



Spatial contrast sensitivity: effects of reliability, test–retest repeatability and sample size using the Metropsis software

Thiago Paiva Fernandes^{1,2} · Natalia Leandro de Almeida^{1,2} · Pamela D. Butler³ · Natanael Antonio Santos^{1,2}

Received: 20 September 2018 / Revised: 30 April 2019 / Accepted: 11 May 2019 / Published online: 5 June 2019
© The Royal College of Ophthalmologists 2019

Abstract

Purpose The goals of the study were to further assess contrast sensitivity to (1) investigate the existence of monocular vs. binocular differences; (2) observe possible differences between sample sizes; (3) investigate the effects of test–retest repeatability.

Methods Contrast sensitivity measurements were obtained by presenting eight horizontal sine-wave gratings (ranging from 0.2 to 20 cycles per degree). A three-up-one-down method was used to obtain thresholds with a criterion of 79.4% correct responses for each spatial frequency. The mean of 12 reversals was used for obtaining thresholds, and the two-alternative forced-choice method was used. Data were recorded in 55 naive observers from 20 to 45 years. All participants were free from identifiable ocular disease and had normal visual acuity.

Results We observed the absence of differences on CSF for both monocular and binocular observers, as well as the absence of differences between large sample sizes. The latter investigation revealed a high degree of repeatability across time (baseline to 6 months later) with the higher test–retest for low and high spatial frequencies.

Conclusions Our results indicated that spatial contrast sensitivity measurements were little influenced by variables, such as binocular summation, eye dominance, sample size and time using the Metropsis test. The results obtained here have significance for basic and clinical vision science.

Introduction

Several types of measurements have been established over decades to characterise visual processing in individuals, such as visual acuity and contrast sensitivity functions (CSF) [1]. The CSF defines the threshold between the visible and invisible, measuring thresholds for stimuli with a wide range of contrast intensity and spatial frequency. CSF assessment is a core tool in vision science, presenting a high degree of reliability and is potentially useful in clinical settings [2–5].

Contrast thresholds at different spatial frequencies establish the CSF [6]. This function can be measured by psychophysical criteria [7]. Specifically, contrast sensitivity assesses different visual channels, where each channel is sensitive to a narrow and discrete range of the visual scene spectrum [4]. Physiologically, these channels refer to neuronal populations involved in the selective processing of spatial frequencies [8]. These frequencies are often classified into low, medium and high spatial frequencies and provide a description of a visual scene. Contrast sensitivity is one foundation of higher-order visual processing, such as orientation discrimination and motion detection. Identification of an abnormal CSF may provide insights into the bases of additional visual disturbances [9, 10].

The CSF has been widely studied [10–15]. As stated, the CSF has essential significance for basic and clinical vision research. In view of this, the Metropsis software was developed to provide a comprehensive evaluation of the CSF in retinal, cortical and subcortical populations. The Metropsis vision testing has been used in several studies [12, 15–19]. Despite reliable results, the effects of eye dominance, sample size and repeatability of Metropsis have

✉ Natalia Leandro de Almeida
natalialalmeida@hotmail.com

¹ Department of Psychology, Federal University of Paraiba, Joao Pessoa, Brazil

² Perception, Neuroscience and Behaviour Laboratory, Federal University of Paraiba, Joao Pessoa, Brazil

³ Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, USA

not been studied. As its use continues to increase, repeatability, reproducibility and reliability are needed to ratify the findings.

Costa et al. [20] have already investigated such an approach using the Cambridge Colour Test (from the same manufacturer of the Metropsis, the Cambridge Research Systems). The results indicated the absence of eye dominance and binocular summation. It is important to note, therefore, that this study did not investigate sample size or repeatability in normal trichromats. Psychophysical studies can benefit from repeated measures design and the sample size from some studies is generally small, but it is interesting to look whether or not differences in sample sizes can influence the results or repeatability in Metropsis. In view of this, the purpose of this study was threefold: (i) to investigate the existence of monocular vs. binocular differences; (ii) observe possible differences between sample sizes; (iii) investigate the effects of test–retest repeatability.

Some measures for perceptual evaluation, such as the Lanthony D-15d, for example, may present a low degree of replication [21]. In view of this, we intend to answer some questions. Repeatability was estimated by the intraclass correlation coefficient (ICC) statistics to investigate the genuine consistency between measures. In addition, Bland–Altman (BA) was used to measure the accuracy of contrast sensitivity detection in Metropsis. The coefficient of repeatability (COR) is a useful metric for reliability, because it details the 95% confidence intervals for the fluctuation in test–retest data, and assesses a bias between the mean differences [22].

Materials and methods

General materials and methods

The subjects were required to have good ocular health, with no abnormalities on fundoscopic or optical coherence tomographic examination. All of the participants had normal or corrected-to-normal vision measured by the Snellen eye chart (at least 20/20). All participants were randomised to their groups. After the first study, all experiments were performed using both eyes.

In addition, participants underwent a series of neuropsychological measures during the experiment to minimise possible intervening variables. Measures such as the Stroop Test, Trail-Making Test A and B, Mini Mental State Exam and Hamilton Scale for Depression were used. All of the groups were matched for gender and level of education. The subjects participated in the study on a voluntary basis.

The exclusion criteria were <25 or >48 years old, current history of neurological disorder, cardiovascular disease, history of head trauma, history of contact with substances such

as solvents, current or previous drug or substance abuse and current use of medications that may affect visual processing and cognition. Female participants who used oral contraception were only tested outside their menstrual period to minimise confounds associated with hormonal differences.

Participants were students or staff members of the Federal University of Paraiba. The participants had no psychiatric disorders according to the Structured Clinical Interview for the DSM [23]. This study followed the ethical principles of the Declaration of Helsinki and was approved by the Committee of Ethics in Research of the Health Sciences Centre of Federal University da Paraiba. Written informed consent was obtained from all of the participants.

Participants

Data of 29 participants were excluded, because they reported or revealed any neurological or neuropsychiatric disorders during the experiments (from baseline, T0, to 6 months, T1). Naive participants participated in each study.

Study 1—assessment of the differences between monocular vs. binocular testing

Data of 20 observers of CSF measurements were included in the further analysis of the effects of monocular vs. binocular testing.

Ocular dominance was assessed by the hole-in-the-card test [24, 25]. A card with a small hole in the centre (of about 3.0-cm diameter hole) was presented to the observers. When they aligned the hole in the card, they were instructed to alternatively close the eyes or slowly draw the opening back to the head to determine which eye is viewing the object (this is the dominant eye). This test was performed twice to ensure reliability in the assessment of ocular dominance. This study was performed by ten monocular observers ($M = 32$ years; $SD = 7.73$ years) and 10 binocular observers ($M = 37$ years; $SD = 6.70$ years).

Study 2—does sample size really matter?

CSF measurements from 55 naive observers were included in the analysis of the influence of sample size. The participants were clustered in groups of 5 individuals ($M = 32.2$ years; $SD = 4.32$ years), 10 individuals ($M = 36.3$ years; $SD = 7.02$ years) and 40 individuals ($M = 35.2$ years; $SD = 8.14$ years). They performed this part using both eyes.

Study 3—repeatability of contrast sensitivity function using Metropsis

Twenty healthy observers ($M = 33.7$ years; $SD = 7.80$ years) performed the study using both eyes.

Cognitive measures

Stroop colour-word interference

The stroop colour-word interference test [26] was used to measure executive functions, such as attention, cognitive flexibility, inhibition and information processing speed. A series of colour words was presented. Participants were asked to name the colour of the words instead of reading the word. Conflict occurs when the colour of the word and the name are different. We used four colours (red, blue, yellow and green) in several combinations that were randomly displayed on a computer screen. The measure was the number of elements that were properly named. Fewer errors indicated better performance.

Trail-making test

This test was used to investigate cognitive operations, such as visual search, psychomotor speed, cognitive flexibility and sustained attention [27]. The participant was presented with a sheet of randomly placed circles and instructed to draw a line that connected numbers and letters in the correct ascending sequence. A maximum time limit of 300 s was adopted. A quicker reaction time indicated better performance.

Mini-mental state examination

This test was used to screen for possible cognitive impairment. The mini-mental state examination (MMSE) is used to detect changes in mental state and allows the observation of verbal learning and memory, verbal and spatial working memory and semantic memory. The maximum score is 30. A score below 25 suggests impairment [28].

Stimuli and apparatus

Stimuli were presented on a 19-inch LG CRT monitor with 1024 × 786 resolution and a 100-Hz refresh rate. Stimuli were generated using a VSG 2/5 video card (Cambridge Research Systems, Rochester, Kent, UK), which was run on a Precision T3500 computer with a W3530 graphics card. The average luminance was 40 cd/m². All of the procedures were performed in a room at 26 ± 1 °C. The walls of the room were covered in grey to better control luminance during the experiments. The luminance of the monitor and chromatic calibrations were performed using a ColCAL MKII photometer (Cambridge Research Systems, Rochester, UK).

The Metropsis software (Cambridge Research Systems Ltd., Rochester, UK) determined contrast sensitivity. The stimuli for the CSF were linear, vertically oriented and

sine-wave gratings with spatial frequencies of 0.2, 0.6, 1.0, 3.1, 6.0, 8.8, 13.2 and 20.0 cycles per degree (cpd). The stimuli consisted of equiluminant gratings with dimensions of 5° of visual angle and were presented on the monitor at 2.5° spatial offset from the central cross-shaped fixation point. An extensive description of Metropsis and of the stimuli used here can be found in several studies [12, 29].

Procedures

The procedures were performed in two stages. In the first stage, the participants were referred to our laboratory where we conducted the cognitive tasks. A specialist performed the neuropsychological tests. This procedure was performed in a quiet, comfortable and reserved room; the approximate time was 35–40 min for each participant. In the second stage, each participant performed the contrast sensitivity measurements. For all tests, the participants were encouraged to take breaks between each block. Each session lasted from 35 to 40 min and the participants were encouraged to take breaks between each block to avoid lack of motivation.

Prior to the start of the tests, instructions on the operation and tasks that individuals should perform were provided. The participants performed a short training with high-contrast stimuli, including each spatial frequency to familiarise participants with the procedure and to avoid misunderstanding. Accuracy over speed was emphasised. The Metropsis software incorporates a check on the validity of the data by using catch trials (suprathreshold stimuli) to detect and avoid random responding.

Measurements were performed at a distance of 150 cm from the computer monitor. The participants had to respond whether the grating was presented on either the left or right side of the computer screen (Fig. 1). A number of catch trials (commonly used in perception studies to investigate whether or not the participant grasped the task) were randomly intermixed with the test trials, in order to detect false-negative or false-positive responses. The psychometric function of Metropsis computed those answers and estimated thresholds without interference from the participant's guessing or learning effects. The order of the spatial frequencies was randomised within a session. A three-down one-up logarithmic staircase with dynamic steps was used to derive a contrast threshold, with a level of accuracy of target detection of 79.4% on a psychometric function [30]. Initially, the contrast values appeared at the suprathreshold level, for which we expected correct responses. Thus, after three consecutive correct responses, contrast decreased by 0.7 dB. After every incorrect response, contrast increased by 1.0 dB. Each stimulus had an exposure time of 600 ms. After responding, the next trial started after 300 ms. The session ended after 12 contrast reversals occurred. A higher CSF curve indicates better discrimination. An extensive

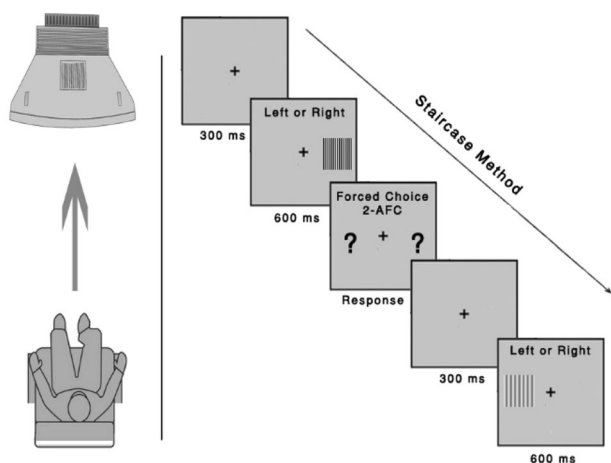


Fig. 1 Contrast sensitivity function task. The task was to identify, using a remote control response box, whether the gratings were presented either on the left or right side of the computer screen. Each stimulus had an exposure time of 600 ms, with an inter-trial interval of 300 ms. The Metropsis algorithm randomises spatial frequencies (low, medium and high) and contrast values

description of the procedure used in this study can be found in the literature [12, 15, 19].

Statistical analyses

The statistical analysis was performed using SPSS 23.0 software. The data from both groups presented a non-normal distribution; thus, nonparametric statistical tests were used to analyse the data for CSF. Despite the fact that Metropsis has no tolerance limits reported in the literature, we used the magnitude of the BA for repeatability.

For pairwise comparisons, the Mann–Whitney U test was used. Regarding CSF, each spatial frequency is considered, individually, as a dependent variable. However, in order to avoid inflation of type 1 error we performed Bonferroni correction. For categorical variables, the chi-squared test was used. Paired comparisons were made by the Wilcoxon signed-rank test. Nonparametric regression analyses (bootstrapping) were performed to evaluate relationships between cognitive performance and visual sensitivity between groups and participants. The bootstrapping regression used the general equation: $B_r = 1999 \equiv [A + B_1x_1 + \dots + B_kx_{ik}]$ and the residual for each observation was calculated based on $E_* = A_i - \hat{A}$. Bootstrapping is taken from estimated residuals to investigate an estimator [31]. The advantage of using bootstrapping in nonparametric regression is the possibility of using a smaller sample size, since this type of analysis considers mainly the residuals. For the separate analysis, the mean of 0.2–1.0 was classified as low-spatial frequencies (LSF), those of 3.1–5.0 were classified as medium spatial frequencies (MSF) and frequencies of 8.8–20.0 were defined as high spatial frequencies (HSF).

ICC statistics were used to measure the absolute agreement of the tasks across time. An ICC score of 1 denotes complete agreement, while an ICC score of 0 denotes no agreement. For reference, the following ranges can be used for ICC interpretation: ICC < 0.20 indicates poor agreement; 0.21–0.40 indicates fair agreement; 0.41–0.60 indicates moderate agreement; 0.61–0.80 indicates strong agreement; > 0.80 indicates almost perfect agreement [32]. Bland–Altman analysis included the coefficient of repeatability (COR), which is 1.96 times the standard deviation for the difference between the test and the retest scores (which are contrast measurements within the participant). The effect size (r) was estimated based on z -score conversion [33]. Effect sizes > 0.50 were considered medium-to-large effect sizes.

Results

Study 1—effects of monocular vs. binocular testing

Sample characteristics and cognitive assessment

Twenty observers participated in this study (binocular observers, $n = 10$; monocular observers, $n = 10$). The groups did not differ in age ($U = 29$, $p = 0.110$), level of education ($U = 49.50$, $p = 0.969$) or the ratio of males to females ($\chi^2_1 = 1.288$, $p = 0.525$).

With regard to the cognitive assessments, the Mann–Whitney U test revealed no statistically significant difference for baseline measurements in Trail-Making A task scores ($U = 48$, $p = 0.880$) and in Trail-Making B task scores ($U = 43$, $p = 0.597$) between groups. When comparing the measurements across time (T0–T1), there were no significant differences within ($Z = -1.274$, $p = 0.203$; binocular observers), ($Z = -0.764$, $p = 0.445$; monocular observers) and between groups ($Z = -0.261$, $p = 0.794$) for Trail-Making A; and within ($Z = -0.153$, $p = 0.878$; binocular observers), ($Z = -0.663$, $p = 0.508$; monocular observers) and between groups ($Z = -0.411$, $p = 0.681$) for Trail-Making B test, respectively. The same pattern was observed in the Stroop test for congruent ($U = 58$, $p = 0.670$) and incongruent stimuli ($U = 69.76$, $p = 0.751$). When comparing the measurements across time (T0–T1), there were no significant differences within ($Z = -0.474$, $p = 0.635$; binocular observers), ($Z = -0.919$, $p = 0.358$; monocular observers) and between groups ($Z = -0.886$, $p = 0.376$) for congruent stimuli. There were no differences within ($Z = -0.459$, $p = 0.649$; binocular observers), ($Z = -0.561$, $p = 0.575$; monocular observers) and between groups ($Z = -0.093$, $p = 0.926$) for incongruent stimuli, respectively. There were no differences between Hamilton Rating Scale for Depression ($p > 0.05$) and mini mental state exam ($p > 0.05$) for both T0 and T1.

Contrast sensitivity function

The results of the CSF measurements are shown in Fig. 2. There was no statistically significant difference between the groups, $p = 0.685$; Roy's largest root = 0.268. There was not a significant effect when age and level of education were added as covariates (all p values > 0.05).

Bootstrapping regression analysis

There were no significant differences across time for LSF ($p > 0.05$), MSF ($p > 0.05$) and HSF ($p > 0.05$). Thus, a global average (including T0 and T1) was organised for LSF, MSF and HSFs. We took this decision to investigate the existence of predictive effects of cognitive performance on visual measurements between groups. Thus, nonparametric bootstrapping regression analysis was conducted. No significant predictors of visual processing were found in the Stroop Colour-Word Interference, Trail-Making Test A and B and MMSE for LSF [F [5, 14] = 0.336, $p = 0.883$ and adjusted $R^2 = .030$; $\beta = 0.222$, $t = 0.862$ and $p = 0.402$], MSF [F [5, 14] = 0.685, $p = 0.642$ and adjusted $R^2 = 0.090$; $\beta = 0.040$, $t = 0.152$ and $p = 0.881$] and HSF [F [5, 14] = 0.751, $p = 0.599$ and adjusted $R^2 = 0.070$; $\beta = 0.095$, $t = 0.381$ and $p = 0.709$].

Study 2—influence of sample size on contrast detection

Sample characteristics

There were no statistically significant differences between age [χ^2 (2) = 4.365, $p = 0.113$], level of education [χ^2 (2) = 1.444, $p = 0.486$] or gender χ^2 (2) = 0.682, $p = 0.771$.

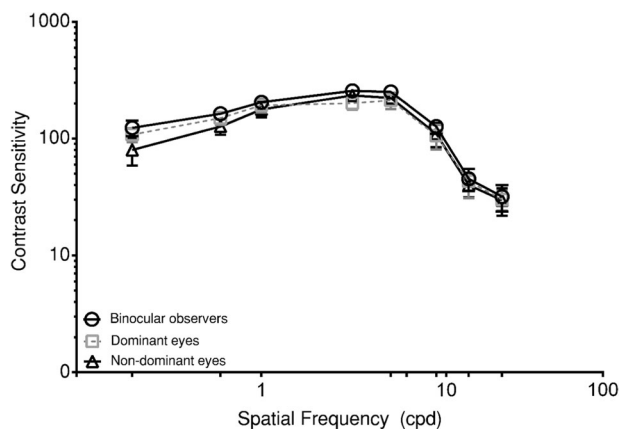


Fig. 2 Contrast sensitivity curves as a function of spatial frequency (cpd) in healthy observers. Each data point represents the mean sensitivity (reciprocal of contrast threshold). Error bars represent the standard deviation (SD) of the mean sensitivity based on 1000 bootstrapping resampling. Binocular observers, $n = 10$; monocular observers, $n = 10$

Contrast sensitivity function

The Kruskal–Wallis H test revealed no statistically significant differences between groups for spatial frequencies of 0.2 (χ^2 (2) = 5.176, $p = 0.075$), 0.6 (χ^2 (2) = 1.955, $p = 0.370$), 1.0 (χ^2 (2) = 5.287, $p = 0.071$), 3.1 (χ^2 (2) = 3.446, $p = 0.179$), 5.0 (χ^2 (2) = 5.169, $p = 0.075$), 8.8 (χ^2 (2) = 0.744, $p = 0.689$), 13.2 (χ^2 (2) = 1.692, $p = 0.429$) and 20.0 (χ^2 (2) = 4.588, $p = 0.101$). However, further analyses revealed that the group with five observers had a slight reduction in CSF when compared with the group with 10 observers ($p = 0.020$; overall reduction of 2.82) and 40 observers ($p = 0.001$; overall reduction of 2.44) mainly for LSF. There were not significant differences between the group with 10 and 40 observers ($p > 0.05$). The CSF measurements are presented in Fig. 3.

Study 3—repeatability of contrast sensitivity function using Metropsis

Sample characteristics and cognitive assessment

There were no statistically significant differences between age ($p = 0.221$), level of education ($p = 0.175$) or gender χ^2 (2) = 0.686, $p = 0.710$.

The cognitive measurements of this study were considered T0 and T1. A Wilcoxon signed-rank test revealed that after six months, the measurements did not elicit a statistically significant change in the Trail-Making A test scores ($Z = -1.480$, $p = 0.193$); Trail-Making B test scores ($Z = -0.281$, $p = 0.831$); stroop congruent ($Z = -1.831$, $p = 0.087$); stroop incongruent ($Z = -0.215$, $p = 0.830$); MMSE ($Z = -1.796$, $p = 0.072$); and Hamilton depression ($Z = -0.317$, $p = 0.751$) tests.

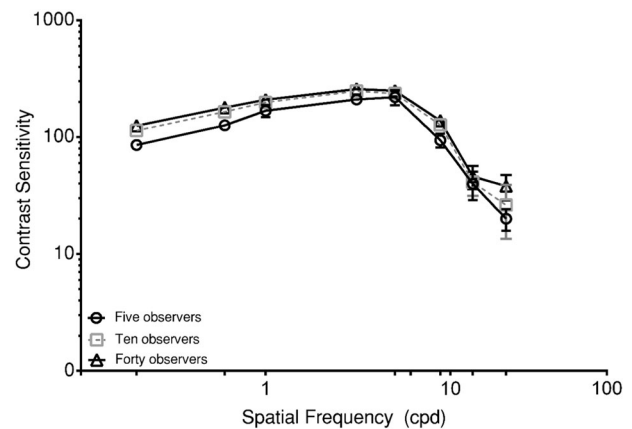


Fig. 3 Contrast sensitivity curves as a function of spatial frequency (cpd) in different sample sizes. Each data point represents the mean sensitivity (reciprocal of contrast threshold). Error bars represent the standard deviation (SD) of the mean sensitivity based on 1000 bootstrapping resampling

Whilst the data from the first study did not indicate predictive effects of cognitive variables in the measurement of visual contrast sensitivity, we chose not to perform regression analysis in this study.

Contrast sensitivity function

Comparatively with the first study, the spatial frequencies were grouped into a low spatial frequency (LSF), medium spatial frequency (MSF) and high spatial frequency (HSF). Intraclass correlation coefficient scores for all spatial frequencies ranged from 0.60 to 0.80. The ICC values for LSF, MSF and HSF were .63 (T0 vs. T1), .80 (T0 vs. T1) and .71 (T0 vs. T1), respectively. Bland–Altman plots are presented in pairs and exemplified in Fig. 4 for LSF, MSF and HSFs

Discussion

The main objective of this study was to evaluate the specificity of spatial contrast sensitivity using Metropsis as a tool for measuring the contrast detection thresholds in healthy individuals. Our results confirmed our initial hypothesis and showed the absence of binocular summation, absence of sample size effects (>5 individuals) and that CSF measurements are reliable across time.

Binocular summation

Binocular summation is a factor that can indicate the superiority of binocular over monocular performance in some visual tasks and was the subject of several psychophysical studies [20, 34, 35]. The results were often mixed. Nevertheless, they pointed to the conclusion that under photopic conditions, differences would not exist between binocular and monocular observers [35–37]. The lack of summation indicates that spatial filtering (mainly in detection tasks) does not indicate superiority of binocular over monocular performance [38]. One may argue that at low spatial frequencies (LSFs), the binocular summation can be observed because of the differences in orientation and the smaller size of these stimuli [39, 40]. In fact, we observed a slight difference between LSFs in the nondominant monocular observers (Fig. 2). However, we used a photopic condition, and we cannot extrapolate our data to other luminance conditions (e.g., mesopic). In addition, the assumption of binocular summation was revisited and counter-argued in a recent study [41]. As stated by Costa et al. [20], lower binocular thresholds can be interpreted as binocular summation, due to binocular fusion (the fusion of the images) mechanisms in the primary visual cortex [42]. Some recent results about the effects of ocular dominance on contrast sensitivity revealed that under the same

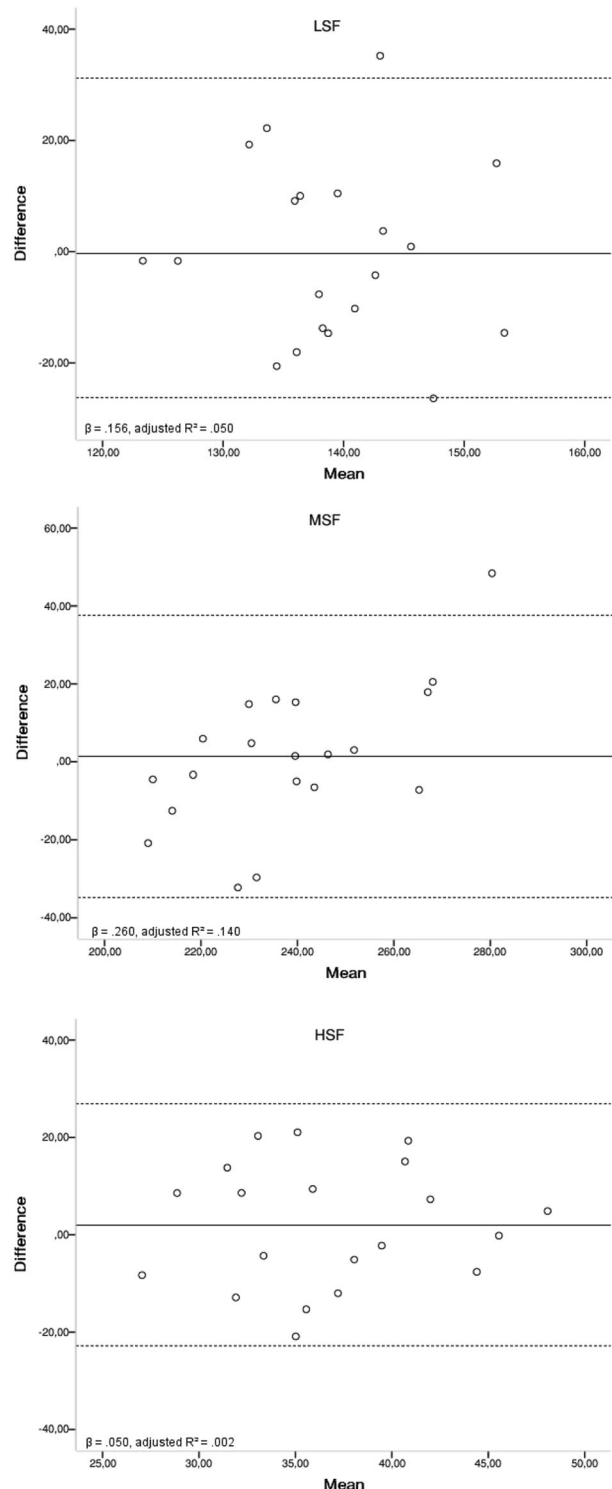


Fig. 4 Bland–Altman plots for contrast sensitivity measurements

conditions that we used here, the CSF of dominant vs. nondominant eyes was similar at all spatial frequencies [41]. In our study, we also did not observe differences for dominant vs. nondominant eyes (Fig. 2). The psychophysical approach used here may be limited by peripheral or

central noise, but our data suggested that ocular dominance and/or monocular rivalry had limited effects on binocular performance for a wide range of spatial frequencies used here. However, this should be further examined in future studies.

Sample size

As a de facto standard in psychophysical studies, the use of reduced sample size provided essential and strong insights about the visual processing [43–47]. These classic studies investigated contrast detection using small sample sizes and the results can still be observed in the literature, indicating that maybe the sample size for psychophysical studies was not a matter of large debate due to its design [3, 5, 15, 48]. This important aspect of psychophysical studies may be related to the repeated measures, when the same observer (or observers) perform several trials to obtain a threshold. The CSF measurements in computerised tasks also use this same paradigm [5], but a question remains: Does the sample size matter? This is quite common when one intends to obtain or present CSF data. However, the use of sample size in CSF should be an estimator, not a definer of the population. The influence of the sample size on modern contrast detection techniques is underreported. Here, we investigated if the sample size influenced the results, and we observed that there was a slight difference when the sample size was too small ($n = 5$; Fig. 3). However, when we increased the sample ($n = 10$), the results were not different between 10 and 40 participants. Thus, for the Metropsis test, the use of >5 participants is fundamental. Although the differences between the curves for 5–10 observers and 5–40 observers were narrow, the results can be related to any effect that we were not able to report here. Further studies using the Metropsis should investigate a wide range of observers' thresholds in a cluster-based analysis (e.g., 3, 5, 7, 10 and 15 observers).

Repeatability

The test–retest for contrast detection is well defined in the literature [5, 49–51]. Several studies reported that many tests presented interesting COR and ICC values for contrast test–retest. However, as stated in the 'Introduction' section, there were no reports of repeatability of the Metropsis test. The average COR for Metropsis' test–rest (0.71) compares favourably with previous studies, such as CSV-1000 (0.19), Pelli–Robson contrast test (.18), Vistech FDT (0.36) and Miller–Nadler (0.36) [5, 52]. The Metropsis' CORs for test–retest were higher than those previously reported for other devices. It seems that the psychometric function, the staircase method—or another aspect of Metropsis—differ in this software from the others. It is

important to note that the Metropsis algorithm incorporates a staircase with 79.4% of criteria, used a forced-choice alternative and this can serve as a matter of debate when comparing with the other tests. Although we can only speculate why the Metropsis had a higher COR when compared with other tests, there are some limitations about our study. We used only photopic conditions in a limited range of spatial frequencies (some of them should be properly investigated in scotopic conditions). We should mention here that all measurements occurred between morning and afternoon hours. Unfortunately, however, we cannot observe the relationship between measure time and visual performance, which serves as insights for further studies. In addition, for a better comprehension of test–retest, further studies using Metropsis should use different luminance conditions combined with different approaches (e.g., imaging).

Overall, the data presented here showed that the inverted U-shaped curve seemed similar for both Study 1 and Study 2. In Study 3, the ICC data confirmed repeatability values of CSF, as can be seen in the BA plots and COR. This emphasises the idea that the use of Metropsis has the potential of being a non-invasively diagnosing tool for clinical practice, being an alternative or adjuvant to the existing tools, providing a thorough investigation of the initial visual processing in quick time (about 30 min). These results suggest that Metropsis can be used to characterise the visual processing of healthy individuals and those affected by neuropsychiatric disorders.

Summary

What was known before

- The contrast sensitivity function has essential significance for basic and clinical vision science. Despite reliable results, the repeatability of Metropsis was not extensively studied. Our study investigated reproducibility, repeatability and reliability of spatial contrast sensitivity.

What this study adds

- Our results indicated that spatial contrast sensitivity measurements were little influenced by the variables, such as binocular summation, time and sample size using the Metropsis.

Author contributions TM conceived and designed the experiment, and performed the critical review of the paper. NL helped draft the paper, collected data and performed statistical analysis. PB performed the critical review of this paper and helped in the revised version. NA was

responsible for the direction, guidance and critical review of this paper. Both authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors did not receive (or will receive) any benefit from the Cambridge Research Systems.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Woods RL, Wood JM. The role of contrast sensitivity charts and contrast letter charts in clinical practice. *Clin Exp Optom*. 1995;78:43–57.
2. Dorr M, Elze T, Wang H, Lu Z-L, Bex PJ, Lesmes LA. New precision metrics for contrast sensitivity testing. *IEEE J Biomed Health Inform*. 2018;22:919–25.
3. Hou F, Lesmes LA, Kim W, Gu H, Pitt MA, Myung JI, et al. Evaluating the performance of the quick CSF method in detecting contrast sensitivity function changes. *J Vis*. 2016;16:18.
4. Pelli DG, Bex P. Measuring contrast sensitivity. *Vis Res*. 2013;90:10–4.
5. Pomerance GN, Evans DW. Test-retest reliability of the CSV-1000 contrast test and its relationship to glaucoma therapy. *Invest Ophthalmol Vis Sci*. 1994;35:3357–61.
6. Shapley RM, Lam DM-K. Contrast sensitivity. London: MIT Press; 1993. p. 370.
7. De Valois RL, Morgan H, Snodderly DM. Psychophysical studies of monkey Vision-III. Spatial luminance contrast sensitivity tests of macaque and human observers. *Vis Res*. 1974;14:75–81.
8. Owsley C. Contrast sensitivity. *Ophthalmol Clin N Am*. 2003;16:171–7.
9. Silverstein SM. Visual perception disturbances in schizophrenia: a unified model. *Neb Symp Motiv Neb Symp Motiv*. 2016;63:77–132.
10. Fernandes TP, Shaqiri A, Brand A, Nogueira RL, Herzog MH, Roinishvili M, et al. Schizophrenia patients using atypical medication perform better in visual tasks than patients using typical medication. *Psychiatry Res*. 2019;275:31–8.
11. Bulens C, Meerwaldt JD, van der Wildt GJ, Keemink CJ. Visual contrast sensitivity in drug-induced Parkinsonism. *J Neurol Neurosurg Psychiatry*. 1989;52:341–5.
12. Fernandes TM, de P, Almeida NL, de, Santos NAdos. Effects of smoking and smoking abstinence on spatial vision in chronic heavy smokers. *Sci Rep*. 2017;7:1690.
13. Jindra LF, Zemon V. Contrast sensitivity testing: a more complete assessment of vision. *J Cataract Refract Surg*. 1989;15:141–8.
14. Kéri S, Antal A, Szekeres G, Benedek G, Janka Z. Spatiotemporal visual processing in schizophrenia. *J Neuropsychiatry Clin Neurosci*. 2002;14:190–6.
15. Fernandes TMP, Andrade MJO de, Santana JB, Nogueira RMTBL, Santos NA dos. Tobacco use decreases visual sensitivity in schizophrenia. *Front Psychol*. 2018;288:1–13.
16. Andrade LCO, Souza GS, Lacerda EMC, Nazima MT, Rodrigues AR, Otero LM, et al. Influence of retinopathy on the achromatic and chromatic vision of patients with type 2 diabetes. *BMC Ophthalmol*. 2014;14:104.
17. Andrade MJO, Silva JA, Santos NA. Influência do Cronotipo e do Horário da Medida na Sensib ao Contraste Visual. 2015; 28:522–31.
18. Fernandes TMP, Silverstein SM, Butler PD, Kéri S, Santos LG, Nogueira RL, et al. Color vision impairments in schizophrenia and the role of antipsychotic medication type. *Schizophr Res*. 2019; 204:162–70.
19. Fernandes TP, Silverstein SM, Almeida NL, Santos NA. Visual impairments in tobacco use disorder. *Psychiatry Res*. 2018; 271:60–7.
20. Costa MF, Ventura DF, Perazzolo F, Murakoshi M, Silveira LC, de L. Absence of binocular summation, eye dominance, and learning effects in color discrimination. *Vis Neurosci*. 2006;23:461–9.
21. Good GW, Schepler A, Nichols JJ. The reliability of the lanthony desaturated D-15 test. *Optom Vis Sci Publ Am Acad Optom*. 2005;82:1054–9.
22. Vaz S, Falkmer T, Passmore AE, Parsons R, Andreou P. The case for using the repeatability coefficient when calculating test–retest reliability. *PLOS ONE*. 2013;8:e73990.
23. American Psychiatric Association. Structured Clinical Interview for DSM-5 (SCID-5); 2015.
24. Ehrenstein WH, Arnold-Schulz-Gahmen BE, Jaschinski W. Eye preference within the context of binocular functions. *Graefes Arch Clin Exp Ophthalmol*. 2005;243:926–32.
25. Mapp AP, Ono H, Barbeito R. What does the dominant eye dominate? A brief and somewhat contentious review. *Percept Psychophys*. 2003;65:310–7.
26. Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psychol*. 1935;18:643–62.
27. Tombaugh TN. Trail making test A and B: normative data stratified by age and education. *Arch Clin Neuropsychol*. 2004;19:203–14.
28. Folstein MF, Folstein SE, McHugh PR. Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189–98.
29. Fernandes TM, de P, Souza RM, da Ce, Santos NAdos. Visual function alterations in epilepsy secondary to migraine with aura: a case report. *Psychol Neurosci*. 2018;11:86–94.
30. Levitt H. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*. 1971;49(Suppl 2):467.
31. Härdle W, Bowman AW. Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *J Am Stat Assoc*. 1988;83:102–10.
32. Montgomery AA, Graham A, Evans PH, Fahey T. Inter-rater agreement in the scoring of abstracts submitted to a primary care research conference. *BMC Health Serv Res*. 26 de 2002;2:8.
33. Field A. Discovering statistics using IBM SPSS statistics. London: SAGE; 2013. p. 953.
34. Castro JJ, Soler M, Ortiz C, Jiménez JR, Anera RG. Binocular summation and visual function with induced anisocoria and monovision. *Biomed Opt Express*. 2016;7:4250–62.
35. Von Grünau M. Binocular summation and the binocularity of cat visual cortex. *Vis Res*. 1979;19:813–6.
36. Johansson J, Pansell T, Ygge J, Seimyr GÖ. The effect of contrast on monocular versus binocular reading performance. *J Vis*. 2014;14:8.
37. Lesmes LA, Kwon M, Lu Z-L, Dorr M, Miller A, Hunter DG, et al. Monocular and binocular contrast sensitivity functions as clinical outcomes in amblyopia. *Invest Ophthalmol Vis Sci*. 2014;55:797.
38. Frisén L, Lindblom B. Binocular summation in humans: evidence for a hierarchic model. *J Physiol*. 1988;402:773–82.
39. Pardhan S, Gilchrist J. The importance of measuring binocular contrast sensitivity in unilateral cataract. *Eye Lond Engl*. 1991;5 (Pt 1):31–5.

40. Schneck ME, Haegerstrom-Portnoy G, Lott LA, Brabyn JA. Binocular summation and inhibition for low contrast among elders. *Invest Ophthalmol Vis Sci.* 2007;48:5496.
41. Pekel G, Alagöz N, Pekel E, Alagöz C, Yılmaz ÖF. Effects of ocular dominance on contrast sensitivity in middle-aged people international scholarly research notices; 2014: 2014. p. 903084.
42. Cumming BG, Parker AJ. Binocular mechanisms for detecting motion-in-depth. *Vis Res.* 1994;34:483–95.
43. Campbell FW, Maffei L. Contrast and spatial frequency. *Sci Am.* 1974;231:106–14.
44. Hubel DH. Effects of deprivation on the visual cortex of cat and monkey. *Harvey Lect.* 1978;72:1–51.
45. Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. *J Physiol.* 1968;195:215–43.
46. Livingstone M, Hubel D. Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science.* 1988;240:740–9.
47. Movshon JA, Kiorpes L. Analysis of the development of spatial contrast sensitivity in monkey and human infants. *J Opt Soc Am A.* 1988;5:2166–72.
48. Cadenhead KS, Dobkins K, McGovern J, Shafer K. Schizophrenia spectrum participants have reduced visual contrast sensitivity to chromatic (red/green) and luminance (light/dark) stimuli: new insights into information processing, visual channel function, and antipsychotic effects. *Front Psychol.* 2013;4:535.
49. Haymes SA, Roberts KF, Cruess AF, Nicoleta MT, LeBlanc RP, Ramsey MS, et al. The letter contrast sensitivity test: clinical evaluation of a new design. *Invest Ophthalmol Vis Sci.* 2006;47:2739–45.
50. Kelly S, Pang Y, Engs C, Foley L, Salami N, Sexton A. Test-retest repeatability for contrast sensitivity in children and young adults. *Invest Ophthalmol Vis Sci.* 2011;52:1895.
51. Simpson TL, Regan D. Test-retest variability and correlations between tests of texture processing, motion processing, visual acuity, and contrast sensitivity. *Optom Vis Sci Publ Am Acad Optom.* 1995;72:11–6.
52. Elliott DB, Sanderson K, Conkey A. The reliability of the Pelli-Robson contrast sensitivity chart. *Ophthalmic Physiol Opt J Br Coll Ophthalmic Opt Optom.* 1990;10:21–4.