledgment

We would like to thank Prof. T. Kammer and another anonymous reviewer for their fruitful comments on the original version of the manuscript.

## References

- [1] Sandrini M, Umiltà C, Rusconi E. The use of transcranial magnetic stimulation in cognitive neuroscience: a new synthesis of methodological issues. *Neurosci Biobehav Rev* 2011;35:516–36. <http://dx.doi.org/10.1016/j.neubiorev.2010.06.005>.
- [2] Perini F, Cattaneo L, Carrasco M, Schwarzbach JV. Occipital transcranial magnetic stimulation has an activity-dependent suppressive effect. *J Neurosci* 2012;32:12361–5. <http://dx.doi.org/10.1523/JNEUROSCI.5864-11.2012>.
- [3] Schwarzkopf DS, Silvanto J, Rees G. Stochastic resonance effects reveal the neural mechanisms of transcranial magnetic stimulation. *J Neurosci* 2011;31:3143–7. <http://dx.doi.org/10.1523/JNEUROSCI.4863-10.2011>.
- [4] Abrahamyan A, Clifford CWG, Arabzadeh E, Harris JA. Improving visual sensitivity with subthreshold transcranial magnetic stimulation. *J Neurosci* 2011;31:3290–4. <http://dx.doi.org/10.1523/JNEUROSCI.6256-10.2011>.
- [5] Abrahamyan A, Clifford CWG, Arabzadeh E, Harris JA. Low intensity TMS enhances perception of visual stimuli. *Brain Stimul* 2015;1–8. <http://dx.doi.org/10.1016/j.brs.2015.06.012>.
- [6] Walsh V, Ashbridge E, Cowey A. Cortical plasticity in perceptual learning demonstrated by transcranial magnetic stimulation. *Neuropsychologia* 1998;36:363–7.
- [7] Rossini PM, Barker AT, Berardelli A, Caramia MD, Caruso G, Cracco RQ, et al. Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an IFCN committee. *Electroencephalogr Clin Neurophysiol* 1994;91:79–92. [http://dx.doi.org/10.1016/0013-4694\(94\)90029-9](http://dx.doi.org/10.1016/0013-4694(94)90029-9).
- [8] Rothwell JC. Paired-pulse investigations of short-latency intracortical facilitation using TMS in humans. *Electroencephalogr Clin Neurophysiol* 1999;51:113–9.
- [9] Hanajima R, Wang R, Nakatani-Enomoto S, Hamada M, Terao Y, Furubayashi T, et al. Comparison of different methods for estimating motor threshold with transcranial magnetic stimulation. *Clin Neurophysiol* 2007;118:2120–2. <http://dx.doi.org/10.1016/j.clinph.2007.05.067>.
- [10] Antal A, Nitsche MA, Kincses TZ, Lampe C, Paulus W. No correlation between moving phosphene and motor thresholds: a transcranial magnetic stimulation study. *Neuroreport* 2004;15:297–302. <http://dx.doi.org/10.1097/01.wnr.0000099600.41403.ec>.
- [11] Marzi CA, Mancini F, Savazzi S. Interhemispheric transfer of phosphenes generated by occipital versus parietal transcranial magnetic stimulation. *Exp Brain Res* 2009;192:431–41. <http://dx.doi.org/10.1007/s00221-008-1496-4>.
- [12] Mazzi C, Mancini F, Savazzi S. Can IPS reach visual awareness without V1? Evidence from TMS in healthy subjects and hemianopic patients. *Neuropsychologia* 2014;64:134–44. <http://dx.doi.org/10.1016/j.neuropsychologia.2014.09.026>.
- [13] Bagattini C, Mazzi C, Savazzi S. Waves of awareness for occipital and parietal phosphenes perception. *Neuropsychologia* 2015;70:114–25. <http://dx.doi.org/10.1016/j.neuropsychologia.2015.02.021>.
- [14] Silvanto J. Transcranial magnetic stimulation and vision. third ed., vol. 116. Elsevier B.V.; 2013. <http://dx.doi.org/10.1016/B978-0-444-53497-2.00052-8>.
- [15] Marg E, Rudiak D. Phosphenes induced by magnetic stimulation over the occipital brain: description and probable site of stimulation. *Optom Vis Sci* 1994;71:301–11.
- [16] Deblicq C, Thompson B, Iacoboni M, Wu AD. Correlation between motor and phosphene thresholds. *A Transcranial Magn Stimul Study* 2008;670:662–70. <http://dx.doi.org/10.1002/hbm.20427>.
- [17] Boroojerdi B, Meister IG, Foltys H, Sparing R, Cohen LG, Töpper R. Visual and motor cortex excitability: a transcranial magnetic stimulation study. *Clin Neurophysiol* 2002;113:1501–4.
- [18] Tyrrell RA, Owens DA. A rapid technique to assess the resting states of the eyes and other threshold phenomena: the Modified Binary Search (MOBS). *Behav Res Methods, Instrum Comput* 1988;20:137–41. <http://dx.doi.org/10.3758/BF03203817>.
- [19] Anderson AJ, Johnson CA. Comparison of the ASA, MOBS, and ZEST threshold methods. *Vis Res* 2006;46:2403–11. <http://dx.doi.org/10.1016/j.visres.2006.01.018>.
- [20] Kammer T, Beck S, Erb M, Grodd W. The influence of current direction on phosphene thresholds evoked by transcranial magnetic stimulation. *Clin Neurophysiol* 2001;112:2015–21.
- [21] Abrahamyan A, Clifford CWG, Ruzzoli M, Phillips D, Arabzadeh E, Harris J a. Accurate and rapid estimation of phosphene thresholds (REPT). *PLoS One* 2011;6:19554. <http://dx.doi.org/10.1371/journal.pone.0022342>.
- [22] Rossi S, Hallett M, Rossini PM, Pascual-Leone A. Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clin Neurophysiol* 2009;120:2008–39. <http://dx.doi.org/10.1016/j.clinph.2009.08.016>.
- [23] Kammer T, Puls K, Erb M, Grodd W. Transcranial magnetic stimulation in the visual system. II. Characterization of induced phosphenes and scotomas. *Exp Brain Res* 2005;160:129–40. <http://dx.doi.org/10.1007/s00221-004-1992-0>.
- [24] Meyer B, Diehl R, Steinmetz H, Britton T, Benecke R. Magnetic stimuli applied over motor and visual cortex: influence of coil position and field polarity on motor responses, phosphenes, and eye movements. *Electroencephalogr Clin Neurophysiol* 1991;43:121–34.
- [25] Kammer T, Beck S. Phosphene thresholds evoked by transcranial magnetic stimulation are insensitive to short-lasting variations in ambient light. *Exp Brain Res* 2002;145:407–10. <http://dx.doi.org/10.1007/s00221-002-1160-3>.
- [26] Brainard DH. The psychophysics toolbox. *Spat Vis* 1997;10:433–6.
- [27] Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 1997;10:437–42.
- [28] Kleiner M, Brainard DH, Pelli DG. What's new in Psychtoolbox-3? *Percept 36 ECPV abstr suppl*. 2007.
- [29] Brückner S, Kammer T. High visual demand following theta burst stimulation modulates the effect on visual cortex excitability. *Front Hum Neurosci* 2015;9:591. <http://dx.doi.org/10.3389/fnhum.2015.00591>.
- [30] Brückner S, Kammer T. Modulation of visual cortex excitability by continuous theta burst stimulation depends on coil type. *PLoS One* 2016;11:1–14. <http://dx.doi.org/10.1371/journal.pone.0159743>.
- [31] Prins N, Kingdom FAA. Palamedes: Matlab routines for analyzing psychophysical data. 2009. <http://www.palamedestoolbox.org>.
- [32] Kontsevich LL, Tyler CW. Bayesian adaptive estimation of psychometric slope and threshold. *Vis Res* 1999;39:2729–37.



- [33] Wichmann FA, Hill NJ. The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept Psychophys* 2001;63:1293–313. <http://dx.doi.org/10.3758/BF03194544>.
- [34] Prins N. The psychometric function: the lapse rate revisited. *J Vis* 2012;12:25. <http://dx.doi.org/10.1167/12.6.25>.
- [35] Merabet LB, Theoret H, Pascual-Leone A. Transcranial magnetic stimulation as an investigative tool in the study of visual function. *Optom Vis Sci* 2003;80:356–68.
- [36] Terhune DB, Murray E, Near J, Stagg CJ, Cowey A, Cohen Kadosh R. Phosphene perception relates to visual cortex glutamate levels and covaries with atypical visuospatial awareness. *Cereb Cortex* 2015;25:4341–50. <http://dx.doi.org/10.1093/cercor/bhv015>.
- [37] Summerfield C, de Lange FP. Expectation in perceptual decision making: neural and computational mechanisms. *Nat Rev Neurosci* 2014;1–12. <http://dx.doi.org/10.1038/nrn3838>.
- [38] Engbert K, Wohlschläger A, Haggard P. Who is causing what? The sense of agency is relational and efferent-triggered. *Cognition* 2008;107:693–704. <http://dx.doi.org/10.1016/j.cognition.2007.07.021>.
- [39] Qi F, Wu AD, Schweighofer N. Fast estimation of transcranial magnetic stimulation motor threshold. *Brain Stimul* 2011;4:50–7. <http://dx.doi.org/10.1016/j.brs.2010.06.002>.
- [40] Goldsworthy MR, Hordacre B, Ridding MC. Minimum number of trials required for within- and between-session reliability of TMS measures of corticospinal excitability. *Neuroscience* 2016;320:205–9. <http://dx.doi.org/10.1016/j.neuroscience.2016.02.012>.
- [41] Komssi S, Kähkönen S, Ilmoniemi RJ. The effect of stimulus intensity on brain responses evoked by transcranial magnetic stimulation. *Hum Brain Mapp* 2004;21:154–64. <http://dx.doi.org/10.1002/hbm.10159>.