UC Berkeley
UC Berkeley Previously Published Works

Title

Rapid measurement and machine learning classification of colour vision deficiency.

Permalink

Journal

Authors

He, Jingyi
Bex, Peter
Skerswetat, Jan

Publication Date

DOI

Rapid measurement and machine learning classification of colour vision deficiency

Jingyi He[1], **Peter J. Bex**[1,*], **Jan Skerswetat**[1]

[1]Department of Psychology, Northeastern University, Boston, Massachusetts. USA

Abstract

Colour vision deficiencies (CVDs) indicate potential genetic variations and can be important biomarkers of acquired impairment in many neuro-ophthalmic diseases. However, CVDs are typically measured with tests which possess high sensitivity for detecting the presence of a CVD but do not quantify its type or severity. In this study, we introduce FInD (Foraging Interactive D-prime); a novel computer-based, generalizable, rapid, self-administered vision assessment tool and apply it to colour vision testing. This signal detection theory-based adaptive paradigm computed test stimulus intensity from d-prime analysis. Stimuli were chromatic gaussian blobs in dynamic luminance noise, and participants clicked on cells that contained chromatic blobs (detection) or blob pairs of differing colours (discrimination). Sensitivity and repeatability of FInD colour tasks were compared against the Hardy-Rand-Rittler and the Farnsworth-Munsell 100 hue tests in 19 colour-normal and 18 inherited colour-atypical, age-matched observers. Rayleigh colour match was also completed. Detection and discrimination thresholds were higher for atypical than for typical observers, with selective threshold elevations corresponding to unique CVD types. Classifications of CVD type and severity via unsupervised machine learning confirmed functional subtypes. FInD tasks reliably detect inherited CVDs, and may serve as valuable tools in basic and clinical colour vision science.

Keywords

colour detection; colour discrimination; colour vision deficiency; cone-isolating directions; K-means clustering; unsupervised machine learning; vision diagnostics

Introduction

Conventional phenotypical categories for inherited colour vision deficiency (CVD) are anomalous trichromats (AT), dichromats and monochromacy, with mild to strong colour vision defects, respectively. They can be further referred to as protan, deutan or tritan types, with L-, M- and S-cone relevant deficiencies, respectively.[1] Identification and diagnosis of CVDs are critical in many respects. In clinical applications, abnormal colour vision may reveal a hereditary CVD, or an acquired CVD which could signify neural pathway or

systemic disease.[2] However, widely employed colour vision tests are not ideal for detecting and monitoring acquired CVD progression or remediation, owing to their insensitivity, long testing time, difficulty of administration and interpretation of their results.

Rayleigh colour matches conducted with an anomaloscope have been considered a gold standard for the precise diagnosis of red-green CVD.[3] An anomaloscope enables red-green Rayleigh matching and blue-green Moreland matching tests, in which the observer adjusts the mixed light side of the bipartite field to match the light in the reference side.[4] However, anomaloscopes are expensive and a full examination of the colour matches requires extensive instruction, expert administration and exhaustive testing time. In addition, matching ranges of an extreme anomalous trichromat can be indistinguishable from those of dichromats. Pseudoisochromatic plates are capable of rapid screening and classification of CVD. Ishihara plates contain pseudoisochromatic numbers and curved lines, as well as vanishing plates with stimuli seen by only CVD patients. Hardy-Rand-Rittler (HRR) plates[5] exceed Ishihara plates, in that they contain universal colour symbols instead of Arabic numbers, have a three-step severity scale for each CVD subtype, use a two-step psychometric protocol (i.e., what symbol do you see in which quadrant) to reduce the probability of guessing and are able to classify tritan defects.[6] However, printed pseudoisochromatic plates have a number of shortcomings. They require a trained clinician to administer the test, and are generally insensitive to detect subtle changes of colour detection due to the absence of a severity scale (Ishihara), a coarse severity scale (HRR) and the lack of cone-isolated colours. The Farnsworth-Munsell 100 hue test (FM100) provides a relatively complete colour discrimination measurement for Protans, Deutans and Tritans, by asking participants to arrange 85 caps according to their colour, but the task is extremely time-consuming, as is the analysis of FM100 error scores. D-15, an abridged version of FM100, is able to classify CVD types rapidly. However, interpretation of the error scores of both tests can be vague. CVD patients can also rely on luminance cues to pass the arrangement task or improve their scores.[7]

Attempts to improve the performance of traditional colour vision tests have been made with computer-based testing. The Cambridge Color Test (CCT) adopted the pseudoisochromatic pattern from the printed tests described above with a coloured Landolt C orientation-identification stimulus embedded in static luminance noise, and utilises forced-choice adaptive procedures.[8] The Color Assessment and Diagnosis (CAD) test employs a moving chromatic square superimposed onto dynamic luminance noise.[9] The Rabin Cone Contrast Test (RCCT) adopted cone-isolating colours to screen for CVDs,[10,11] in which coloured letters with varying cone contrast were presented without any background noise. Resulting thresholds of these tests in vector length units can be used to classify the type and severity of CVD. In the clinical context, testing duration is crucial. The CCT Trivector test shortened the testing procedure by assessing only the three confusion lines. Faster and simpler tablet-based tests, which measure detectability along the three confusion lines, were developed for better accessibility and in young children.[12] The trade-off between testing speed and information collected has been the main disadvantage in these colour vision tests.

Here, we introduce and validate a generalizable, rapid and self-administered computer-based procedure named FInD (Foraging Interactive D-prime), and adapt it to assess colour

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

detectability and discriminability. FInD uses an adaptive algorithm that measures visual performance thresholds for a wide range of visual functions, e.g., contrast sensitivity, motion- and form-coherence.[13,14] The algorithm selects stimulus strength derived from d', a signal-to-noise ratio metric referring to the detectability of the stimulus, and calculates a new d' based on previous responses.

FInD Color Detection measures detection thresholds to L, M and S cone-isolating stimuli, reflecting the performance of individual cone types, while FInD Color Discrimination measures hue discrimination thresholds around six directions on a HSV (Hue, Saturation, Value) or eight directions on an equiluminant colour plane. The two colour spaces are used for different purposes – while equiluminant colours satisfy accurate measurement of colour discrimination, HSV colours do not require careful calibrations so can be readily applied.[15] The detection and discrimination tasks together interrogate an observer's colour vision system: the detection task classifies photoreceptor-level colour sensitivity, and the discrimination task further quantifies the resolution of colour perception. Unlike RCCT, FInD Color tasks use Gaussian blobs, thereby removing high frequency chromatic signals and avoiding optotype familiarity. Dynamic luminance contrast noise was also added to mask potential luminance artefacts. Our first aim was to validate FInD Color detection and discrimination in terms of thresholds, testing duration and reliability in both colour-normal (CN) and CVD groups. The second aim was to determine CVD type and severity using unsupervised machine learning classification with FInD Color detection and discrimination thresholds. The classification of CVD subtypes is difficult because genetic testing can be laborious and sometimes expensive, and results of current behavioural tests for thresholds vary due to different stimulus types and tasks. Personalised threshold results generated with FInD enable the deployment of unsupervised machine learning approaches to determine groups based purely on the behavioural performance of detectability and discriminability. The third aim was to compare FInD results against those of conventional clinical tools— HRR and FM100.

Methods

The experiment protocol was approved by Northeastern University Institutional Review Board, and followed the principles in the Declaration of Helsinki.

Participants

Nineteen participants (mean ± SD age: 26.2 ± 10.0 years; age range: 18–50 years; 9 females) with self-reported normal colour vision and 18 (mean ± SD age: 23.1 ± 7.8 years; age range: 18–54 years; 1 female) with self-reported inherited CVD were recruited after providing informed consent and completing a demographic and ocular history questionnaire. All participants had normal or corrected-to-normal (at least 0.00logMAR (6/6)) visual acuity and no history of eye diseases except for one participant with strabismus and two with amblyopia; one amblyopic participant also had defective colour vision. Two participants were diagnosed with attention deficit disorder and depression, respectively, and both received medical treatment.

Apparatus

Experimental procedures were programmed by Psychtoolbox[16] in MATLAB (MathWorks, mathworks.com), and presented on a 32" 4K LG display (lgdisplay.com) with a resolution of 3840×2160. Colour resolution of the graphics card was 8 bits per colour channel. The display was gamma-corrected with a Datacolor SpyderX elite colorimeter (datacolor.com) and the spectra and luminance were measured with a Photo Research PR-650 spectroradiometer (jadaktech.com). Luminance of the mid-grey (x=0.321, y=0.334) background was 90.3 cd/m^2. Participants viewed the screen binocularly at a distance of 111 cm, subtending a visual angle of 35°× 20°, with head position stabilised in a chin rest. Standard illumination for HRR and FM100 administration was provided by a Sol•Source daylight lamp (117V, 50/60 Hz) manufactured by GretagMacbeth (xrite.com). The time was recorded using a standard mobile phone timer application or by the computer for FInD tests.

Tasks and Stimuli

Four methods were compared in this study: Hardy-Rand-Rittler (HRR) Pseudoisochromatic Plates (4th Edition; bernell.com/product/RP396/Index_H), Farnsworth-Munsell 100 hue test (FM100; xrite.com/categories/visual-assessment-tools/fm-100-hue-test), Rayleigh colour match and FInD Color Detection and Discrimination tasks.

Hardy-Rand-Rittler (HRR) Pseudoisochromatic Plates—HRR was conducted by the experimenter flipping the plates and participants reporting the shape and location of the test colour symbols under controlled lighting.

Farnsworth-Munsell 100 hue test (FM100)—Participants were asked to arrange the 85 caps according to reference colours in each testing case. The four cases were completed in random order.

Rayleigh Colour Match—Rayleigh colour matching data were collected on 14/19 CN and 6/18 CVD with an Oculus HMC Anomaloscope (Oculus, us.oculus.de/us/products/visual-test-equipment/hmc-anomaloskop/functions/). Participants were asked to complete eight measurements (four with each eye) and an additional matching range identification trial.

FInD Color detection—Stimuli used in FInD Color Detection task were cone-isolated gaussian blobs (σ=1°, support diameter = 4°) (Figure 1a left). L-, M- and S-cone isolating directions in the RGB unit were calculated by integrating Stockman-Sharpe cone fundamentals[17,18] and the measured display spectra, then weighted by cone contrasts (detailed computation steps can be found in He et al.[19] See the supplementary materials for exact values. Stimulus contrast was adaptively controlled by the FInD algorithm.

FInD Color discrimination: HSV colour space—The FInD Color Discrimination task measured discriminability of a pair of two small gaussian blobs (σ=0.6°, support diameter=3.6°) having different colours (Figure 1a right). Each stimulus contained a pair of colours that were selected from the HSV colour space (Figure 1b). Six hues (H = red, yellow, green, cyan, blue, magenta), three primary and three confusion axes, in the

half-brightness (V=0.5) HSV plane were tested separately with two saturation levels (100% or 50%). For each selected hue axis, two stimulus colours were selected at the same angular distance away from this hue axis in opposite directions. The angular distance between test colours in HSV space was adaptively controlled by the FInD algorithm.

FInD Color discrimination: equiluminant colour plane—The task was the same as using HSV colours, but the stimuli differed. The equiluminant plane contained four primary colour directions, red-green axis (L-M, M-L) and purple-yellow axis (S+, S−), with red (R), purple (P), green (G) and yellow (Y) being 0°, 90°, 180° and 270°, respectively. The four intermediate angles were also tested (RP, PG, GY, YR; Figure 4 left). Colour contrasts were kept at 8 x threshold (See details in Supplementary material; *FInD Color discrimination task with an equiluminant colour plane*).

Procedures

Each participant completed the four tasks once (test session) or twice (test-retest sessions) in random order after the optometric screening.

The FInD adaptive algorithm measures *d'* to estimate thresholds efficiently. In short, *d'* is a measure of detectability or discriminability in signal detection theory. As shown in Figure 1c, *d'* is the distance between the noise and signal distribution means, and the location of the criterion (λ) directly affects the proportion of responses corresponding to "hit", "miss", "false alarm" and "correct rejection". The initial stimulus contrast (detection) or angle distance (discrimination) was determined by pre-defined approximate estimates of slope and threshold to derive the test range to span difficult (*d'*=0.1) to easy (*d'*=4.5) stimuli. The independent variable of the stimuli for the first chart was sampled in log steps across this test range. Subsequent charts used the combined responses across all previous charts to re-estimate *d'* (Figure 1e).

Stimuli were presented in charts (Figure 1a) containing 4×4 cells (6°×6°) at the centre of the screen, with each cell containing one stimulus at pre-selected contrasts (detection) or hue difference (discrimination) embedded in 8 Hz dynamic luminance noise, with check size of 10 arcmin and ±20% luminance contrast. A high stimulus-intensity example and instructions were provided at the upper left corner of the screen to help participants identify the targets to be foraged. Participants were informed that target stimuli were present in some but not all cells, and the number of targets varied from chart to chart. They were instructed to click on any cells that contained "faint versions" of the stimulus (detection) or on cells that contained blob pairs having different colours (discrimination). The chart stayed on the screen until the observer completed the current chart by clicking on the exemplar. Then, *d'* was calculated based on hit/ miss/ correct rejection/ false alarm classifications as a function of stimulus contrast (detection) or hue difference (discrimination). The estimate of *d'* from all previous responses was used to generate the stimulus range (contrast or hue difference) in the next chart. A series of three charts for each stimulus were presented in interleaved order (Figure 1d). The proportion of "yes" responses were fitted by a psychometric function as depicted in Equation 1 (Figure 1e).

$$Proportion\ Yes = 1 - \Phi\left(\Phi^{-1}(1 - FalseAlarmRate) - \frac{5 \times \left(\frac{TestLevel}{Threshold}\right)^{Slope}}{\sqrt{(5^2 - 1) + \left(\frac{TestLevel}{Threshold}\right)^{2 \times Slope}}}\right),$$

(1)

where Φ is standard normal cumulative distribution function.

Results

Data analyses were performed in MATLAB (mathWorks.com). Descriptive statistics of FM100 and FInD tasks for the CN group are summarised in Table S1. The analyses below are based on the test session data only, except for the repeatability analysis.

HRR

All CN observers passed the first ten demonstration and screening plates of the HRR test. The mean time recorded for 12 CNs to pass was 37 seconds. All those who self-reported CVDs failed HRR and were classified as protan or deutan with a mild, medium or strong defect (Table 1). The mean time to complete all 24 HRR plates for the 13 CVDs was 2.88 min. One CVD observer with mild red-green deficiency achieved equal protan and deutan scores, and thus could not be classified.

FM100

Total error scores (TES) and right-half mid-point (MP) were reported for the FM100. TES is calculated as the sum of the error scores for each colour with two subtracted from each error score. For CNs, TES is expected to range from 0 to 100 where superior, average and low discrimination abilities are indicated by error scores ranging from 0~16, 20~100 and > 100, respectively.[20] In our sample 42%, 53% and 5% of the CN observers, respectively, fell into these three categories (Table 1). Mean and standard deviation of the TES for the CN group are reported in Table S1, and the average error score pattern is shown in the left panel of Figure 2 for the CN group. Note that in the radial axis range, where the centre of the error score pattern is two, indicating perfect responses, with the largest error score scale being 3.5. Our TESs for CNs (2.42±0.24) were comparable to those reported in Knoblauch et al.[21] The error score pattern of one CVD participant (CVD#5) is shown in the right panel of Figure 2 (note radial axis range is 2–16). Patterns for each CVD observer are provided in Figure S1. The TES for 11% of the CVD observers exceeded the error score range of our CN group, while 89% of CVDs had TES > 100. The right-half MP contains the median of the error scores for cap sequence 43 to 84, and was used to identify the type of defect. The classification criteria are indicated in the right panel of Figure 2 according to the FM100 manual: protans, deutans and tritans had right-half mid-points ranges of 62~70, 56~61 or 46~52, respectively. TES and MP scores for the CVD group are reported in Table S2. The average test time was 13.53 min for 18 CNs and 13.38 min for 17 CVDs. The average time for plotting the pattern and calculating TES and MP manually was 3.28 min and 14.80 min

for CNs and CVDs, respectively. Computer-based data analysis algorithms can improve the efficiency of FM100 test by reducing the data processing time.

Rayleigh Colour Match

Average duration for Rayleigh match with an anomaloscope was 10.64 min for 14 CNs to complete eight measurements (four with each eye) and 15.07 min for 6 CVDs to complete all eight measurements with an additional matching range identification trial. Detailed testing procedures and results are reported in the Supplementary material. All 14 CN participants produced normal anomalous quotients and all eight CVDs produced atypical quotients. Six CVD classifications agreed with the HRR classification and four agreed with the FM100 classifications (Table 1).

FInD Color Detection and Discrimination Thresholds and K-means Clustering

FInD Color detection thresholds for each CVD observer are shown in the top panels of Figure 3, while low-saturation discrimination thresholds with the HSV colours are shown in bottom panels (Figure 3b; high-saturation discrimination thresholds are shown in Figure S2). For both tasks, data from CNs tend to cluster at lower thresholds and have relatively small variance, whereas there are large individual differences among CVD observers with different patterns and degrees of selective threshold elevation likely corresponding to CVD types. Average (standard deviation) durations for CN and CVD participants to complete all trials in the detection task were 4.95 min (1.98) and 5.13 min (1.75), respectively and 19.95 min (5.55) and 18.07 min (5.53) to complete all trials of the discrimination task, respectively. Given that two saturation levels were investigated in the discrimination task, one test level takes only half of the time reported <10 min).

To assess discrimination performance in an equiluminant colour plane (Figure 4 left), a subset of participants (6 CVDs and 14 CNs) were tested. Resulting discrimination thresholds along eight axes show that response patterns using the equiluminant plane were similar to that of HSV space: both protans and deutans had significantly elevated thresholds at yellow, green and blue/purple axes, with thresholds for red and other intermediate axes being more variable (Figure 4 right; see p-value tables in Supplementary material). Note that the adopted equiluminant plane was generated based on a theoretical standard observer with normal colour vision, so it would produce luminance cues for protan and duetan observers.

To set up diagnostic criteria for the FInD Color results, we sought to achieve an automated classification of CVD type and severity using unsupervised machine learning. To that end, a K-means clustering algorithm was applied with inputs being subsets of the detection and discrimination threshold datasets. A two-step classification was performed (see details in Supplementary section: K-means classification). The first step took LM detection thresholds as inputs and was able to segregate a large group. Observers in this group include CN and potential anomalous trichromats (AT) with low LM thresholds (Figure 5a green circles). We refer to this group of participants who have defective colour vision but low detection thresholds as 'potential AT' for simplicity in the following paragraphs. Three smaller clusters were also identified, one with high L and low M thresholds (CVD# 1, 5, 14; likely protanopes; Figure 5a blue diamonds), one with high M and low L thresholds (CVD#

2, 4, 6; likely deuteranopes; Figure 5a cyan squares), and one with intermediate M and low L thresholds (CVD# 3, 18; likely extreme deuteranomaly or deuteranopes; Figure 5a orange triangles). Observers who were classified in the three small groups received consistent classifications as from HRR, FM100 and anomaloscope data if available (Table 1). The second classification step took LMS detection thresholds and low-saturation yellow (Y), blue (B) and magenta (MA) hue discrimination thresholds as inputs, and divided the large group from the first step to four smaller clusters: one CN group (Figure 5 b–f: "1"s surrounded by green circles), one deutan group (likely deuteranomaly; Figure 5 b–f: "3"s surrounded by red squares; two received consistent HRR and FM100 classifications, and one received inconsistent HRR and FM100 classifications but was classified as deuteranomaly by the anomaloscope), one protan group (likely protanomaly; Figure 5 b–f: "4"s surrounded by red squares; two received consistent HRR and FM100 classifications, and one received inconsistent HRR and FM100 classifications) and one unknown AT group (Figure 5 b–f: "2"s surrounded by red squares; two received inconsistent HRR and FM100 classifications, and were classified as protanomaly and deuteranomaly, respectively, by the anomaloscope, and one received consistent HRR and FM100 classifications as protan). With the lack of genetic measurements available for the participants, however, we are not able to relate the subgroups to any specific genotype.

Test-retest reliability

Thirteen CN and eight CVD participants completed the retest session. All CNs passed HRR on both tests. HRR categorisation of two CVD observers changed: one observer (CVD#14) stayed the same type (protan) but with different severity (medium to strong), and the other observer (CVD#12) changed in both type and severity (strong protan to medium deutan). Bland-Altman analyses[22] showed no significant learning effect and bias (see details in the Supplementary material).

Comparison of the methods

Classification results with three methods for the 18 CVD observers are compared in Table 1. As we used consistent Rayleigh match as well as HRR and FM100 CVD type classifications as references in the two-step unsupervised classification, then agreement between FInD Color and the other tasks should be expected. For the observers classified as either strong or medium protan or deutan (likely protanope or deuteranope) using FInD Color LM detection thresholds, CVD type classification of the three methods agree perfectly (CVD#1, 2, 3, 4, 5, 6, 14, 18). However, L-cone and M-cone detection thresholds alone were not able to distinguish between CN and AT as they both have relatively low cone specific detection thresholds. With the addition of the hue discrimination thresholds, the algorithm successfully distinguished CN observers from anomalous trichromats (AT), and further assigned potential AT to CVD subtypes. The second-step classification results agreed for all 19 CN observers and five out of nine AT observers. The remaining four AT observers (CVD#7, 11, 15, 17) received inconsistent diagnoses from HRR and FM100; therefore for FInD classification, we assigned them to the CVD categories in which the members in their clusters fell, which necessarily agreed with one of their HRR or FM100 diagnoses, but not both. It is worth noting that for one cluster (CVD#:12, 15, 17), although one of the participants (CVD#12) received consistent HRR and FM100 diagnoses in the test session,

the retest HRR diagnosis changed type, so that all three participants received confusing HRR and FM100 diagnoses when the retest session results were considered. This cluster is referred to as the unknown AT group. As for testing durations, a FInD Color detection task with three testing directions and three trials per direction (CN: 4.95 min, CVD: 5.13 min) took longer than HRR testing (CN: 0.62 min, CVD: 2.88 min), but quick screening with only one trial was shorter (1.72 min) for the CVD subjects. A single level FInD Color discrimination task (CN: 9.98 min CVD: 9.03min) was quicker than FM100 (CN: 13.53 min, CVD: 13.38 min). FM100 testing can be further prolonged due to the additional time spent on data processing.

Discussion

The current study introduced and measured the performance of the newly designed FInD Color detection and discrimination tasks, and compared them against HRR, FM100 hue and Rayleigh matching tests. FInD is rapid, self-administered and easy to use, without strict operating regulations other than the normal use of a computer, and so can be readily grasped by patients (including children).[23] The combination of colour detection and colour discrimination performance also provides richer information about colour perception in CVDs, and enables classification of colour vision subtypes using unsupervised machine learning classification. In comparison, the HRR test, although rapid and easy, provides limited information about atypical colour perception, and is not able to classify CVD type for very mild cases (CVD#7). FM100 measures colour discrimination but does not have strict and precise diagnostic metrics. Even in such a small sample, some CVD observers obtained lower error scores than those with normal colour vision but poor discrimination. The discrimination pattern required further analysis in addition to TES for CVD diagnosis.[24] Moreover, the actual error score pattern shows a large degree of variation rather than a clearly classifiable pattern. FInD Color tasks afford a remedy for these shortcomings, which may be critical for detecting and tracking progression or remediation of acquired CVD in neuro-ophthalmic disease.[2] Furthermore, the FInD discrimination task significantly reduced testing and analysis time compared with FM100, while the FInD detection task, although taking a similar amount of time to the HRR for CVD testing, was more informative.

In the present implementation of FInD Color detection, we calibrated the monitor to generate and test detection thresholds for short-, medium- and long-wavelength sensitive cone-isolating stimuli, based on standard cone fundamentals (see Methods). For FInD Color discrimination, we employed HSV colour space to generate stimulus pairs of differing colours. HSV is based on subjective estimates of distance in perceptual colour space, which is convenient for the measurement of colour perception on standard computer displays. However, different colour hues in HSV may not be equiluminant, and therefore participants could utilise luminance differences instead of colour differences to select cells that contained blobs of differing colour. We attempted to mask any such cues with 20% contrast dynamic luminance noise, and in the Supplementary materials we report the magnitude of luminance artefacts in our display. The results show that artefacts differed with test hue angle in a pattern that was inconsistent with differences in discrimination threshold, and were less than 6 cd/m^2 for all test axes at the highest discrimination difference for CN participants. A separate luminance-matched achromatic blob discrimination task was conducted to evaluate

the magnitude of luminance artefacts, and thresholds for luminance difference are reported in Table S7. Sensitivities of CN observers to low-saturation luminance differences never exceeded the artefact generated at a threshold level of 5° hue angle. However, sensitivities to high-saturation luminance differences along the three hue axes (R, C, G) exceeded the threshold at a 5° hue angle. In this case, the results of the FInD Color discrimination test are conservative, and we would expect colour discrimination thresholds to be even higher for CVD participants. When compared with the equiluminant colour discrimination thresholds, the response patterns were similar for the two spaces, suggesting that the HSV colours, although not equiluminant, were equally diagnostic as the equiluminant colours.

Widely used CVD categories adopted by many colour vision screening tests were predicated on a colour vision phenotype which allows rapid detection of abnormal performance. Albeit simple, the conventional categories disguise the variable nature of CVD. The well-known large individual variations in CVD make the diagnostic groups less distinguishable, and their relationship to genotypes remains unclear.[25,26] For instance, in addition to the protanomaly category, one CVD type (Pseudo-protanomaly), while having an LM photopigment peak similar to protanopes, is able to produce trichromacy with differing optical density.[27] The story can become more complicated when interactions between post-receptoral signals are considered. Brain plasticity seems to elicit compensatory adjustments so that the actual perceptual colour ability loss for CVD can be reduced, further blurring boundaries between discrete phenotypes.[28] Unsupervised machine learning has the ability to group individuals with similar features to the same category, and has previously been used to successfully identify subtypes of other diseases.[29] Finding categories for groups is a non-trivial challenge for the following reasons: using prior information such as self-report of CVD or clinical results to classify data is dependent on the methods used in the clinical tests; Furthermore, treating CN and CVDs as one distribution, and using multiple standard deviations away from the mean of CNs does not consider the known CVD subtypes that belong in separate distributions. Hence, we decided to use an unsupervised machine learning technique — k-means, to address these classification problems without assigning a group *a priori*. The novel application of using a k-means clustering algorithm to establish CVD classification criteria in the present work sheds light on the competence of machine learning techniques in capturing obscure response patterns in psychophysical measurements of typical and atypical colour vision, thus possessing the potential to reveal the continuous and variable nature of CVD. As a result, the CVD group can be better categorised into subtypes, and inconclusive CVD cases that received confusing diagnoses from traditional tests can be better understood and treated.

The addition of hue discrimination thresholds in the second-step classification was essential to discern potential ATs, as detection thresholds alone did not effectively differentiate potential anomalous trichromats from normal trichromats (Note that the LMS cone isolating directions used in the detection task are computed based on a standard observer with normal colour vision; therefore, a certain degree of individual differences should always be considered). This finding suggests complexity in AT phenotypes, echoing the previous evidence.[25] We caution that inferences about AT features made based upon current results should remain provisional until larger datasets are acquired, which would also benefit the comprehensive evaluation of the FInD method. As discussed above, even though

the HSV plane in the discrimination task is not equiluminant, our classification results demonstrate its ability to assay response pattern differences between CN and CVD, with significantly different discrimination patterns observed, and the pattern of findings was the same for an equiluminant colour space. Apart from HSV, the FInD tasks implemented in this study required careful display calibration to generate cone-isolating directions (for an ideal observer model) and equiluminant colour planes. The same concerns prevail for the anomaloscope and for the light source and pigment decay of the HRR and FM100 tests. Therefore, deployment of these approaches will depend on calibration of the model specifications. The ability of FInD Color tasks to detect and classify tritan deficiencies, whether inherited or acquired, was not tested in the current study. Future efforts are needed to verify whether the unused features, S detection and the rest of the hue discrimination thresholds, are critical for the categorisation of tritans.

In summary, this proof-of-concept study has shown that FInD Color tasks can provide continuous colour detection and discrimination threshold estimates that may track changes in CVD and could serve as a rapid and easy-to-use tool for clinical monitoring and diagnosis. The HSV colours are equally diagnostic compared with the equiluminant colours. FInD also has potential in basic colour vision research, in that the combination of FInD Color tasks with an unsupervised machine learning technique might provide insights about hidden structures in the data and further assist the understanding of defective colour mechanisms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Sharpe LT, Stockman A, Jägle H, Nathans J. Opsin genes, cone photopigments, color vision, and color blindness. Color vision: From genes to perception. 1999;351:3–52.

2. Simunovic MP. Acquired color vision deficiency. Survey of Ophthalmology. 2016;61(2):132–55. [PubMed: 26656928]

3. Birch J Classification of anomalous trichromatism with the Nagel anomaloscope. In: D B, editor. Colour Vision Deficiencies XI: Kluwer Academic Press, Netherlands; 1993. p. 19–24.

4. Zabel J, Przekoracka-Krawczyk A, Olszewski J, Michalak KP. Variability of Rayleigh and Moreland test results using anomaloscope in young adults without color vision disorders. Plos one. 2021;16(5):1–15.

5. Bailey JE, Neitz M, Tait DM, Neitz J. Evaluation of an updated HRR color vision test. Visual Neuroscience. 2004;21(3):431–6. [PubMed: 15518225]

6. Cole BL, Lian Ky, Lakkis C. The new Richmond HRR pseudoisochromatic test for colour vision is better than the Ishihara test. Clinical and Experimental Optometry. 2006;89(2):73–80. [PubMed: 16494609]

7. Evans BE, Rodriguez-Carmona M, Barbur JL. Color vision assessment-1: Visual signals that affect the results of the Farnsworth D-15 test. Color Research & Application. 2021;46(1):7–20.

8. Mollon JD, Regan BC. Cambridge Colour Test Handbook. (Cambridge Research Systems Ltd, 2000).

9. Barbur JL, Harlow J, Plant GT. Insights into the different exploits of colour in the visual cortex. Proceedings of the Royal Society of London Series B: Biological Sciences. 1994;258(1353):327–34.

10. Rabin J Quantification of color vision with cone contrast sensitivity. Visual neuroscience. 2004;21(3):483–5. [PubMed: 15518234]

11. Rabin J Cone-specific measures of human color vision. Investigative ophthalmology & visual science. 1996;37(13):2771–4. [PubMed: 8977494]

12. Tang T, Alvaro L, Alvarez J, Maule J, Skelton A, Franklin A, et al. ColourSpot, a novel gamified tablet-based test for accurate diagnosis of color vision deficiency in young children. Behav Res Methods. 2022;54(3):1148–60. [PubMed: 34463952]

13. Bex P, Skerswetat J. FInD-Foraging Interactive D-prime, a rapid and easy general method for visual function measurement. Journal of Vision. 2021;21(9):2817.

14. Merabet LB, Manley CE, Pamir Z, Bauer CM, Skerswetat J, Bex PJ. Assessing Motion and Form Coherence Processing in Cerebral Visual Impairment. Developmental Medicine & Child Neurology. In Press.

15. Burger W, Burge MJ. Principles of digital image processing: fundamental techniques: Springer Science & Business Media; 2010.

16. Kleiner M, Brainard DH, Pelli D. What's new in Psychtoolbox-3? Perception. 2007;36(Suppl.):14.

17. Stockman A, Sharpe LT. The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. Vision Research. 2000;40(13):1711–37. [PubMed: 10814758]

18. Stockman A, Sharpe LT, Fach C. The spectral sensitivity of the human short-wavelength sensitive cones derived from thresholds and color matches. Vision research. 1999;39(17):2901–27. [PubMed: 10492818]

19. He J, Taveras-Cruz Y, Eskew RT Jr. Modeling individual variations in equiluminance settings. Journal of Vision. 2021;21(7):1–16.

20. Farnsworth D The Farnsworth-Munsell 100 Hue Test Manual (revised ed). Baltimore: Munsell Color Company. 1957.

21. Knoblauch K, Saunders F, Kusuda M, Hynes R, Podgor M, Higgins KE, et al. Age and illuminance effects in the Farnsworth-Munsell 100-hue test. Applied Optics. 1987;26(8):1441–8. [PubMed: 20454340]

22. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135–60. [PubMed: 10501650]

23. Merabet LB, Manley CE, Pamir Z, Bauer CM, Skerswetat J, Bex PJ. Motion and form coherence processing in individuals with cerebral visual impairment. Dev Med Child Neurol. 2023;00:1–8.

24. Mahon LE, Vingrys AJ. Scoring the Farnsworth-Munsell 100-hue for vocational guidance. Optometry and vision science. 1995;72(8):547–51. [PubMed: 8539021]

25. Bosten J The known unknowns of anomalous trichromacy. Current Opinion in Behavioral Sciences. 2019;30:228–37.

26. Neitz J, Neitz M. The genetics of normal and defective color vision. Vision Res. 2011;51(7):633–51. [PubMed: 21167193]

27. Neitz J, Neitz M, He JC, Shevell SK. Trichromatic color vision with only two spectrally distinct photopigments. Nature Neuroscience. 1999;2(10):884–8. [PubMed: 10491608]

28. Isherwood ZJ, Joyce DS, Parthasarathy MK, Webster MA. Plasticity in perception: insights from color vision deficiencies. Faculty Reviews. 2020;9:8. [PubMed: 33659940]

29. Eshaghi A, Young AL, Wijeratne PA, Prados F, Arnold DL, Narayanan S, et al. Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. Nature communications. 2021;12(1):1–12.

Key points:

- We introduced a new computer-based method, Foraging Interactive D-prime or FInD colour detection and discrimination that rapidly measured patients' colour perception and generated personalised models of their visual performance.

- The method was compared to clinical tests, including Hardy-Rand-Rittler pseudoisochromatic plates, Farnsworth-Munsell 100 hue test and an anomaloscope, and showed good agreement with these tests in detecting atypical colour vision.

- FInD enabled cluster analysis via machine learning, and was used to classify normal and atypical colour vision deficiency categories.
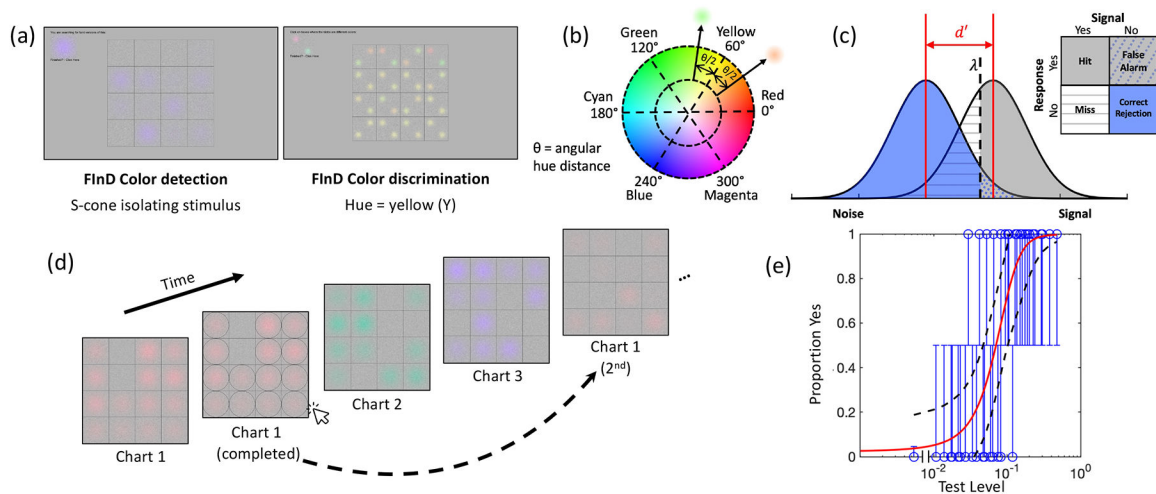
Figure 1.

Illustration of FInD stimuli and experimental procedures. (a) FInD detection (left) and discrimination (right) task interfaces. (b) The colour wheel shows a cross-section of the Hue, Saturation, Value (HSV) space for V=1, from which six hues (0° to 300° in 60° steps) and two saturation levels (0.5 and 1) were chosen and used in the discrimination task. If, for instance, yellow (60°) discrimination was tested, then two colours were symmetrically selected the same angular hue distance (θ/2) away from yellow with a fixed saturation level. (c) Illustration of signal detection theory. The noise distribution (blue) and signal distribution (grey) bell curves lie on the normalised Z-score abscissa. Detectability or discriminability (d') and criterion (λ) are depicted. The areas under the curves correspond to "hit", "miss", "false alarm" and "correct rejection", respectively, according to stimulus presentation and response. d' can be calculated by *z(false alarm)–z(hit)*. (d) FInD experimental procedures with cone isolating direction detection stimuli as an example. The dashed arrow represents the adaptive procedure that selects a range of stimulus intensities on the second chart based on analysis of the responses to stimuli on the first chart. (e) An example of a typical psychometric function: blue data show the probability that the observer reported the presence of a stimulus as a function of intensity; vertical lines show binomial standard deviation; the red curve shows the best fitting function for equation 1 and black dashed lines represent the upper and lower 95% confidence intervals. The leftmost data point, which is on the left of the break on the horizontal axis, indicates the false alarm rate (2.4% in this case).
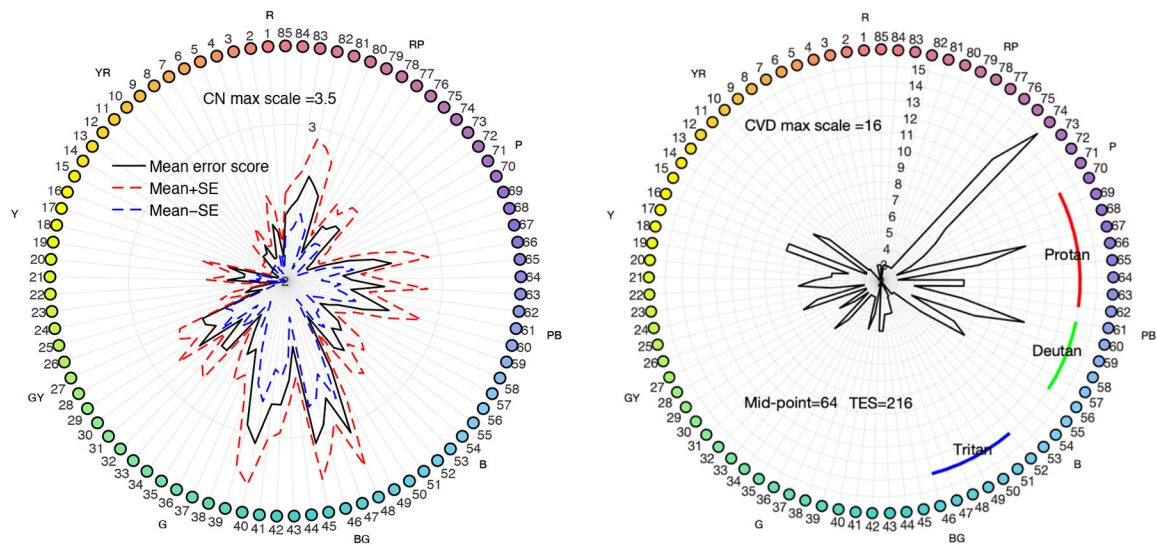
Figure 2.
FM100 hue test results. Left: average error score pattern of 19 colour normal (CN). Hues of coloured caps are numbered from 1 to 85 with the corresponding mean error score (black line) and standard error range indicated along the radial coordinates for each hue. Upper and lower standard error ranges are depicted by the red and blue dashed lines, respectively. The outermost circle where the colour dots reside represents an error score of 3.5, and the centre error score is 2, indicating the lowest possible error score. Right: error score pattern (black line) of an example colour vision deficient (CVD) observer (CVD#5). Note that the largest radial scale is 16. The mid-point and total error score (TES) of this observer as well as diagnostic curves (colour arcs) are also shown. The mid-point of this observer fell in the protan range.
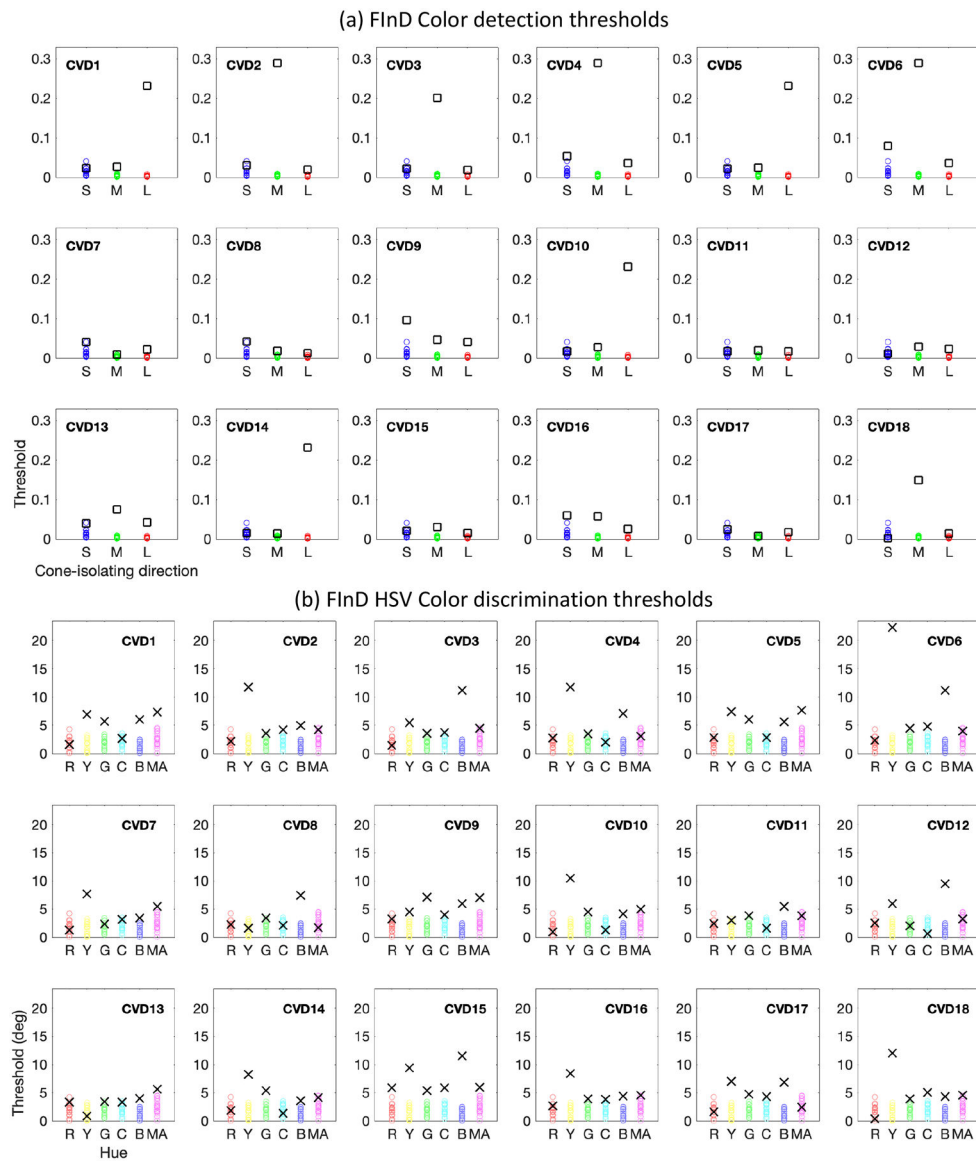
(a) FInD Color detection thresholds

(b) FInD HSV Color discrimination thresholds

Figure 3.

FInD Color detection and discrimination thresholds. Thresholds of all colour normal (CN) participants are plotted as coloured circles in all panels as references, and thresholds of each colour vision deficient (CVD) observer are denoted by black squares (detection) or crosses (discrimination) in separate panels. (a) FInD Color detection thresholds (upper panels) are plotted as cone contrast vector length and (b) low-saturation discrimination (lower panels) thresholds are plotted in degrees of Hue, Saturation, Value (HSV) colour space angle.
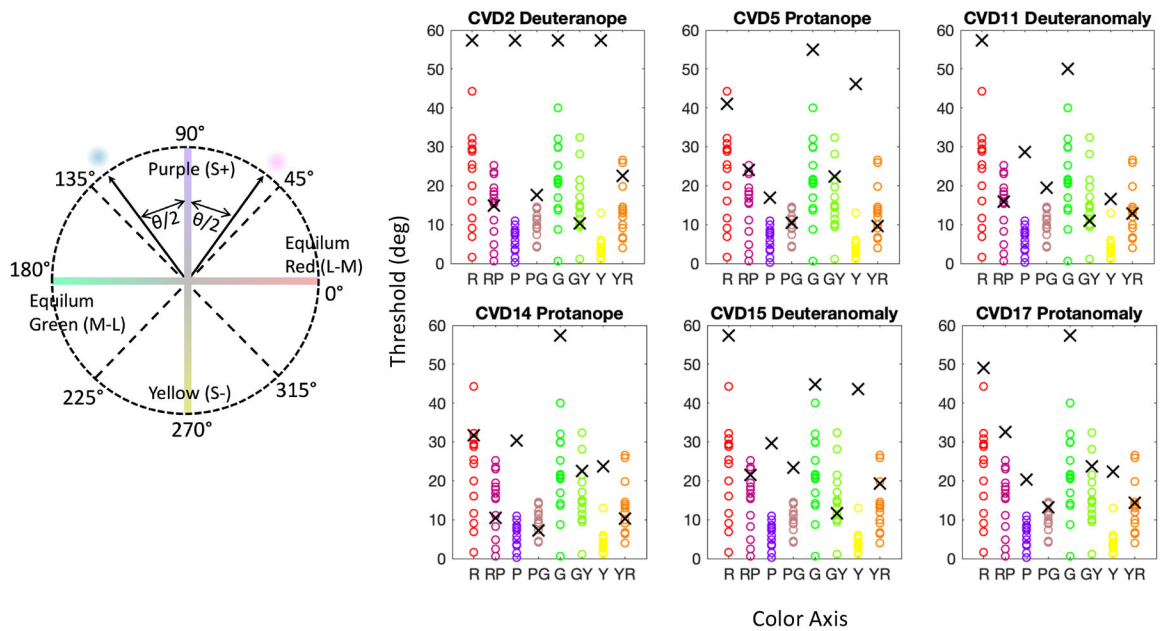
Figure 4.
Equiluminant colour discrimination. The colour wheel on the left shows the equiluminance plane with four primary axes representing the red-green and blue-yellow postreceptoral mechanisms. For example, in the purple (S+, 90°) discrimination condition, two colours are symmetrically selected at the same angular distance (θ/2) away from purple. The panels on the right show results of 6 colour vision deficient (CVD) participants for the FInD Color discrimination task with equiluminant stimuli. Thresholds of all colour normal (CN) participants are plotted as coloured circles in all panels as references, and thresholds of each CVD observer are denoted by black crosses in separate panels. The anomaloscope diagnosis for each participant is indicated.
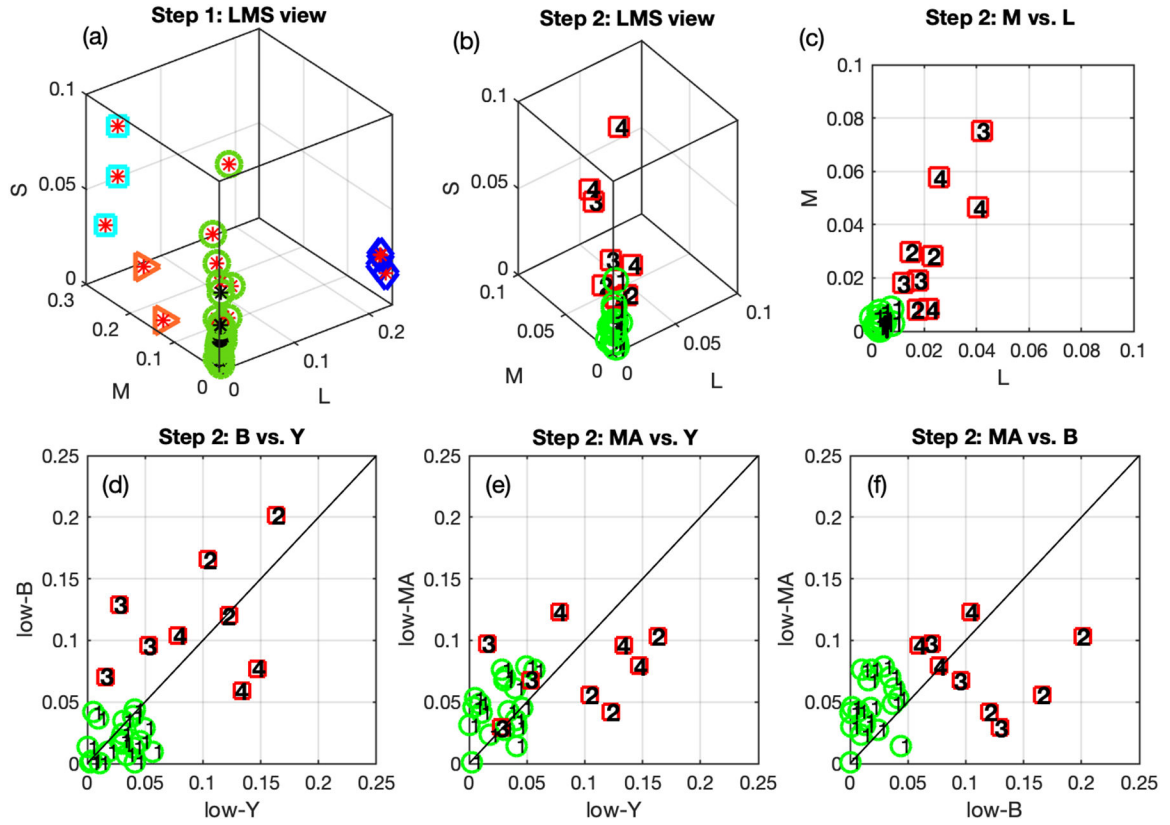
Figure 5.

Classification of colour normal (CN) and colour vision deficient (CVD) results. (a) Step one clustering results illustrated in LMS detection threshold space. All 37 individual data points are shown. CVD and CN individuals are represented by red and black asterisks, respectively. Four clusters, denoted in differently shaped and coloured symbols (green circles, blue diamonds, cyan squares and orange triangles) around the asterisks, are identified. (b)-(f) show step two clustering results. Only individual points surrounded by green circles (n=28) in (a) are plotted. These thresholds were clustered as four groups denoted by numbers. Red squares and green circles represent CVD and CN, respectively.

Table 1.

Identification of colour vision deficiency (CVD) type and severity for 18 CVD observers for the three methods. AT, anomalous trichromat; FM100, Farnsworth-Munsell 100 hue test; HRR, Hardy-Rand-Rittler pseudoisochromatic plates.

CVD#	Anomaloscope	HRR		FM100	FInD Color K-means	
		Type	Severity	Type	Type	Severity
1		Protan	Strong	Protan	Protan	Strong
2	Deuteranopia	Deutan	Strong	Deutan	Deutan	Strong
3		Deutan	Strong	Deutan	Deutan	Medium
4		Deutan	Strong	Deutan	Deutan	Strong
5	Protanopia	Protan	Strong	Protan	Protan	Strong
6		Deutan	Strong	Deutan	Deutan	Strong
7		Pro/Deu	Mild	Deutan	Protan	Mild
8		Deutan	Mild	Deutan	Deutan	Mild
9		Protan	Strong	Protan	Protan	Mild
10		Protan	Strong	Protan	Protan	Strong
11	Deuteranomaly	Deutan	Mild	Protan	Deutan	Mild
12		Protan	Strong	Protan	unknown AT	Mild
13		Deutan	Strong	Deutan	Deutan	Mild
14	Protanopia	Protan	Medium	Protan	Protan	Strong
15	Deuteranomaly	Deutan	Mild	Protan	unknown AT	Mild
16		Protan	Medium	Protan	Protan	Mild
17	Protanomaly	Deutan	Mild	Protan	unknown AT	Mild
18		Deutan	Strong	Deutan	Deutan	Medium