ASSESSMENT OF PSYCHOPHYSICAL METHODS FOR MEASURING THE CRITICAL FLICKER FUSION FREQUENCY IN YES/NO TASKS

Maydel Fernandez-Alonso[1], Christos Kaspiris-Rousellis[1], Will Innes[1,2], Jenny C. A. Read[1]

1 Institute of Neuroscience, Newcastle University, 2 Newcastle Eye Centre, Royal Victoria Infirmary

ABSTRACT

The Critical Flicker Fusion (CFF) threshold is widely used to evaluate the limits of visual temporal processing and has important practical applications in the field of display technologies. In this study, we evaluate the suitability of a novel adaptive psychophysical procedure for measuring CFF thresholds in a YES/NO task. Our results indicate that while the adaptive staircase procedure has high repeatability and is of shorter duration when compared to the more robust constant stimuli method, its accuracy is lower, giving thresholds that were significantly higher ($p<0.01$) by approximately 15Hz.

Index Terms **—** critical flicker fusion, visual temporal resolution, psychophysics methods, YES/NO tasks

INTRODUCTION

The Critical Flicker Fusion (CFF) is the lowest frequency at which an intermittent light appears to be completely steady to the average human observer. This threshold is used to characterize the limits of the temporal-resolving ability of the human visual system and it has important practical applications in multiple fields, including in visual display technologies. Consequently, the psychophysical methods by which the CFF is measured and the characterization of their accuracy, reliability and duration, are relevant to the field and could be useful for determining which method to use for different applications. Furthermore, exploring novel adaptive procedures might allow to reduce the time it takes to obtain a CFF measure without sacrificing the robustness of more traditional psychophysics methods.

The most commonly used procedures for measuring detection thresholds such as the CFF, are YES/NO and forced-choice tasks. In YES/NO tasks, either a target frequency or no stimulus is presented in each trial, and the observer must answer "yes" or "no" to indicate whether the signal was present or absent. The target frequencies are presented in half of the trials, while a supra-threshold frequency is used as no-stimulus in the other half, with the threshold usually being defined as the frequency at which the observer answers "yes" correctly in 50% of the trials. Forced-choice tasks on the other hand, usually present two or more alternatives in each trial, and the observer must choose the one containing the signal. Within the different forced-choice procedures, two-interval forced choice (2IFC) is particularly suitable for measuring the CFF since it allows to present all stimuli in the same retinal location, any variation of which would cause unintended changes to the CFF. In such tasks, two alternatives are presented sequentially in time, with the observer indicating which one contained the signal, and the threshold often being defined as the frequency at which they gave correct answers in 75% of the trials.

Previous studies have shown that YES/NO tasks have higher statistical efficiency than forced-choice tasks [1], in addition to several practical advantages over 2IFC, namely: the duration of each trial is halved since only one stimulus is being presented, stimuli presented in intervals can interact [2], and discrepancies can arise between naïve and experienced psychophysical observers in 2IFC, with thresholds estimates for the former showing poorer reliability and sensory determinacy [3]. For all these advantages, YES/NO tasks have one important drawback: they are criterion-dependent, that is, they are susceptible to the criterion adopted by the observer on how strong the internal signal must be before they give an affirmative answer. Hence, it is important to be able to measure the observer's bias in order to obtain an accurate measure of the CFF in YES/NO tasks.

This can be achieved through the method of constant stimuli (MCS), where a predefined set of stimulus levels, some above and some below the threshold, are presented in random order. From the proportion of "hits" (affirmative answers when the signal is present) and "false alarms" (affirmative answers when the signal is absent), it's possible to calculate the observer bias and the sensitivity index or d'. The latter is a measure of the internal strength of the signal that is independent of observer criterion and can be used to calculate an unbiased proportion correct, that is, the performance the observer would achieve if their criterion was neutral. While the MCS is one of the most robust in psychophysics, it can be time-consuming and fatiguing for the subject. A more commonly used alternative is adaptive procedures, which attempt to increase efficiency by presenting stimuli at levels where there is more information to be gained about the parameters of interest. However, for YES/NO tasks, there are not many adaptive methods available in which the false alarm rate is measured so that d' can be calculated and an unbiased threshold obtained [4]. Furthermore, the few methods available have not been evaluated for their suitability to measure CFF thresholds accurately and reliably.

One of such methods is the Quick Yes-No (qYN) algorithm by [5]. In their study they proposed a family of methods that combine elements of signal detection theory and Bayesian adaptive inference to broadly sample the psychometric function and estimate both sensitivity and decision parameters simultaneously. Their qYN method for YES/NO detection tasks directly estimates the decision criterion, as well as the steepness and threshold (defined as d'= 1) of the sensitivity function, while fixating its asymptote parameter (d'= 5). To efficiently select the stimulus sampling they use Bayesian adaptive inference to improve the gain of information over a multi-dimensional space of decision and sensitivity parameters. For this, a probability density function represents prior knowledge of these parameters, and during the experiment, a search algorithm evaluates a space of potential stimuli and their possible outcomes to select those that improve the information gained about the parameters of interest. The staircase terminates when a predefined number of trials have been completed.

The authors [5] validated the qYN method using a contrast grating detection task with both experiments in real observers and simulated psychophysical observers. Results from their simulations showed that, within 25 trials, the threshold estimates converged for all the false alarm rates tested with high accuracy (difference < 0.05 log unit from true threshold). Within the same number of trials, the variability was low ($\leq$ 0.1 log unit) for the smaller false alarm rates tested (2.5% and 10%), while for the highest false alarm rate (40%), the precision achieved was lower

(~0.15 log unit). However, when testing the qYN method in real observers and comparing the results to those obtained through the MCS, the authors found an increase in bias of estimated thresholds as the number of trials increased, with a ~0.10 log unit difference that persisted up to 100 trials. Likewise, threshold variability increased from 50 to 100 trials by ~0.05 log units. This decrease in accuracy and increase in variability was inconsistent with their simulations results, and the authors attributed it to the MCS providing only an approximation rather than the true thresholds of the observer.

In this study we aim to assess the suitability of the qYN method for measuring the CFF in real observers, by comparing the results to those obtained through the more classical MCS. A previous study [6] compared three methods for measuring the CFF: the method of limits, MCS and a staircase procedure, with results indicating greater agreement between last two. However, the task used in this study was 2IFC, the possible drawbacks of which have been discussed here. By using a novel adaptive procedure that delivers a criterion-free threshold estimate, the time required to measure the CFF under multiple conditions could be significantly reduced, while maintaining the higher statistical efficiency of YES/NO tasks.

METHODS

We developed a CFF testing system based on a microcontroller and a custom electronic circuit to drive a set of high-power Light Emitting Diodes (LEDs). The LEDs were placed behind a diffuser to create a uniform field. Using several LEDs allowed us to achieve a higher overall brightness and to produce stimuli with a wider range of luminance values through the use of neutral density filters. The latter was key in obtaining a varied range of threshold values, since the CFF is known to increase linearly with the logarithm of the luminous intensity [7].

The stimuli presentation and processing of observers' answers were controlled from custom written MATLAB code [8], which followed one of the two psychophysics methods tested respectively. For the qYN method, we developed our own implementation of [5]'s proposal. The results were processed using the Palamedes toolbox [9] in combination with custom written MATLAB code.

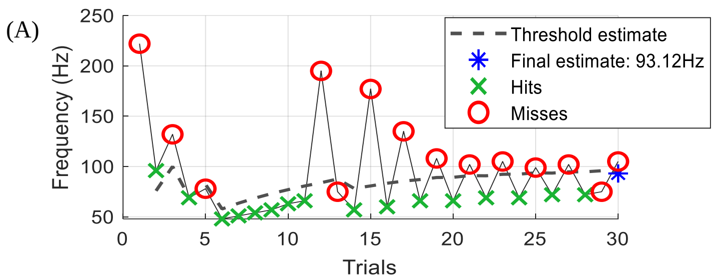
Five healthy participants (2 females, 3 males; 31 ± 6.6 years old) with normal uncorrected vision, including three of the authors, took part in the experiment. In each session, their left eye was occluded, and their right pupil was dilated using eye drops of Cyclopentolate at 1% to keep a constant retinal illuminance throughout each experimental block. Several exclusion criteria were used to minimize the risk of suffering negative side effects from the medication or from being exposed to a flickering light. All participants signed an informed consent form, and the study was reviewed and approved by the University Ethics Committee for Newcastle University.

The CFF was measured peripherally under 8 different luminance conditions, with the test stimuli placed at 35 degrees eccentricity on the right temporal visual field. A fixation cross was placed centrally in front of the right eye to aid fixation. At each luminance level, participants first completed a 30-trials run of the qYN staircase. The threshold estimate obtained was then used to determine the range of stimuli levels to be used for the MCS. Ten equally spaced periods of flicker were selected and presented 28 times each, together with 28 trials of a supra-threshold frequency, all in random order. To further explore the performance of the staircase procedure, one subject participated in six more sessions, where they did five repetitions of 60 trials of the qYN staircase, and five repetitions of the MCS, with seven different luminance levels being tested each time.

RESULTS

We collected data from five participants using the qYN method under eight different luminance conditions to obtain a varied range of CFF values. Concurrently, we collected data from the same participants under the same conditions using the MCS to obtain an estimate of "ground truth". Figure 1 presents an example of the results obtained from one run of the qYN staircase for one participant at one luminance level. Figure 1.A shows the frequencies of flicker presented and the correct responses (hits) and incorrect responses (misses) given by the observer, and in Figure 1.B, the proportion of correct responses is plotted as a function of the frequency presented, with the size of the symbols indicating how many trials of each frequency were used. The final threshold estimate obtained through this method is also plotted (blue dotted line), in addition to the corresponding threshold obtained through the MCS as a reference (red dotted line). Note that although the MCS threshold was determined from an independent set of measurements on this observer, it gives a better description of the frequency of seeing curve for the data obtained with the qYN staircase.

To analyze the results from the MCS, d' values and the criteria of the observer is calculated for each stimulus level from the proportion of false alarms and the proportion of hits at each stimuli level[1]. Once the d' values were obtained, an unbiased proportion correct was calculated[2], which indicates the performance the observer would achieve when adopting a neutral criterion. This unbiased proportion correct as a function of the period of flicker presented was then fitted with a Quick psychometric function through a maximum likelihood fitting procedure [10], and an estimate of the threshold, slope and lapse rate (i.e. the upper asymptote) was obtained. Note that, although



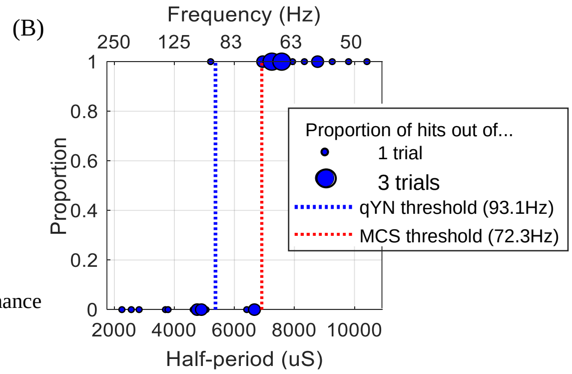


Figure 1. (A) Example of the qYN staircase procedure for one subject at one luminance level. (B) Frequency of seeing curve and final CFF estimate obtained, with the corresponding estimate from the MCS is included as reference.

[1] *The observer criteria (C) and sensitivity (d') were calculated as:* $C = -[z(pH) + z(pF)]/2$; $d' = z(pH) - z(pF)$; *where pH represents the proportion of hits and pF the proportion of false alarms* [10].

[2] *The unbiased proportion correct* (Pc_{max}) *was calculated as:* $Pc_{max} = \Phi(d'/2)$ [10].

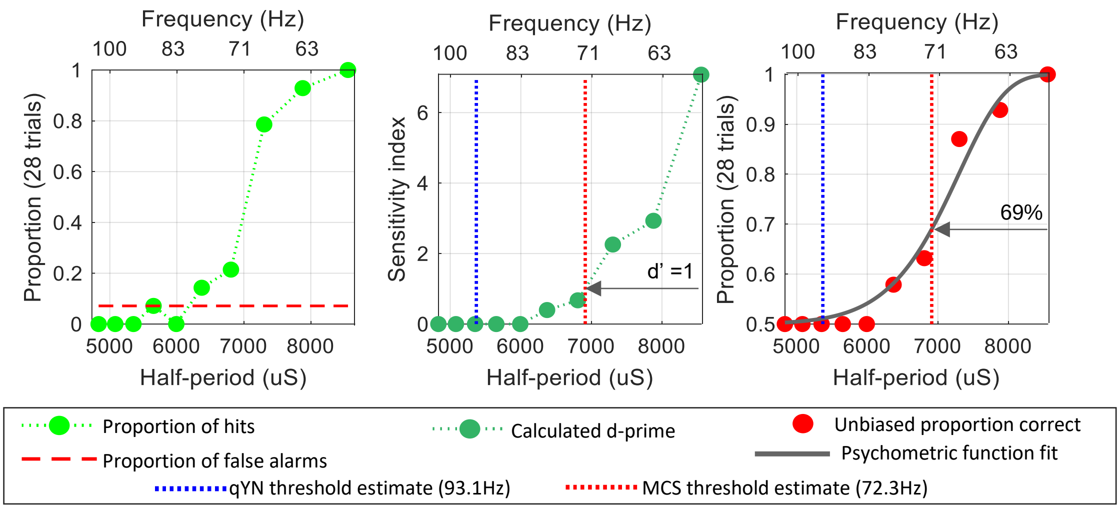


Figure 2. (A) Example of the results obtained through the MCS for one participant at one luminance level, showing the proportion of hits as a function of the frequency of flicker presented and the false alarm rate (proportion of "yes" responses for a 150Hz stimulus). (B) Sensitivity or d' values as a function of frequency, with the MCS threshold and the qYN threshold estimates plotted for reference. (C) Unbiased proportion correct calculated from the sensitivity values as a function of frequency and fitted psychometric function, with the resulting threshold estimate and the qYN estimate plotted.

the Quick psychometric function normally evaluates to 0.75 at threshold when the guess rate is 0.5, we modified the threshold so it would correspond to a proportion correct of 0.69. This value is equivalent to a d' of 1, the same definition of threshold used in the qYN method.

Figure 2 presents an example of the results from this procedure applied to the data obtained from the same participant and at the same luminance level as Figure 1. Figure 2.A shows the proportion of hits and false alarms as a function of stimulus intensity; Figure 2.B, the sensitivity values obtained; and Figure 2.C, the unbiased proportion correct with the fitted psychometric function and threshold estimate. The CFF values obtained through this method (red dotted line) and the qYN staircase procedure (blue dotted line), are plotted in the center and right graphs as references.

Method agreement. To assess method agreement, the correlation and differences between the 40 CFF values obtained from five participants through qYN and MCS are presented in Figure 3. We found a strong and statistically significant correlation between both methods ($r = 0.92$, $p = 2.08e{-17}$, $p<0.01$), however, as seen in Figure 3.A, most of the values fall above the identity line. In order to evaluate the differences between the CFF values obtained we performed a Bland-Altman analysis. The results are shown in Figure 3.B, where the difference between paired CFF measurements is plotted against their mean. We found a significant difference ($p = 5.45e{-17}$, $p<0.01$) between the CFF values obtained through the two methods, with the qYN values being on average 15.16Hz (±6.66Hz) higher than the MCS thresholds. This difference seems to be constant throughout range of CFF values tested. The limits of agreement indicate that at least 95% of CFF measurements obtained through the qYN method will be between 1.9 and 28Hz higher than those obtained through the MCS.

To further explore method agreement, Figures 4 and 5 present the mean and standard deviation of the difference between paired MCS and qYN thresholds as a function of trial number in the qYN adaptive run. This allow us to assess the performance of the staircase procedure in estimating the "true threshold" of the participant (or its closest approximation available) as the number of trials increases. Figure 4 presents the difference between paired MCS threshold estimates and qYN estimates taken at each trial, obtained for five participants that completed one staircase run of 30 trials at 8 different luminance levels. Figure 5 shows the difference between the average of four MCS threshold estimates and paired qYN estimates at each trial for one participant that completed 5 staircase runs of 60 trials at 7 luminance levels.

The results in Figure 4 show that, while the standard deviation of the differences decreases largely by 16 trials, indicating that on average, this number of trials suffices to approximate the true threshold; the mean of the differences shows an upward trend that persists up to 30 trials, with the final estimate showing the fixed bias discussed above. Furthermore, Figure 5 shows that even with a larger number of trials, the staircase shows similar performance, with the bias persisting up to 60 trials.

Finally, because the MCS allows us to measure the full psychometric and sensitivity functions, we evaluated these functions at the corresponding threshold estimate obtained through the qYN staircase. For the five participants evaluated initially, the psychometric function of unbiased proportion correct had a median value of 0.50 (±0.07) at the qYN threshold estimate,

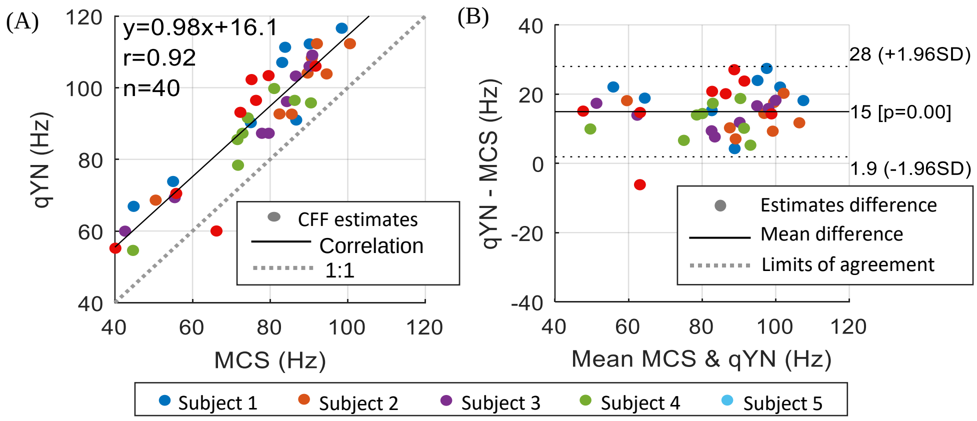


Figure 3. Correlation and agreement between the MCS and qYN methods. (A) The correlations between CFF values obtained by the two methods for all luminance levels for 5 subjects. (B) Bland–Altman plots of CFF values, where the difference between the estimates obtained from the two methods are plotted as a function of their mean.

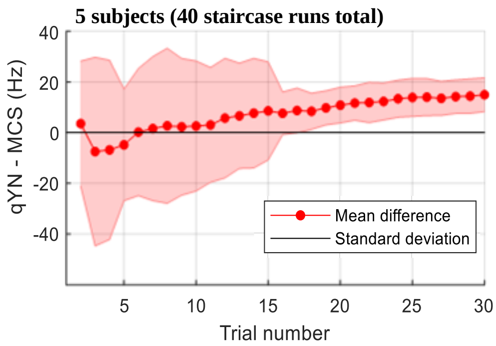


Figure 4. Mean and standard deviation of the difference between the CFF estimates from both methods as a function of trial number in the qYN staircase run, for 5 subjects at 8 luminance levels repeated once (n=40).

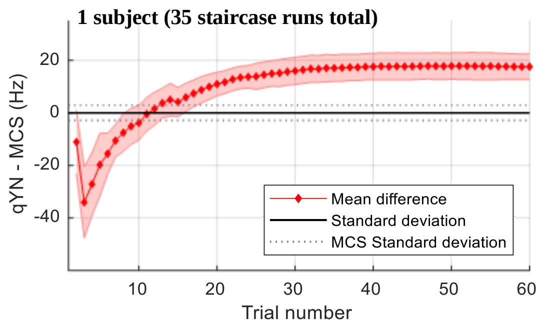


Figure 5. Mean and standard deviation of the difference between the CFF estimates from methods as a function of trial number in the qYN staircase run, for 1 subject at 7 luminance levels repeated 5 times each (n=35).

while the sensitivity index had a median value of 0.00 (±0.51), so this method was not succeeding in targeting a d' value of 1.

Method variability. To evaluate the variability of the qYN method when compared to the standard MCS, CFF measurements were obtained for one participant at 7 luminance levels, with the MCS being repeated four times and the qYN method five times at each luminance level. These results, presented in Table 1, allow us to evaluate how repeatable the estimates obtained through each method are, given the same observer and luminance conditions.

As seen, despite the differences in the mean of the CFF values obtained through both methods, their mean variability is comparable: 3.37Hz for the MCS and 3.78Hz for the qYN.

Table 1. Mean and standard deviations of CFF measurements obtained through qYN and MCS for one participant.

Condition	qYN		MCS	
	mean (Hz)	SD (Hz)	mean (Hz)	SD (Hz)
1	116.35	4.40	92.82	6.72
2	110.90	2.41	91.57	1.51
3	111.47	4.96	95.51	0.99
4	105.67	3.21	88.47	6.66
5	105.5	5.20	82.84	4.34
6	103.65	4.11	92.82	2.07
7	78.317	2.15	59.90	1.33
		3.78 (±1.2)		3.37 (±2.5)

DISCUSSION

In this study, we present a comprehensive comparison between two methods for measuring CFF thresholds in YES/NO tasks: the MCS and the qYN adaptive procedure developed by [5]. While the MCS is very robust and is considered the gold standard in many psychophysical studies [6], it can be time consuming to implement and fatiguing for the subject. Faster and more modern adaptive procedures are widely used in psychophysics nowadays, but these are mainly implemented with forced-choice tasks such as 2IFC, since such tasks are considered to be free from observer bias. YES/NO tasks can have several advantages over 2IFC, including a higher statistical efficiency and shorter trial duration, but not many methods have been available that allow to measure the observer decision criteria and obtain a bias-free estimate of the threshold. A recent study [5] developed such a procedure and validated it with simulations and psychophysical results obtained with a contrast grating detection task. Our purpose here is to evaluate the suitability of this qYN adaptive procedure for measuring CFF thresholds.

Our results demonstrated that the qYN method has a variability comparable to that of the standard MCS, showing similar repeatability over multiple sessions for one participant. High correlation but not good agreement was observed between both methods. A significant difference was found when comparing the CFF measurements obtained for five participants with 30-trials staircase runs, with the qYN method yielding values on average 15Hz higher than those obtained through the standard MCS. This bias was present across all luminance conditions tested and showed an increase as a function of trial number in the adaptive run. Further testing with one subject showed that this overestimation consistently persisted up to 60 trials, making it unlikely that a larger number of trials would give more accurate estimates.

These results are consistent with those found by the original study with real psychophysical observers, with their bias persisting up to 100 trials [5]; but inconsistent with the results from their simulated observers. A possible reason stated by the authors for this discrepancy is that the MCS gives only an approximation of the threshold of the participant, while their true sensitivity remains unknown. One advantage of the MCS is however, that it allows to sample the full psychometric and sensitivity functions when the stimuli levels are properly selected. Hence, to explore this further, we evaluated both functions at the qYN threshold estimate for all participants and at all luminance levels. Our results showed that, at the qYN threshold, the sensitivity index had a median value of 0 and the unbiased proportion correct a median value of 0.5, thus, failing to target a d' of 1 as set by the method (equivalent to a proportion correct of 0.69). Therefore, the discrepancy found between both methods cannot be explained solely by the uncertainty associated with the MCS, since the sensitivity of observers at the qYN threshold is very low and their performance is close to chance, meaning that this frequency of flicker is likely not visible to them.

We conclude that, while the qYN method has high repeatability and is of shorter duration when compared to the MCS, its accuracy is lower and might not suffice for applications where a precise CFF threshold is required. Future studies may explore whether modifying certain assumptions made in the development and implementation of this method could improve its suitability for obtaining accurate CFF measurements.

REFERENCES

[1] S. P. McKee, S. A. Klein and D. Y. Teller, "Statistical properties of forced-choice psychometric functions: Implications of probit analysis.," *Perception & psychophysics,* vol. 37, no. 4, pp. 286-298, 1985.

[2] Y. Yeshurun, M. Carrasco and L. T. Maloney, "Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model," *Vision research,* vol. 48, no. 17, pp. 1837-1851, 2008.

[3] F. Jäkel and F. A. Wichmann, "Spatial four-alternative forced-choice method is the preferred psychophysical method for naïve observers.," *Journal of Vision,* vol. 6, no. 11, pp. 13-14, 2006.

[4] S. A. Klein, "Measuring, estimating, and understanding the psychometric function: A commentary," *Perception & psychophysics,* vol. 63, no. 8, pp. 1421-1455, 2001.

[5] L. A. Lesmes, Z. L. Lu, J. Baek, N. Tran, B. A. Dosher and T. D. Albright, "Developing Bayesian adaptive methods for estimating sensitivity thresholds (d′) in Yes-No and forced-choice tasks," *Frontiers in Psychology,* vol. 6, pp. 1-24, 2015.

[6] A. Eisen-Enosh, N. Farah, Z. Burgansky-Eliash, U. Polat and Y. Mandel, "Evaluation of critical flicker-fusion frequency measurement methods for the investigation of visual temporal resolution," *Scientific reports*, vol. 7, no. 1, p. 15621, 2017.

[7] C. W. Tyler and R. D. Hamer, "Analysis of visual modulation sensitivity. IV. Validity of the Ferry–Porter law," *Journal of the Optical Society of America,* vol. 7, no. 4, pp. 743-757, 1990.

[8] *The Mathworks Inc.,* Massachusetts, USA.

[9] N. Prins and F. A. A. Kingdom, "Applying the Model-Comparison Approach to Test Specific Research Hypotheses in Psychophysical Research Using the Palamedes Toolbox. , .," *Frontiers in Psychology,* vol. 9, no. 1250, 2018.

[10] N. P. Frederick A.A. Kingdom, Psychophysics: A Practical Introduction, Second Edition, Academic Press, 2016.