

Curveball: A tool for rapid measurement of contrast sensitivity based on smooth eye movements

Scott W. J. Mooney

Burke Neurological Institute, White Plains, NY, USA



N. Jeremy Hill

Burke Neurological Institute, White Plains, NY, USA

Melis S. Tuzun

Burke Neurological Institute, White Plains, NY, USA

Nazia M. Alam

Burke Neurological Institute, White Plains, NY, USA

Jason B. Carmel

Burke Neurological Institute, White Plains, NY, USA

Glen T. Prusky

Weill Cornell Medicine, New York, NY, USA
Burke Neurological Institute, White Plains, NY, USA

The contrast sensitivity function (CSF) is an informative measure of visual function, but current tools for assessing it are limited by the attentional, motor, and communicative abilities of the participant. Impairments in these abilities can prevent participants from engaging with tasks or following an experimenter's instructions. Here, we describe an efficient new tool for measuring contrast sensitivity, Curveball, and empirically validate it with a sample of healthy adults. The Curveball algorithm continuously infers stimulus visibility through smooth eye tracking instead of perceptual report, and rapidly lowers stimulus contrast in real time until a threshold is found. The procedure requires minimal instruction to administer and takes only five minutes to estimate a full CSF, which is comparable to the best existing methods available for healthy adults. Task repeatability was high: the coefficients of repeatability were 0.275 (in \log_{10} units of RMS contrast) within the same session and 0.227 across different days. We also present evidence that the task is robust across illumination changes, well correlated with results from conventional psychophysical methods, and highly sensitive to improvements in visual acuity from refractive correction. Our findings indicate that Curveball is a promising means of accurately assessing contrast sensitivity in previously neglected populations.

Introduction

The contrast sensitivity function (CSF) is a useful measure of visual system performance (see Pelli & Bex, 2013) and is well correlated with many other measures of visual health and disease (Bodis-Wollner, 1972; Bodis-Wollner, Onofrij, Marx, & Mylin, 1986; Chan, Edwards, Woo, & Woo, 2002; Cimmer et al., 2006; Collins & Carney, 1990; Kleiner, Enger, Alexander, & Fine, 1988; Regan, Raymond, Ginsburg, & Murray, 1981; Woo, 1985), but the best means through which to assess the CSF is a matter of ongoing research. Unlike other measures of basic visual function, the CSF describes a continuum of sensitivity thresholds and is consequently more difficult to assess than one-dimensional measures such as visual acuity. It can be constructed from a sequence of independent contrast thresholds computed at different spatial frequencies, but the time required to measure multiple thresholds with conventional psychophysical staircase procedures is unrealistic in clinical settings. Newer variants of the staircase technique reduce the time required to estimate the CSF by using adaptive Bayesian algorithms to repeatedly compute the most informative combination of spatial frequency and contrast to present in each trial (Lesmes, Jeon, Lu, & Doshier, 2006; Lesmes, Lu, Baek, & Albright, 2010; Vul, Bergsma, & MacLeod, 2010). These methods often directly model the CSF as a parameterized function (see Pelli, Rubin, & Legge,

Citation: Mooney, S. W. J., Hill, N. J., Tuzun, M. S., Alam, N. M., Carmel, J. B., & Prusky, G. T. (2018). Curveball: A tool for rapid measurement of contrast sensitivity based on smooth eye movements. *Journal of Vision*, 18(12):7, 1–19, <https://doi.org/10.1167/18.12.7>.



1986) rather than fitting a curve to independent thresholds obtained at different spatial frequencies.

Simultaneous two-dimensional Bayesian approaches can estimate a participant's CSF in less than five minutes (Dorr et al., 2017) and can even be run on a portable device (Dorr, Lesmes, Lu, & Bex, 2013), but they do not address one of the fundamental restrictions of traditional psychophysical tasks: their reliance on extended periods of attention and volitional perceptual report. The tasks can be tedious because they require the repeated presentation of visually uninteresting stimuli such as filtered noise patterns and sinusoidal gratings. Adaptive procedures also naturally become more difficult as the threshold of an observer's ability is approached; by design, these tasks spend as much time as possible presenting stimuli at peri-threshold combinations of contrast and spatial frequency. These drawbacks can be tolerated by healthy adults, but are not well tolerated by observers who are less willing or less able to engage with the test, such as children or individuals with brain injury (Witton, Talcott, & Henning, 2017). Witton et al. suggest that the only available workaround for less motivated subjects has been to add more trials (hence, more time) to the task. This strategy, however, does not help individuals (of any age) who have difficulty sustaining attention, following task instructions, or communicating their responses to an experimenter. Simplified report-based tools such as Pelli-Robson charts (Pelli, Robson, & Wilkins, 1988) can provide fast approximations to full psychophysical measurement of contrast sensitivity for certain ranges of spatial frequencies (Leat & Woo, 1997), but are similarly reliant on intentional feedback from the participant. There are alternatives to report-based tasks for impaired, noncommunicative populations, but they also have significant shortcomings that have prevented them from being routinely used in clinical settings. Preferential looking paradigms such as Teller cards circumvent the need for verbal communication, but they are still highly dependent on the participant's attention span (Teller, McDonald, Preston, Sebris, & Dobson, 1986). Electrophysiological methods such as visual evoked potentials are more sensitive than preferential looking paradigms (Katsumi, Denno, Arai, de Faria, & Hirose, 1997; Riddell et al., 1997) and have the significant benefit of providing objective measures of visual function (Leat, Yadav, & Irving, 2009), but require specialized training to administer and have reduced sensitivity relative to tasks with intentional behavioral report (de Faria, Katsumi, Arai, & Hirose, 1998). The longer time required for setup and measurement may also make it more difficult for participants to consistently attend to the display throughout the task, particularly if they have cognitive impairments. More promising techniques for assessing these populations have emerged that infer stimulus

visibility indirectly through tracking behavior. Bonnen, Burge, Yates, Pillow, and Cormack (2015) had participants move a cursor to track the position of a luminance patch obscured by noise and found that tracking precision could be used to estimate visual uncertainty as accurately as (and more efficiently than) perceptual report. The authors describe their approach as "continuous psychophysics," in which every change in the stimulus constitutes an informative "mini-trial." Continuous measures of visual function have similarly been obtained using eye trackers (see Schütz, Braun, & Gegenfurtner, 2011), which have been rapidly improving in accuracy and accessibility (Gibaldi, Vanegas, Bex, & Maiello, 2017). Most of these studies used eye trackers to assess low-level oculomotor function (such as saccade latency; Engbert & Kliegl, 2003) or to infer attention in subjective preferential looking tasks aimed at higher-order abilities such as object perception (Einhäuser, Spain, & Perona, 2008) and face perception (Cerf, Frady, & Koch, 2009).

As the name implies, eye tracker data can also be used to infer performance in continuous tracking tasks. Dakin and Turnbull (2016) described a gaze-based procedure that estimates contrast sensitivity by exploiting the automated analysis of *optokinetic nystagmus* (OKN). This oculomotor reflex induces smooth conjugate eye movements to stabilize the world on the retina (e.g., during head rotation). It typically does not occur in the absence of a moving visual field (Cohen, Matsuo, & Raphan, 1977) and therefore constitutes reliable evidence of stimulus visibility. Previous studies have assessed visual acuity in infants by using rotating striped drums to induce OKN (Catford & Oliver, 1973; Dobson & Teller, 1978), and others have measured contrast sensitivity in rodents by filling their visual field with drifting gratings and observing which combinations of spatial frequency and contrast elicit the optokinetic response (Douglas et al., 2005; Prusky et al., 2004). Dakin and Turnbull (2016) applied a similar technique to measure contrast sensitivity in humans. They used an eye tracker to detect the direction of OKN in response to full-screen drifting noise on a computer display and found that trial outcomes classified by OKN direction produced sensitivity thresholds that were similar to thresholds obtained from perceptual report. This approach takes promising steps toward contrast sensitivity assessment in noncommunicative populations: the task does not rely on volitional report and can potentially be administered without instruction, provided that the participant attends to the screen. Other elements of their procedure, however, are likely to pose problems for many participants. The procedure takes almost 20 minutes to complete, which is an unrealistic requirement for inattentive or cognitively impaired participants. The authors suggest that this time can be reduced by

combining the procedure with an adaptive staircase, but it is unclear how well this would work, as their OKN matching algorithm has a false positive rate of 50% and the time-sensitive nature of the task's individual trials appear highly vulnerable to lapses in attention.

We designed a novel method for estimating contrast sensitivity using eye movements that minimizes the influences of attention, motivation, and communicative ability on task performance without sacrificing the efficiency of conventional Bayesian staircase methods. The task, called Curveball, expands on the concept of continuous psychophysics by adapting the appearance of the stimulus to the participant's performance in real time. We created an algorithm that detects *smooth pursuit tracking* by comparing the position, velocity, and acceleration of the participant's gaze to the trajectory of a drifting circular patch of band-limited noise. Most smooth pursuits differ from OKN in that they are driven by localized targets and rely more on foveal sensation (Bahill & McDonald, 1983; Schütz et al., 2011), but are similarly unlikely to occur in the absence of a moving stimulus (Barnes, 2008). Here, we refer to the behavior of interest as *smooth tracking* to differentiate it from any pursuits that are not induced by the target stimulus. The Curveball algorithm exploits smooth tracking to infer noise patch visibility on a frame-by-frame basis and continuously decreases the contrast of the noise target while the participant smoothly tracks it. The reciprocal of the target's final contrast value (determined after an interval of no smooth tracking) provides an estimate of the participant's sensitivity threshold at that target's spatial frequency. This approach differs from previous continuous psychophysical procedures, which assessed visual ability by summarizing performance over discrete trials of fixed length (Bonnen et al., 2015; Dakin & Turnbull, 2016). It instead has more in common with the method of adjustment: each "trial" lasts as long as needed to produce a contrast threshold (typically between 5 and 10 seconds), but the real-time contrast adjustment is controlled by an objective algorithm rather than subjective report. The target follows an unpredictable curving path around the display to keep the stimulus visually interesting and to ensure that subthreshold "aftermath" smooth pursuits are not misclassified as target tracking (as these nontracking pursuit movements are unlikely to match the target's future path). The task requires the participant to be able to exhibit this smooth tracking behavior, which may not be the case for all participants (particularly those who present with broad visual impairments and/or diffuse brain injury). It may nevertheless allow inclusion of a previously unmeasurable subpopulation who have adequately preserved eye movements but cannot participate in existing procedures that are

difficult due to their cognitive or communicative deficits.

In this report, we describe a series of experiments that assessed the repeatability, accuracy, and technical limitations of the Curveball procedure in a sample of healthy adult participants with a wide age range and varying acuity and refractive correction. Our goals were to (a) establish the feasibility of algorithm-driven real-time stimulus manipulation as an alternative to the traditional model of discrete psychophysical trials, and (b) demonstrate that such a procedure retains its validity in a realistic clinical setting. To this end, the experiment was designed to replicate the conditions that are likely to occur in a clinical environment. The procedure was conducted in a well-lit room; head movements were not restrained; the only explicit instruction given to participants was to look at what was visible on the screen; gaze was tracked binocularly at a near distance of 62 cm; and eye tracker calibration was kept to a minimum. If the Curveball procedure does not perform adequately under these conditions, then it will not be useful in a wide range of clinical settings or serve participants with neural disability (regardless of whatever it can achieve under "ideal" conditions that might include detailed experimenter instructions, a head-mounted eye tracker, monocular viewing with an eye patch, head restraints, extensive calibration, and/or a darkened room).

We assessed the repeatability of Curveball within and between testing sessions on two different days and tested its validity by comparing its results to eye chart letter acuity measurements as well as a traditional forced-choice staircase procedure. If the Curveball task is a valid and reliable measure of contrast sensitivity, the CSF curves it generates should (a) be systematically related to the curves for both static and moving stimuli generated by the staircase task, and (b) fall off more rapidly at higher spatial frequencies when participants are measured without their corrective eyewear (depending, to an extent, on their resulting loss of acuity). We also assessed the extent to which Curveball's thresholds depend on the ability to smoothly pursue a moving target. The real-time contrast reduction algorithm relies on smooth eye movements, and participants who cannot consistently produce these eye movements may have their contrast sensitivity underestimated. Our analysis of the gaze data will determine whether any loss of Curveball accuracy caused by poor smooth tracking quality is abrupt (and hence easily detected) or more gradual (and hence more difficult to distinguish from genuine changes in contrast sensitivity). Finally, we assessed how Curveball thresholds vary over changes in room illumination and the participant's distance from the display and eye tracker. If Curveball is to be a useful measure of visual function in clinical settings, it is critical that small variations in these

parameters do not have a large or unpredictable impact on the task's performance.

Methods

Participants

Thirty-five healthy adults (19 women, 16 men; mean age 38.6 ± 15.8) with normal or corrected-to-normal vision participated. All other participants were naïve to the aims of the experiment and were recruited as staff volunteers from the Burke Neurological Institute (BNI). Eighteen participants had corrective eyewear, which was worn in all but one condition of the experiment. All participants provided informed consent under a protocol approved by the Burke Rehabilitation Center Institutional Review Board and were not compensated. Experimental data were secured and managed with the REDCap database.

Apparatus

A 27-in. widescreen LCD Lenovo Horizon 2 all-in-one computer was used to present the stimuli (Figure 1). Screen luminance was calibrated with the sRGB profile (gamma of approximately 2.2; Anderson, Motta, Chandrasekar, & Stokes, 1996). Screen luminance was measured with an ILT1700 radiometer (International Light Technologies, Peabody, MA) and ranged linearly from 0.1 (black) to 211.1 (white) cd/m^2 with the room lights off (the dark condition) and 10.0 to 221.1 cd/m^2 with the lights on (all other conditions). The display was mounted on a wheeled stand with an articulated arm and equipped with a USB display-mounted Tobii 4C eye tracker (Tobii Technology, Stockholm, Sweden). The Tobii 4C has an operating distance of 50 to 95 cm and samples mean gaze position at 90 Hz, though its output was only queried at 60 Hz by our Curveball software. Its precision, operating range, and tolerance of head movements are substantially better than the previous Tobii consumer-grade product, the EyeX, which was reviewed by Gibaldi et al. (2017). The Tobii 4C estimates gaze for each eye independently and can therefore be used monocularly or binocularly (by combining the data from both eyes). Only the binocular mode was used in the present experiment to minimize the apparatus requirements for the participant. Stimulus behavior was programmed in Python using the Shady graphics toolbox (Hill, Mooney, Ryklin, & Prusky, 2018) and was updated and rendered at a frame rate of 60 Hz. Gaze data were analyzed in real time using our novel Curveball algorithm, which measures the similarity between gaze

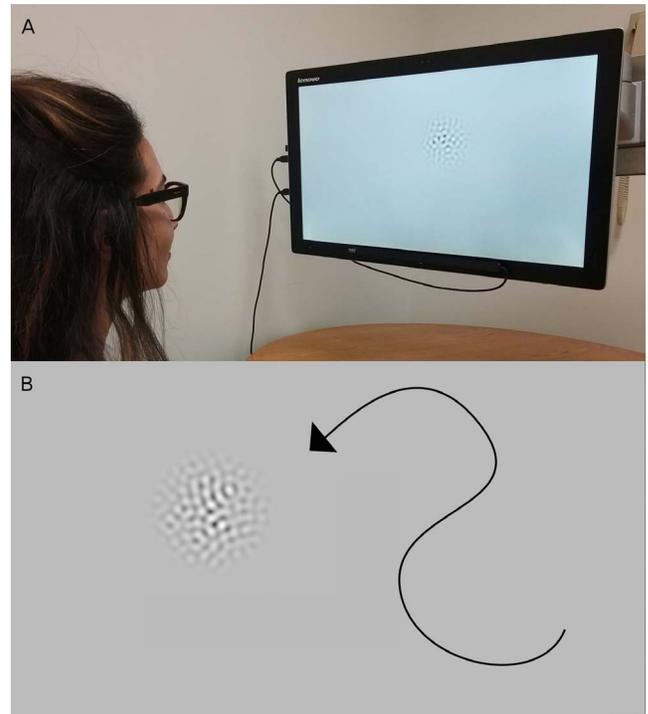


Figure 1. The Curveball procedure. (A) A participant engaging with the Curveball task. An eye tracker (the Tobii 4C) is attached to the bottom of this all-in-one computer, which is held in front of the participant at 62 cm with an articulated arm. The filtered noise target drifts around the screen and decreases in contrast in real time as the participant smoothly tracks it. The noise depicted here is 1 cpd. (B) A screenshot from the Curveball task that demonstrates the size and Hann windowing of the noise target. The noise target depicted is again 1 cpd.

and stimulus trajectories to infer stimulus visibility on a frame-by-frame basis.

Curveball task

Each Curveball run began with the sudden presentation of a white disc with three holes in the center of the display against a uniform gray background (value of 0.5). The same background was used throughout the task. The disc was designed to draw the participant's gaze to a central calibration point without explicit instructions and rotated with increasing angular velocity as the participant looked within 8° of visual angle (hereafter simply $^\circ$) of its position. This phase calibrated for any small offset in gaze position and ensured that the participant was paying attention to the display before launching the main task, which was timed. After 0.5 s of calibration, the disc faded out and the first Curveball trial began.

In each trial, a narrow-band frozen noise patch subtending 12° appeared at a random location on the screen. It then moved around the display, continuously

veering clockwise or counterclockwise in a sequence of smooth random turns while maintaining a fixed speed of $10^\circ/\text{s}$ (identified as the optimal speed for smooth tracking by Burr & Ross, 1982, and Dakin & Turnbull, 2016). The stimulus image was generated by applying a circular-symmetric Hann window to a filtered noise pattern that was regenerated with a new random seed on each trial. The noise started off with a $1/f$ amplitude spectrum in the frequency domain and a random phase spectrum. It was then filtered with an annular band-pass filter centered on the target spatial frequency. The minimum and maximum bounds of the filter were computed by multiplying and dividing the target spatial frequency by 0.9, respectively, which gave the filter a width of approximately 0.34 octaves. The resulting noise had equal power at all orientations but was limited to a narrow band of spatial frequencies. Temporal aliasing at high spatial frequencies was prevented by applying an additional anisotropic filter to the amplitude spectrum of the noise. This filter removed all components with horizontal spatial frequency greater than 2.85 cycles per degree (hereafter cpd), which is 95% of the Nyquist limit (3 cpd) of a stimulus moving at $10^\circ/\text{s}$ on a display with a refresh rate of 60 Hz. The orientation of the noise patch was continuously steered into its direction of motion to keep the anti-aliased direction of this filter facing “forward” at all times. The noise target sharply rebounded whenever it collided with the edge of the screen and simultaneously rotated by 180° to continue facing forward. The rapid variation in stimulus position and rotation also ensured that it was presented at all orientations in all regions of the screen within a single trial. The target’s size (12°) was chosen to make it large enough to display the lowest spatial frequency in the procedure (0.25 cpd) while being small enough that its rotation did not interfere with the smooth tracking detection algorithm if the participant happened to fixate away from its center (where target rotations produce transient higher gaze velocities). Its size was fixed across all spatial frequencies to avoid changing the difficulty of tracking. A screenshot with the target at high contrast is depicted in Figure 1B.

At the start of each trial, a semitransparent cartoon ghost was superimposed on the new noise target and locked to its movement trajectory. We used this cartoon to quickly draw the participant’s attention to the new target’s starting position without explicit instructions. The ghost faded out as soon as the participant’s gaze came within 5° of the new target. After the ghost had fully disappeared, the Curveball algorithm began searching for smooth tracking by continuously comparing the recent 2D trajectories of the participant’s gaze and the roving noise target. The trajectory of the target was defined as the sequence of its eight most recent 2D positions on the screen,

including the current position. An expected gaze trajectory was constructed from this by translating the target trajectory in 2D space so that the current target position was shifted to match the current gaze position. This was accomplished by computing the difference between the current gaze and target positions and subtracting that difference from each data point in the target trajectory. If the participant’s gaze followed the same recent path as the Curveball target, it would be well approximated by this expected trajectory. A tracking hit was recorded if each of the last eight gaze samples was within 0.4° of the corresponding eight points in the expected trajectory. The algorithm’s precise trajectory length (eight frames) and error tolerance (0.4°) were determined through extensive pilot testing of multiple individuals with the Tobii 4C. After five frames (83 ms) of consecutive smooth tracking hits, the root mean square (RMS) contrast of the noise target began to decrease logarithmically as long as smooth tracking continued. The starting RMS contrast of the noise was 0.317; this was above the maximum (~ 0.22) that could be displayed without clipping, but it was intentionally chosen as such for maximum initial visibility. Every frame of ongoing tracking caused its RMS contrast to be multiplied by 0.97. If the participant stopped pursuing the target for a single frame, the contrast reduction halted, and the algorithm again waited for five consecutive frames of tracking before resuming it. Contrast never increased during a trial. Participants instinctively followed the target’s motion on each trial until it had faded beyond their threshold, which took between 5 and 10 s depending on the participant’s sensitivity to that spatial frequency and the consistency of their smooth tracking. The trial was terminated according to a continuously updated deadline: every trial started with a lifespan of 3 s starting from the moment the ghost disappeared, and this lifespan increased by six frames (0.1 s) every time a frame of smooth tracking occurred. Participants therefore needed to pursue the target for at least one in every seven frames, on average, to prevent the trial from terminating. When this lifespan expired, the reciprocal of the noise target’s final RMS contrast in each trial was recorded as a sample of the contrast sensitivity threshold at that target’s spatial frequency. If the final RMS contrast value was above 0.22 (where the stimulus pixel intensities went out of range), no threshold was recorded. Note that less than 0.25 s of tracking was needed to reduce the target’s contrast below this value. The next trial immediately began with full contrast, a new noise stimulus, and the cartoon ghost.

Each participant completed four repeats of six spatial frequencies in a full Curveball run. The spatial frequency values were equally spaced in log units: 0.25, 0.5, 1, 2, 4, and 8 cpd. The lowest two contrast

thresholds for each spatial frequency were averaged to determine the final threshold estimates. This was done to account for participants dropping trials due to false negatives, which could be caused by inattention, poor or infrequent tracking, or other reasons. In lieu of a systematic way of detecting these false negatives, we discarded the worst (highest) 50% of threshold estimates to remove them. The 24 noise patches required were generated on the CPU as the task was initialized, but their visibility, contrast, windowing, gamma correction, “noisy-bit” dithering (Allard & Faubert, 2008), and position were processed in real time on the GPU using the Shady graphics toolbox (Hill et al., in prep). The efficiency of these GPU operations ensured that the task ran at a consistent frame rate of 60 Hz. A video of the Curveball task can be found in the supplemental materials (Supplementary Movie S1).

Staircase task

Thresholds were also obtained using a conventional four-alternative forced choice (4AFC) staircase task for comparison with Curveball. In each trial of this task, a windowed sinusoidal grating subtending 10° was presented against a mid-gray background in one of four cardinal locations around the center of the screen. We used conventional sinusoidal gratings rather than matching the Curveball stimuli to ensure that the between-task comparison could provide validation of the direction-filtered noise patches designed specifically for Curveball’s restrictions, in addition to validating the other novel elements of the task. The orientation of the grating matched the direction of its position (so gratings to the left and right of the center were horizontal and gratings above and below the center were vertical) to aid participants’ responses. Each grating was faded in and out for 0.25 s at the start and 0.25 s at the end of each trial according to a raised cosine function to avoid temporal visibility artifacts. Trials had no time limit and were separated by an interval of 0.75 s. Weighted up-down staircases (Kaernbach, 1991) targeting the 75% log RMS contrast sensitivity threshold were interleaved for 11 stimuli: six static gratings with the same spatial frequency as the Curveball targets (slightly rounded to an even integer number of cycles per image) and five moving gratings with spatial frequency values of 0.125, 0.25, 0.5, 1, and 2 cpd. The moving gratings could not be filtered to avoid temporal aliasing in the same way as the Curveball noise patches and consequently had to be shifted to this lower and more restricted range of spatial frequencies.

Participants were instructed to press the arrow key that matched the direction of the stimulus position in each trial. After a correct response, the contrast of that

stimulus was multiplied by 0.675 for the next presentation of that stimulus and a green spot flashed in the center of the display. After an incorrect trial, the contrast of that stimulus was multiplied by 4.444 for its next presentation and a red spot flashed. The initial RMS contrast of each grating was 0.317, as in the Curveball procedure. Each staircase ran until six reversals in contrast had occurred, at which point that stimulus no longer appeared in the task. A mean of run was computed between each of the four pairs of adjacent reversals excluding the first, and these four means were averaged to calculate the participant’s threshold for that stimulus. The full set of interleaved staircases took between 15 and 25 min to complete. A video of the conventional 4AFC task can be found in the supplemental materials (Supplementary Movie S2).

Procedure

The experiment was conducted in a well-lit laboratory setting in two sessions taking place on different days. In each session, participants were seated in a chair positioned against one wall to allow them to rest their head against the wall. The screen was positioned 62 cm from the participant, which caused the display to subtend a visual angle of 51° horizontally and 30° vertically. Distance was maintained throughout each experiment using depth data provided by the eye tracker: the program automatically paused the task and obscured the display whenever the participant deviated from the target distance by more than 5 cm (i.e., closer than 57 cm or further than 67 cm) and automatically resumed when they moved back into the permitted range. We chose to enforce distance using this approach rather than by restricting head movements to better assess Curveball’s repeatability in a realistic clinical setting. All participants were tested binocularly (i.e., with no eye patch), to ensure that the procedure can estimate a reliable CSF even in conditions with minimal apparatus.

In the first session, participants first had their corrected LogMAR visual letter acuity measured at the same distance (62 cm) with a Tumbling E eye chart chosen at random from a set of six randomly generated charts (Taylor, 1978). The chart had five characters per line and was attached to the screen while the experimenter obscured the participant’s view. Participants read the chart from the top and were permitted to describe the characters as letters (E, M, 3, or W) or using the orientation of the limbs (right, down, left, or up). Acuity was not measured separately for each eye. Two participants inadvertently wore contact lenses in the first session and did not have their letter acuity measured until the second session.

Participants then completed one standard run of Curveball. The only instruction given was to simply look at whatever objects they saw on the screen. This Curveball run was followed by the conventional 4AFC interleaved staircase task. All participants then completed another standard Curveball run, and participants with refractive correction performed an additional standard Curveball run and eye-chart measurement without their corrective eyewear. Two participants who wore contact lenses in the first session chose to defer these uncorrected measurements to the end of the second session instead.

In the second session, participants were seated in the same way as the first session, then completed four variants of Curveball in a different random order for each subject:

- a third standard run;
- a close condition (47 cm viewing distance);
- a far condition (77 cm viewing distance); and
- a dark condition (room lights off, so the display was the only source of illumination).

The two participants who could not remove their corrective eyewear in the first session had their acuity measured and completed an uncorrected Curveball run after the above four variants were completed.

Results

In total, a full run of Curveball took an average time of 5 min and 15 s ($SD = 37$ s) across all observers and conditions. Results for the different metrics and experimental conditions are presented below.

Minimum required smooth tracking ability

The Curveball algorithm requires participants to smoothly track the noise target, and this pursuit behavior must be of sufficient quality to be distinguished from other eye movements (such as saccades) that provide much weaker evidence about target visibility. If a participant cannot pursue a given target smoothly enough to meet the algorithm's minimum requirement, the trial will end prematurely and their sensitivity to that target's spatial frequency will be underestimated (a false negative).

Curveball's analysis protocol accounts for dropped trials by discarding the worst half of thresholds obtained for each spatial frequency (two out of four). Some participants, however, may still track the target too poorly overall to compute any accurate or consistent estimate of sensitivity. We identified these participants by calculating the overall proportion of

frames in which each participant met the Curveball criterion for smooth tracking over all seven runs and conditions of the task (not including the eighth uncorrected-vision condition for applicable participants). We refer to this proportion as the *pursuit score* for that participant. The overall mean pursuit scores for all participants are depicted in Figure 2. Five participants (red triangles) were excluded due to having an overall mean pursuit score below 0.143 (i.e., one out of seven frames; dashed line). This was the minimum pursuit score required to prevent the Curveball trial from terminating. Most participants scored well above this threshold overall, as shown in Figure 2, but some participants fell below the exclusion threshold in specific experimental conditions and were excluded on a condition-by-condition basis. These additional exclusions are noted in the analyses below and the excluded threshold data are faded in corresponding figures. The reliance of Curveball on smooth tracking ability is discussed in further detail below.

One additional participant (red dot in Figure 2) was excluded because the eye tracker had difficulty integrating the gaze estimates from the participant's left and right eyes, which was likely due to strong asymmetric correction for a large astigmatism (cylindrical power of OD -1.25 , OS -5.75). This problem can be avoided in future studies by only taking monocular measurements for participants with incompatible eyewear. Other participants with smaller correction for astigmatism or bifocal lenses were not affected.

Relationship between smooth tracking ability and sensitivity thresholds

The Curveball procedure depends on a minimum quality of smooth tracking ability, but the contrast sensitivity thresholds it produces should not be strongly dependent on the precise quality of each participant's smooth eye movements beyond the required amount. This would suggest that the Curveball task was effectively only measuring smooth tracking ability. We tested this possibility by regressing mean sensitivity across the standard Curveball runs on pursuit score. Mean sensitivity was calculated simply as the average sensitivity thresholds of all six spatial frequencies (as no data were missing).

Mean sensitivity will naturally be related to overall pursuit score, as participants with better contrast sensitivity spend a greater proportion of time tracking the noise target instead of waiting for trials to terminate. We accounted for this conflating factor by only examining pursuit scores calculated over periods in which the noise target was likely to be visible to all participants: a spatial frequency of 1 cpd (the peak

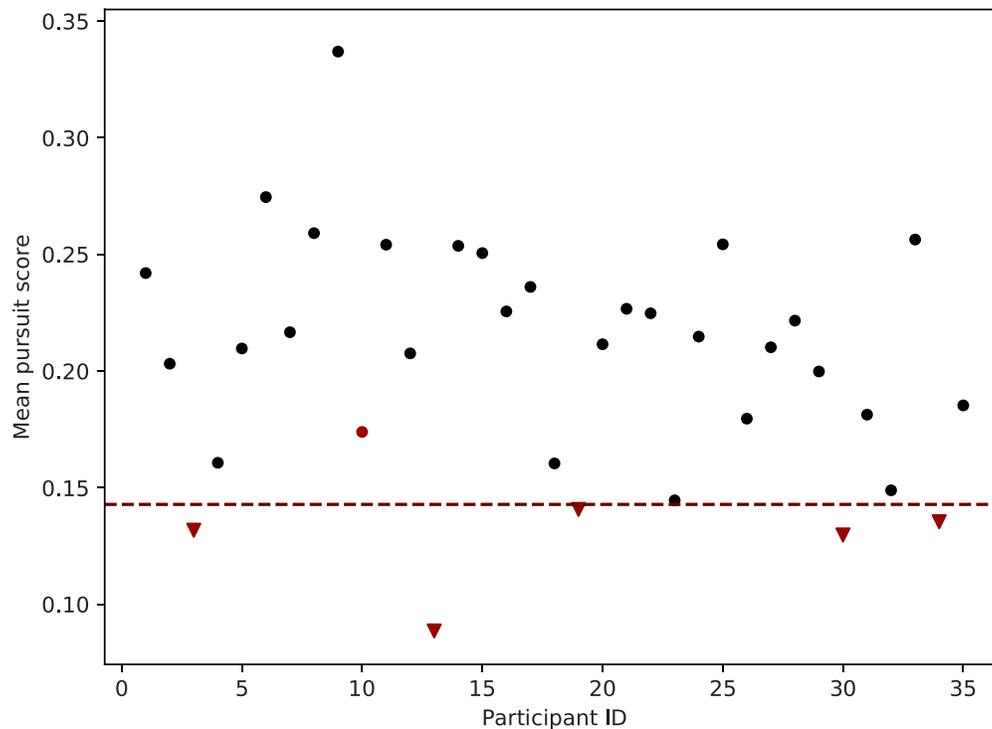


Figure 2. Scatterplot of Curveball mean pursuit scores. The red dashed line (pursuit score = 0.143) depicts the minimum pursuit score required for inclusion in the CSF results. Participants marked by red triangles fell below this threshold and were excluded from all analyses. The red dot represents a participant who was independently excluded due to an incompatibility between their asymmetric corrective lenses and the eye tracker.

sensitivity for most participants) and RMS contrast of 0.01 or greater (log sensitivity of 2). A linear regression revealed that mean sensitivity was weakly but significantly predicted by this pursuit score, $r = 0.488$, $p = 0.007$, with a large standard error of 0.851 log units of sensitivity (approximately half the height of a typical CSF curve). This indicates that participants who were better at smoothly tracking a highly visible target tended to achieve better contrast thresholds, but not to a strong degree. Participants who were better at tracking the target may have been slightly more likely to continue tracking for a short interval after its contrast was reduced below threshold (but before the target could change direction). Alternatively, smooth tracking ability and mean contrast sensitivity may be inherently related through some measure of general visual function. Our data are not able to inform these possibilities.

Repeatability of Curveball thresholds

Curveball CSFs from runs within the same session and across different sessions are depicted in Figure 3 and Figure 4, respectively. We analyzed the same-day repeatability of the standard Curveball task by comparing thresholds estimated during the first

Curveball run (performed before the 4AFC staircases) and the second (performed after) in the first experimental session. These thresholds are plotted together for each of the twenty-nine included participants in Figure 3. The horizontal axis in each subplot represents spatial frequency on a log scale and the vertical axis shows \log_{10} units of RMS contrast sensitivity. The limits and scale of the axes are identical in each subplot. All future figures of CSF data have the same layout and axes as Figure 3.

Same-day repeatability can be visualized for each spatial frequency in the Bland-Altman plot in the top panel of Figure 5 (Bland & Altman, 1986). Each of the six subplots depicts the difference in mean sensitivity to one spatial frequency between Curveball runs for each participant plotted against the mean of the two runs. The horizontal lines represent the 95% limits of agreement, and the spatial frequency and coefficient of repeatability are given above each plot. The overall same-day coefficient of repeatability was 0.275 log units of RMS contrast sensitivity when sensitivity was pooled across the spatial frequencies. Notably, a substantial proportion of the same-day variance was contributed by one participant who performed much better on their second run of the task. The overall coefficient of repeatability decreases to 0.236 if this outlying participant is discounted.

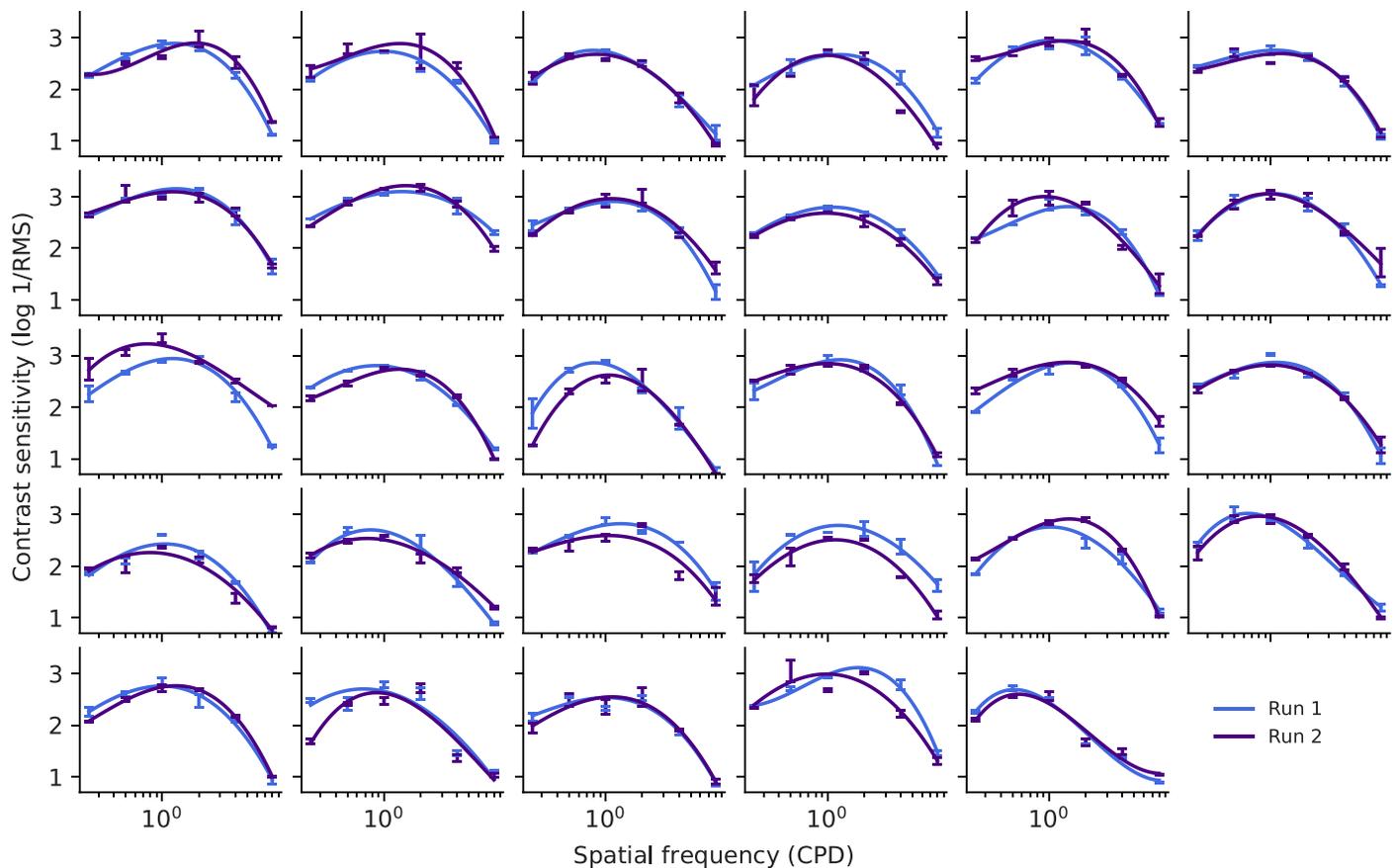


Figure 3. Same-day repeatability of Curveball mean thresholds. Each subplot depicts the CSFs for a participant obtained from the first (blue) and second (purple) Curveball runs within the first testing session. Horizontal axes represent spatial frequency in log scale and vertical axes represent log units of RMS contrast sensitivity. In this and subsequent figures, error bars represent ± 1 SEM and the smooth curves correspond to the optimal cubic polynomial fits.

Different-day repeatability was analyzed in an analogous way to same-day repeatability. Thresholds from the first Curveball run in the first session were compared against thresholds from the standard Curveball run in the second experiment for each participant (Figure 4). The faded subplots indicate two participants whose ability to smoothly pursue the stimulus fell below the exclusion threshold in the second session, which led us to exclude them from this analysis. The bottom panel of Figure 5 depicts separate Bland-Altman plots and coefficients of repeatability for each spatial frequency, with an identical layout to the same-day comparison in the top panel. The overall coefficient of repeatability was 0.227 log units of sensitivity, which is similar to the overall same-day repeatability. There was no significant correlation between mean sensitivity in the second session's run and the run's position (from first to fourth) in the random ordering of conditions in the second session, $r = 0.025$, $p = 0.897$, which suggests that participant fatigue did not affect outcomes in the second session. Furthermore, the within-run error bars are globally small, despite being calculated from just two samples

per point. This suggests that discarding the worst 50% of trials was sufficient to eliminate any false negatives in almost all cases, and that the estimates of contrast sensitivity are both valid and highly consistent within each run as well as between runs. (The grayed-out subplots in Figure 4 demonstrate cases in which this approach evidently did not remove all false negatives.) Overall, our analysis indicates that Curveball is a highly repeatable measure of contrast sensitivity, both within a single testing session and across different days.

Relationship with conventional 4AFC staircases

If Curveball is a valid measure of contrast sensitivity, the CSFs formed from its thresholds at different spatial frequencies should be closely related to the CSFs assessed using conventional report-driven psychophysics. As Curveball uses moving stimuli, this relationship is likely to involve both a horizontal and vertical shift in the CSF curve relative to the conventional static gratings (Burr & Ross, 1982), but the involvement of tracking behavior (which should stabilize the stimulus

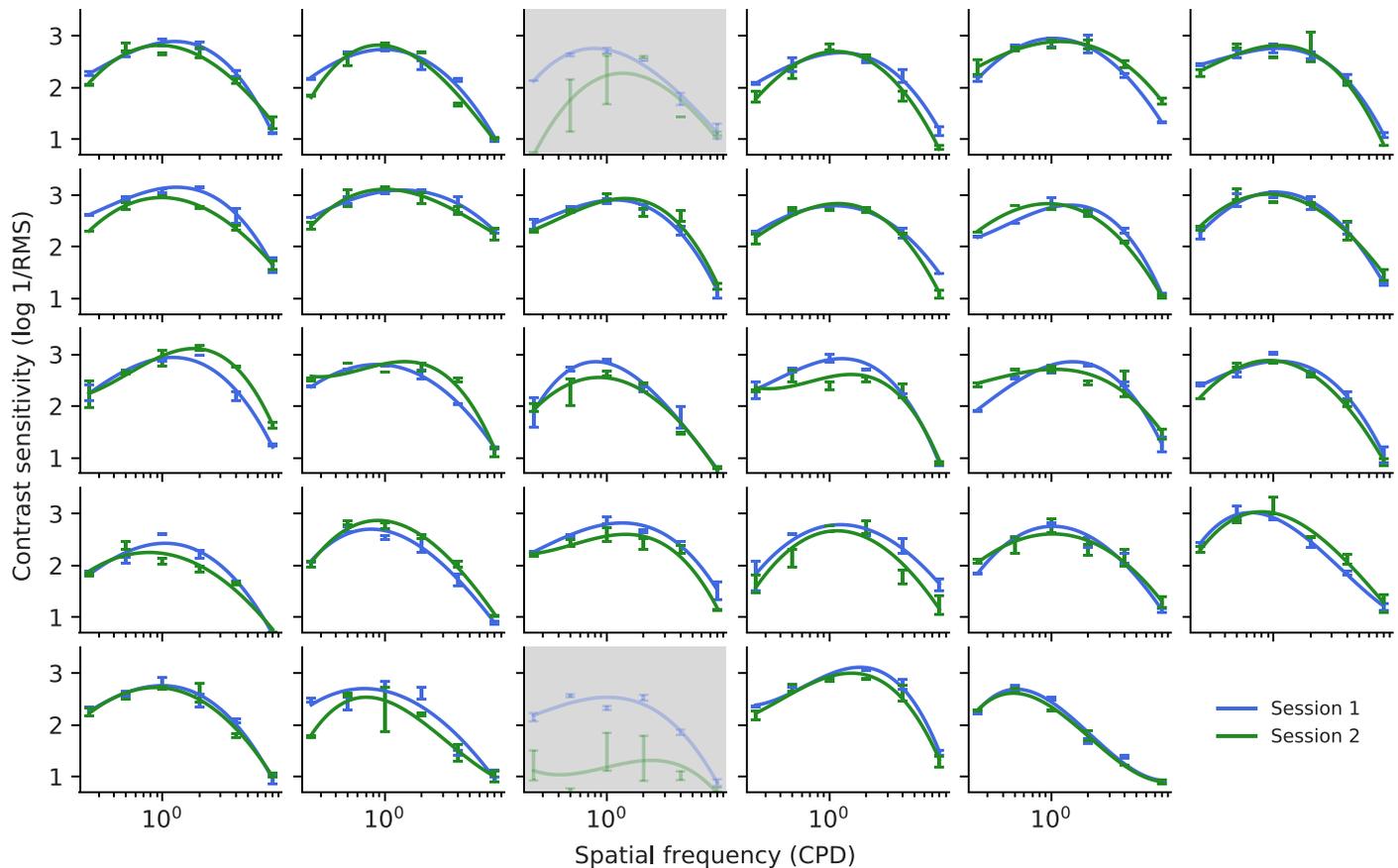


Figure 4. Different-day repeatability of Curveball mean thresholds. The layout is identical to Figure 3, but the lines now represent mean sensitivity in the first (blue) and second (green) testing sessions. The faded subplots represent two participants whose pursuit scores were below the exclusion threshold in the second session's standard Curveball run.

on the retina) and the nature of the filtered noise stimuli may shift the Curveball CSFs relative to the conventional moving gratings as well. We tested this relationship by comparing CSFs estimated using Curveball with CSFs obtained from the traditional 4AFC staircase task completed in the same session. Separate analyses were conducted for the static and moving gratings in the 4AFC task to determine how the CSFs produced by Curveball relate to each of them. One participant was excluded from the comparison with the static 4AFC thresholds due to a sensitivity outlier at 2 cpd, which was likely produced by a run of false positives from correct subthreshold guesses.

The correlations between the raw Curveball thresholds and static 4AFC thresholds are only moderate (mean correlation of 0.681 ± 0.170), but this is not surprising: past work has shown that the CSF elicited by moving stimuli is shifted down in spatial frequency (i.e., horizontally to the left) relative to the CSF for static stimuli (Burr & Ross, 1982). We accounted for this shift in peak sensitivity by allowing the Curveball thresholds to differ from the static 4AFC thresholds by up to an affine transformation. The change in correlation induced by this transformation represents the degree to which the

between-task differences are due to the “lens” of the task itself, rather than the underlying phenomenon being measured (e.g., the *shape* of the curve may be the same in both tasks). The scaling, shearing, and vertical offset parameters of the transformation for each participant were optimized over the pooled thresholds from the remaining 27 participants (i.e., a leave-one-out model). The raw (dotted blue) and transformed (solid blue) Curveball thresholds are plotted together with the static 4AFC thresholds (black) in Figure 6. The affine transformation significantly improved the mean correlation across participants to 0.790 ± 0.154 , $t(27) = 4.044$, $p < 0.001$. This indicates that the global affine transformation captured a large systematic difference in threshold estimates between the Curveball task and static 4AFC task, and that this difference comprised a large proportion of the between-tasks variance. It is not clear whether the remaining unexplained variance is indicative of measurement noise in one or both tasks, a non-affine task difference, or an intrinsic difference in the type of contrast sensitivity being measured by each technique.

The moving gratings in the 4AFC task were necessarily from a lower and more restricted range of

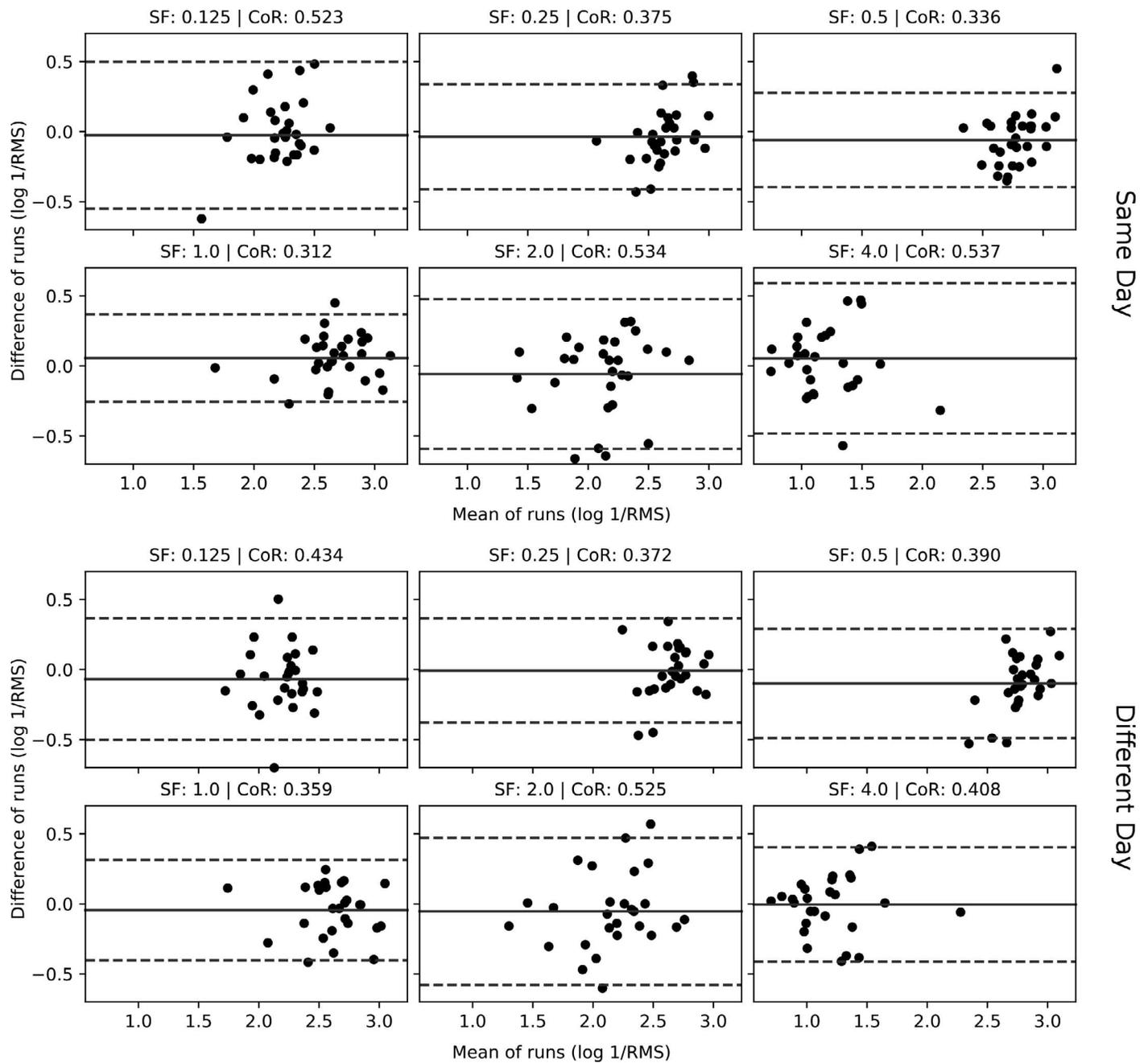


Figure 5. Bland-Altman plots for same-day and different-day repeatability. For each of the six spatial frequencies, contrast sensitivity in the first standard Curveball run from the first testing session has been compared with the second run from the same session (top) and the standard run from the second session on a different day (bottom). The vertical and horizontal axes represent the difference and mean of the two conditions compared in each case. The coefficient of repeatability (CoR) in log units of sensitivity is depicted above each plot, along with the spatial frequency in cycles per degree. The solid black line marks the mean difference and the dashed lines represent the 95% limits of agreement (mean \pm coefficient of repeatability).

spatial frequencies than the static gratings in the same task (which were not aliased by motion) or Curveball noise patches (which were filtered to avoid temporal aliasing). We accounted for this difference before comparing the moving 4AFC thresholds and Curveball thresholds by simply translating the Curveball thresholds to the left by one log unit (i.e., halving each spatial

frequency) and dropping the highest Curveball spatial frequency. This transformation alone was sufficient to determine that the shapes of the Curveball CSFs were highly correlated with the CSFs estimated from the moving gratings in the 4AFC (Figure 7). The mean correlation was 0.907 ± 0.074 and was significant at the 0.05 level for 19 out of 29 participants. Allowing for

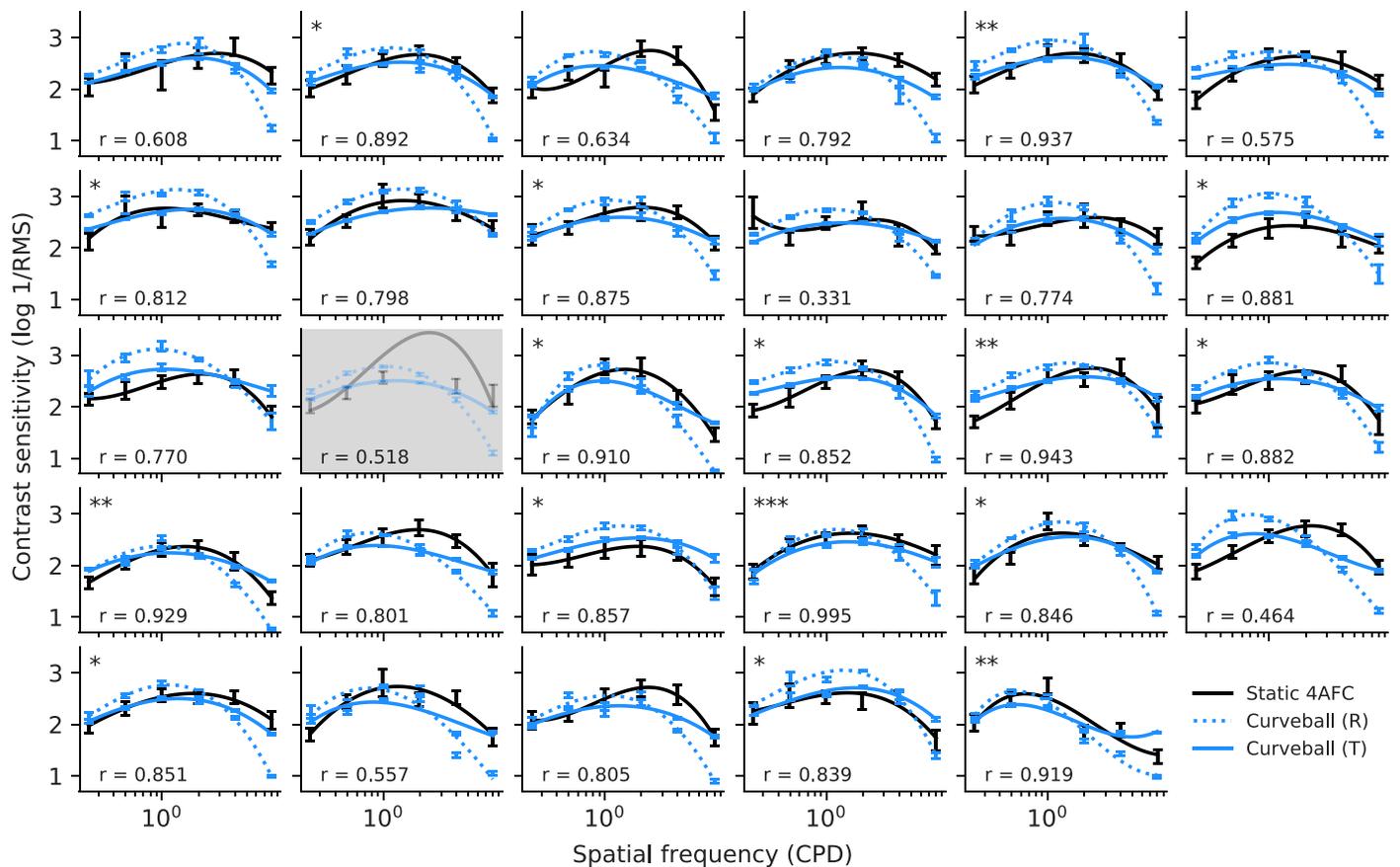


Figure 6. Curveball thresholds versus static 4AFC thresholds. The layout and axes are identical to Figure 3. The dotted and solid blue lines represent the raw and transformed Curveball CSFs, respectively, and the black lines represent the CSFs estimated from the static gratings in the 4AFC staircase task. The faded subplot in the third row denotes one excluded participant who achieved an impossible threshold at 2 cpd in the staircase task. This excluded data was not used to optimize the free transformation parameters for the other participants. Error bars in the 4AFC data here and in Figure 7 represent ± 1 SEM, which was computed using the last four between-reversal runs of each contrast staircase.

an additional global affine transformation in the same way as the static gratings had no significant impact on these correlations, $t(28) = -1.116$, $p = 0.274$.

Overall, these analyses indicate that CSFs obtained using Curveball are well predicted by thresholds obtained from both static and moving gratings in a 4AFC task, particularly after the systematic task-induced shift in the CSF is considered, which in turn suggests that Curveball is measuring the same underlying phenomenon of contrast sensitivity as the conventional tasks. Notably, Curveball's CSFs appear to fall between the curves elicited by static and moving stimuli in conventional discrete psychophysics.

Relationship with corrected acuity differences

The CSFs produced by Curveball should be sensitive to the differences in visual acuity induced by refractive correction. Specifically, participants' contrast sensitivity should decrease more rapidly as a function of spatial frequency as their acuity worsens (i.e., when they remove

their corrective lenses). If this is true, we would expect to find a relationship between the magnitude of the leftward shift in the CSF peak and the difference in eye chart letter acuity measured with and without visual correction. We examined this relationship for the 18 participants with corrected-to-normal vision who performed an additional standard Curveball run without their corrective eyewear. The uncorrected Curveball CSFs for these participants are depicted together with their standard corrected Curveball CSFs in Figure 8.

We quantified the effect of visual correction on the CSF with an affine transformation, but unlike the comparison between Curveball and the 4AFC task, separate transformations were optimized to account for the difference in corrected and uncorrected CSFs for each participant. The shear parameter of this transformation was then used as a measure of the change in the CSF curve: more negative shear indicates that the peak of the CSF shifted further to the left in the uncorrected condition relative to the corrected condition. A linear regression analysis revealed that uncorrected shear was

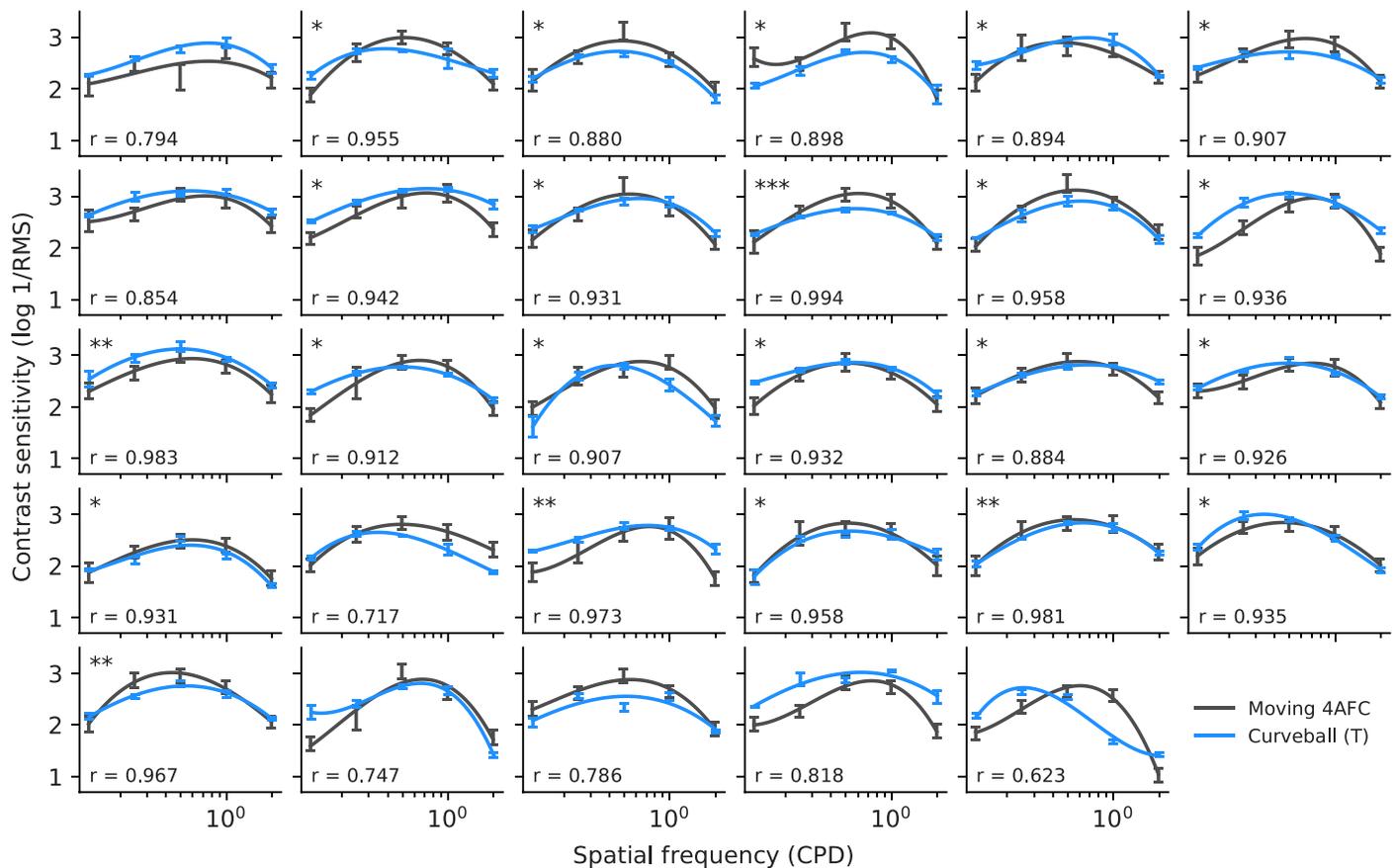


Figure 7. Curveball thresholds versus moving 4AFC thresholds. The layout is identical to Figure 6, but the horizontal axis of each subplot has been shifted to accommodate the lower range of spatial frequencies. The black and blue lines represent the CSFs estimated from moving gratings in the 4AFC staircase task and noise patches in Curveball, respectively. The Curveball spatial frequencies have been reduced by one octave, and the highest value removed, to account for the systematic horizontal shift in the CSF between tasks. No additional participants were excluded from this analysis.

highly and significantly predictive of the change in LogMAR letter acuity measured with the Tumbling E chart, $r = -0.890$, $p < 0.001$, in that more negative shear was associated with a larger loss of acuity from lack of corrective eyewear (as more positive LogMAR values represent worse vision). These data are shown with the line of best fit in Figure 9. Interestingly, the line of best fit was approximately $y = -x$ (slope of -1.153 and intercept of 0.041). These results indicate that Curveball is highly sensitive to changes in visual acuity and refractive correction, which is expected of a useful measure of spatial visual function. It may be possible to estimate an individual's letter acuity from a single Curveball CSF, but this would require an empirical investigation of their absolute relationship that is beyond the scope of this experiment.

Curveball thresholds at different distances

If Curveball is to be a useful measure of vision in a range of clinical settings, it is key that we understand

how dependent the procedure is on participant distance. We assessed the task's reliance on distance by comparing the thresholds and pursuit scores measured from the standard (62 cm), close (47 cm), and far (77 cm) Curveball conditions in the second testing session (Figure 10). The data suggest that deviations from the optimal eye tracker distance resulted in a greater number of dropped trials for several participants. Pursuit score fell below the exclusion threshold for two participants in the close condition and seven participants in the far condition. When these participants were excluded from the appropriate conditions, paired comparisons revealed that there was no significant difference in pursuit score between the standard and close conditions, $t(25) = -0.790$, $p = 0.437$, but there was a significant decrease of 0.018 in pursuit score in the far condition relative to the standard condition, $t(21) = 3.536$, $p = 0.002$. This suggests that the increase in distance in the far condition added enough noise to their gaze data to push multiple participants' smooth tracking ability below the level required for the Curveball algorithm to accurately estimate sensitivity.

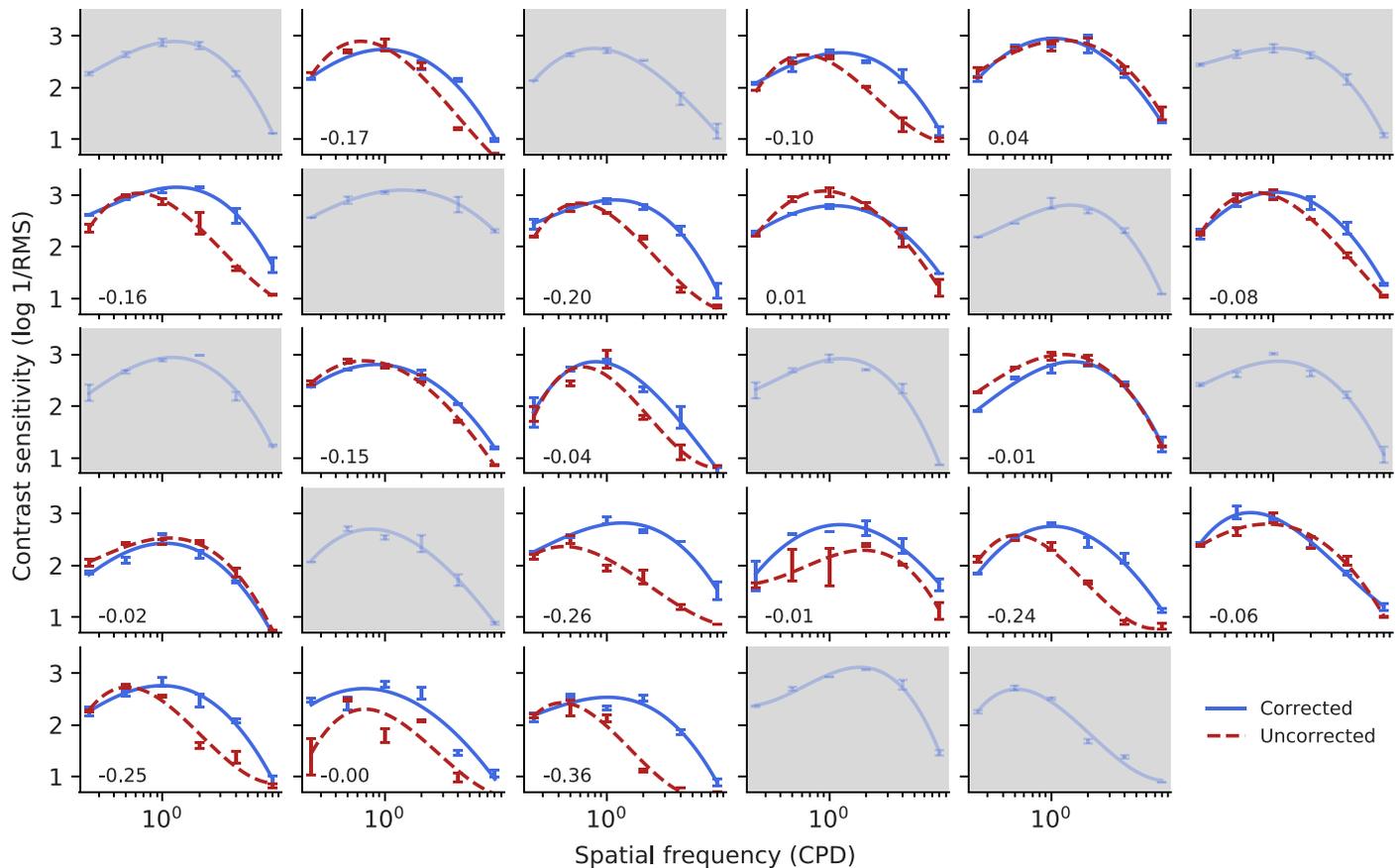


Figure 8. Curveball thresholds measured with and without visual correction. The layout and axes are identical to Figure 3. The solid blue and dashed red lines represent Curveball CSFs obtained with and without corrective eyewear, respectively. The shear in the uncorrected CSF curve relative to the corrected curve is given in the bottom-left corner of each subplot. The standard Curveball thresholds for participants who did not complete the uncorrected condition have been left grayed out in this figure to permit easy comparison across figures. No other participants were excluded.

A repeated measures ANOVA revealed no change in mean sensitivity between the standard and close distance conditions, $F(1, 26) = 0.499$, $p = 0.486$, but did reveal a significant interaction between distance and spatial frequency, $F(5, 130) = 3.036$, $p = 0.013$. A linear trend contrast found that the difference between the standard and close conditions became significantly more positive as a function of increasing log spatial frequency, $t(26) = 2.221$, $p = 0.035$. This is expected: moving closer to the display increases the actual spatial frequency of each stimulus in degrees of visual angle and should shift the CSF to the right, as the presented stimuli are identical.

An analogous repeated measures ANOVA found a significant decrease of 0.135 log units of RMS sensitivity in the far condition relative to the standard condition, $F(1, 20) = 38.981$, $p < 0.001$, but unlike in the close condition, there was no interaction between this distance change and spatial frequency, $F(5, 100) = 0.592$, $p = 0.706$. The expected leftward shift in the CSF may have been masked by the increase in eye tracker noise at greater distances. Participants

may have also found it more difficult to attend to the task in the far condition due to the screen's reduced presence in their field of view, which could explain the reduction in mean sensitivity.

Together, these results suggest that Curveball is more tolerant of decrements in user distance than increments relative to the optimal distance of 62 cm. This is likely a permanent limitation of display-mounted eye trackers, but its effect on the task may decrease as technology improves. For many participants, however, the task appears to remain reliable at a range of distances compatible with the display-mounted eye tracker.

Effect of room illumination on Curveball thresholds

Two participants were excluded from analysis of the dark condition due to a tracking score below the exclusion threshold in that condition. A subsequent

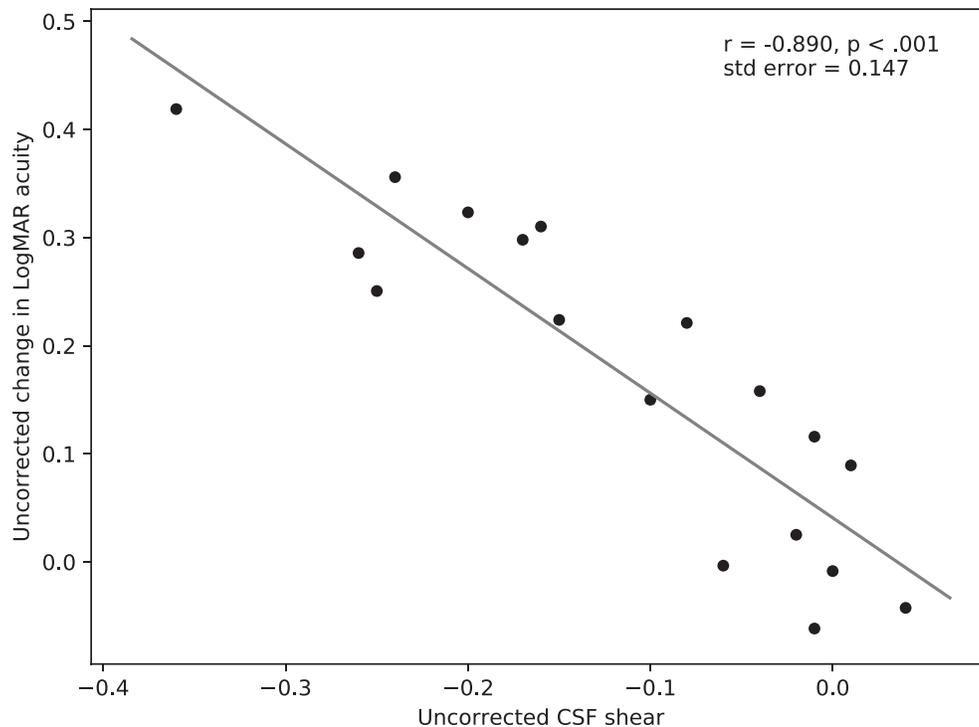


Figure 9. Regression plot of uncorrected change in LogMAR letter acuity against uncorrected CSF shear. Each point represents one of the 18 participants who completed a Curveball run without their corrective eyewear. The change in acuity was significantly predicted by the shear of the uncorrected CSF relative to the corrected CSF in Curveball, in that larger shear values were associated with a greater loss of acuity (i.e. a more positive LogMAR value). The coefficient, p -value, and standard error of the correlation are depicted in the top-right corner and the gray line represents the line of best fit.

repeated measures ANOVA revealed that turning off the room lights had a small significant positive effect on mean sensitivity relative to the standard lights-on Curveball run conducted in the same testing session, $F(1, 26) = 4.670$, $p = 0.040$, but no significant interaction between the change in illumination and spatial frequency, $F(5, 130) = 0.944$, $p = 0.455$. These results suggest that a large change in room illumination (a decrease of 10 cd/m^2) has a minimal effect on Curveball performance. CSFs for the dark condition are not depicted due to their high similarity to the curves from the standard conditions.

Discussion

Our findings provide strong evidence that Curveball is a reliable, accurate, and efficient objective measure of contrast sensitivity at near distances. Task repeatability was high, both within the same session (coefficient of repeatability 0.275) and across different days (coefficient of repeatability 0.227), and its consistency across changes in room illumination suggest that it is suitable for practical clinical settings. The procedure produces CSFs that are (a) systematically related to the CSFs

obtained from both static and moving stimuli in a conventional staircase task and (b) highly predictive of the difference between corrected and uncorrected eye chart letter acuity. Curveball contrast sensitivity estimates are distorted in a predictable way as the user moves closer to the screen and the algorithm's ability to detect smooth tracking appears to degrade only gradually as distance from the eye tracker varies between the optimal and maximum distance allowed by the hardware. This suggests that the participant's distance can be continuously monitored using the eye tracker and used to compute the true spatial frequencies being measured in each trial when estimating the CSF, provided the eye tracker's estimates of user distance are sufficiently accurate and frequently updated. The display-mounted eye tracker used here required only half a second of one-point calibration at the start of the task for our smooth tracking detection algorithm to perform well. Tracking was measured binocularly, and our results suggest that this is a valid means for assessing contrast sensitivity in participants who will not tolerate the apparatus needed for monocular measurement (e.g., an eye patch or glasses). However, there is no reason to expect the procedure will be negatively impacted when used monocularly, as no change occurs in the hardware or software other

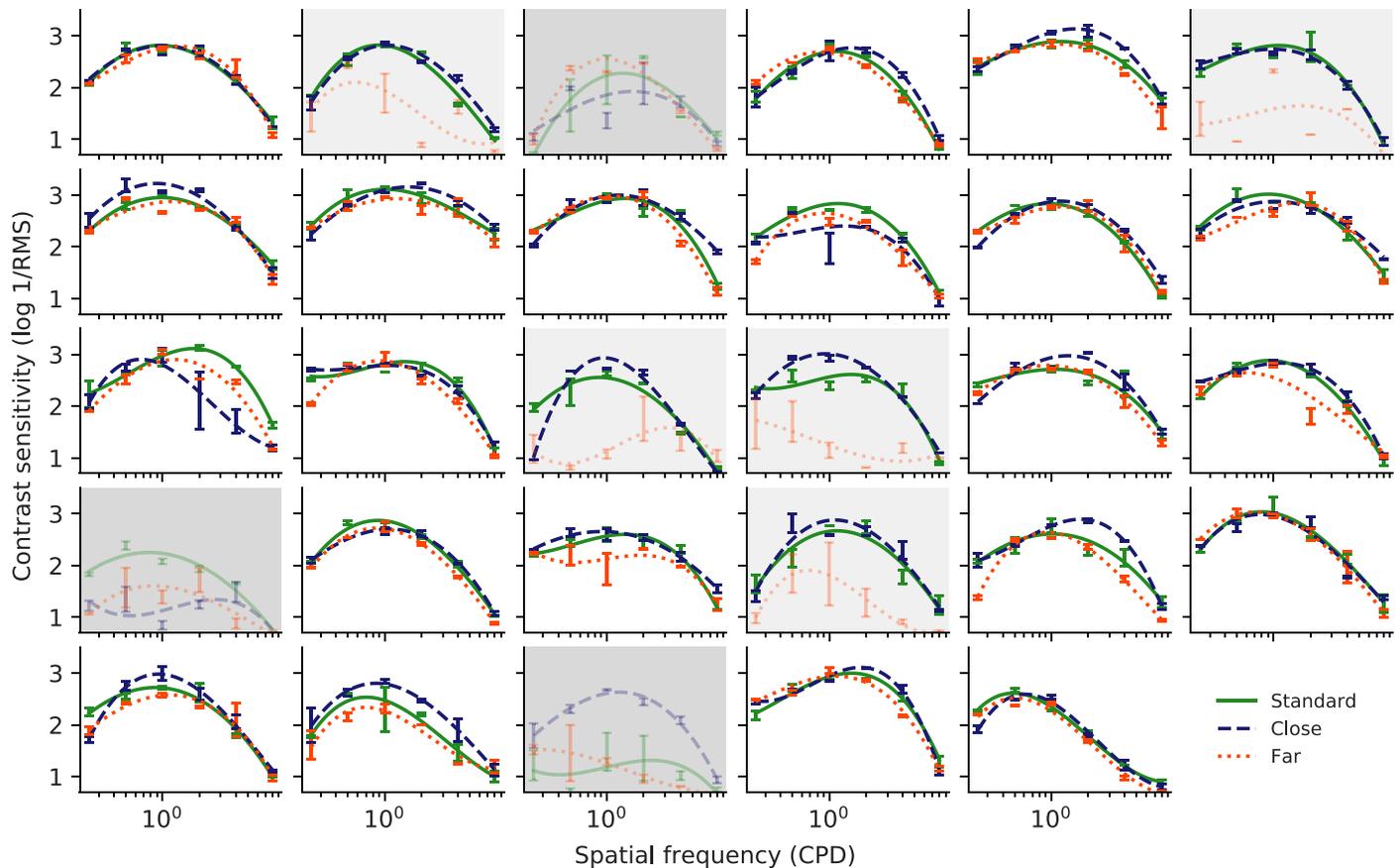


Figure 10. Curveball thresholds measured at different distances. The layout and axes are identical to Figure 3, but correlations are not shown. The solid green, dashed blue, and dotted orange lines represent CSFs measured at standard (62 cm), close (47 cm), and far (77 cm) distances, respectively. Note that the spatial frequency units on the horizontal axis are only correct for the standard distance of 62 cm. The semi-grayed subplots indicate participants who were excluded from the far condition analysis due to an insufficient pursuit score in that condition. The fully grayed subplots indicate participants who were excluded from both distance analyses due to insufficient pursuit scores in at least two of the three conditions.

than the source of the eye tracker data. Separate monocular assessment remains the ideal means of measuring contrast sensitivity.

Critically, Curveball requires no volitional perceptual report and can potentially be administered with no instruction. Many participants reported that it was easier and more engaging than the conventional staircase task and indicated that they preferred the second Curveball-only testing session. Most importantly, the task is no less efficient than the best existing procedures based on perceptual report—even those that use Bayesian statistics and CSF curve parameterization (e.g., Lesmes et al., 2010)—and is potentially more efficient due to its allowance of a flexible number of repeats per threshold. A single threshold estimate for one spatial frequency takes less than 10 s to obtain, and the precision of that estimate rapidly improves as additional repeats are conducted and dropped trials discarded. These dropped trials are likely to cause the trial to end much earlier than it otherwise would, and future implementations of Curveball could potentially

detect these false negatives and respond by adapting the number of repeats needed for that spatial frequency in real time. For example, participants who exhibit a sufficiently low difference between the first two repeats of a given threshold, in addition to a sufficiently high pursuit score, could skip the third and fourth repeats at that spatial frequency.

Curveball minimizes the occurrence of false positives by being contingent on the presence of pursuit-based tracking, but as with all psychophysical tasks, its reliance on an overt motor feedback (eye movements here, rather than verbal communication or key-pressing) introduces the possibility of confounds. Five out of 36 participants (14%) were excluded from our analysis because they did not exhibit overall smooth tracking performance above the minimum consistency needed for Curveball trials to persist, and additional participants were excluded from certain parts of the analysis based on their smooth tracking ability in specific conditions. The data suggest that these participants could not or would not track the noise

target smoothly enough and/or frequently enough for the Curveball algorithm to reliably infer stimulus visibility and adaptively reduce its contrast. These participants were consequently likely to drop too many trials to compute a reliable threshold estimate, particularly in the far condition, where eye tracker noise was likely greater. The result is categorical failure: it is almost always obvious from the data in the figures above where participants dropped more than half of the trials for one or more spatial frequencies. Overestimations of sensitivity, in comparison, were entirely absent from the experiment (other than the systematic distortions in sensitivity produced by the distance variants).

Curveball was developed primarily based on the needs of cognitively impaired populations for a robust, objective, and fast measurement of contrast sensitivity, and it is likely that some of these individuals will have similar difficulties pursuing the Curveball target as seen in several participants here. We are currently investigating several possible techniques to reduce false negatives from poor tracking without compromising Curveball's low false positive rate. Widening the algorithm's definition of "tracking" could cause incidental saccades consistent with the direction of the target to trigger a contrast reduction, and the definition would need to be widened considerably to capture the repeated catch-up saccades that these participants are likely exhibiting in place of smooth eye movements. The noise stimulus could be converted from a windowed patch to a full-screen pattern to elicit the optokinetic nystagmus reflex, as in previous work (Dakin & Turnbull, 2016; Suner, Prusky, Carmel, & Hill, 2015), but this would preclude the algorithm from using gaze position as part of its detection algorithm and would therefore also be likely to promote false positives due to the loosened success criteria. Participants would additionally need to continually readjust their point of fixation whenever it happened to fall off the screen. The Curveball algorithm could potentially be modified to detect saccade-heavy following behavior and exploit it as evidence of visibility in the same way as smooth tracking, but such an algorithm would need to accomplish the difficult task of reliably distinguishing true catch-up saccades from incidental eye movements that happen to land on the moving target. Our present findings suggest that this is an avenue now worth investigating, given Curveball's success at rapidly measuring contrast thresholds in most healthy adults. Until this can be achieved, both healthy and impaired participants who have difficulty exhibiting smooth tracking may need to simply perform more repeats at each spatial frequency (potentially over multiple sessions) to produce the same quantity of accurate threshold estimates as participants who can consistently track the target. Our initial testing sessions with brain-injured children, however, indicate that there is a large population of impaired individuals with intact smooth

pursuits who will benefit from a task that circumvents their cognitive and communicative impairments.

A final advantage of Curveball (and gaze-based tasks in general) that deserves mention is the ability to extract other information about the participant's visual function from the eye tracking data collected during the procedure. This could make the task even more useful for testing participants with brain injury or other cognitive impairments, as these individuals are likely to exhibit low-level ocular or cortical dysfunction that can be measured from Curveball even if accurate contrast thresholds cannot be obtained. The ability to smoothly pursue a target, for example, is a useful dimension of visual function that Curveball already exploits to determine stimulus visibility. Curveball data could be further leveraged to determine how pursuits and saccades depend on stimulus orientation, movement direction, and location in the visual field, all of which naturally vary as the target moves around the display. Catch-up saccade latency could be inferred from the participant's response when the target appears at the start of a new trial or abruptly rebounds off the edge of the display. Specific dysfunctions, such as pathological nystagmus, could also be detected and quantified from the gaze data. It may even be possible to detect which Curveball false negatives were caused by inattention by examining when the participant was and was not looking at the screen. We are currently investigating these avenues and are designing additional objective continuous psychophysics tasks that could complement Curveball and help form a gaze-driven assessment of visual ability across all populations, both healthy and impaired.

Keywords: contrast sensitivity function, eye tracking, smooth pursuit, continuous psychophysics, curveball

Acknowledgments

This research was funded by the National Institute of Health (award EY026753), Blythedale Children's Hospital, and the National Center for Advancing Translational Sciences (award UL1TR002384).

Commercial relationships: none.

Corresponding author: Scott W. J. Mooney.

Email: scm2011@med.cornell.edu.

Address: Burke Neurological Institute, White Plains, NY, USA.

References

Anderson, M., Motta, R., Chandrasekar, S., & Stokes, M. (1996, January). Proposal for a standard default

- color space for the internet—sRGB. In *Color and Imaging Conference, 4th Color and Imaging Conference, Final Program and Proceedings (Vol. 1996, No. 1, pp. 238–245)*. Scottsdale, AZ: Society for Imaging Science and Technology.
- Allard, R., & Faubert, J. (2008). The noisy-bit method for digital displays: Converting a 256 luminance resolution into a continuous resolution. *Behavior Research Methods, 40*(3), 735–743.
- Bahill, A. T., & McDonald, J. D. (1983). Smooth pursuit eye movements in response to predictable target motions. *Vision Research, 23*(12), 1573–1583.
- Barnes, G. R. (2008). Cognitive processes involved in smooth pursuit eye movements. *Brain and Cognition, 68*(3), 309–326.
- Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 327*(8476), 307–310.
- Bodis-Wollner, I. (1972, November 17). Visual acuity and contrast sensitivity in patients with cerebral lesions. *Science, 178*(4062), 769–771.
- Bodis-Wollner, I., Onofrij, M. C., Marx, M. S., & Mylin, L. H. (1986). Visual evoked potentials in Parkinson's disease: Spatial frequency, temporal rate, contrast, and the effect of dopaminergic drugs. *Frontiers of Clinical Neuroscience, 3*, 307–319.
- Bonnen, K., Burge, J., Yates, J., Pillow, J., & Cormack, L. K. (2015). Continuous psychophysics: Target-tracking to measure visual sensitivity. *Journal of Vision, 15*(3):14, 1–16, <https://doi.org/10.1167/15.3.14>. [PubMed] [Article]
- Burr, D. C., & Ross, J. (1982). Contrast sensitivity at high velocities. *Vision Research, 22*(4), 479–484.
- Catford, G. V., & Oliver, A. (1973). Development of visual acuity. *Archives of Disease in Childhood, 48*(1), 47–50.
- Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision, 9*(12):10, 1–15, <https://doi.org/10.1167/9.12.10>. [PubMed] [Article]
- Chan, J. W., Edwards, M. H., Woo, G. C., & Woo, V. C. (2002). Contrast sensitivity after laser in situ keratomileusis: One-year follow-up. *Journal of Cataract & Refractive Surgery, 28*(10), 1774–1779.
- Cimmer, C., Szendi, I., Csifcsák, G., Szekeres, G., Kovács, Z. A., Somogyi, I., ... Kéri, S. (2006). Abnormal neurological signs, visual contrast sensitivity, and the deficit syndrome of schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 30*(7), 1225–1230.
- Cohen, B., Matsuo, V., & Raphan, T. (1977). Quantitative analysis of the velocity characteristics of optokinetic nystagmus and optokinetic after-nystagmus. *The Journal of Physiology, 270*(2), 321–344.
- Collins, J. W., & Carney, L. G. (1990). Visual performance in high myopia. *Current Eye Research, 9*(3), 217–224.
- Dakin, S. C., & Turnbull, P. R. (2016). Similar contrast sensitivity functions measured using psychophysics and optokinetic nystagmus. *Scientific Reports, 6*: 34514, 1–14.
- de Faria, J. M. L., Katsumi, O., Arai, M., & Hirose, T. (1998). Objective measurement of contrast sensitivity function using contrast sweep visual evoked responses. *British Journal of Ophthalmology, 82*(2), 168–173.
- Dobson, V., & Teller, D. Y. (1978). Visual acuity in human infants: A review and comparison of behavioral and electrophysiological studies. *Vision Research, 18*(11), 1469–1483.
- Dorr, M., Lesmes, L. A., Elze, T., Wang, H., Lu, Z. L., & Bex, P. J. (2017). Evaluation of the precision of contrast sensitivity function assessment on a tablet device. *Scientific Reports, 7*:46706, 1–11.
- Dorr, M., Lesmes, L. A., Lu, Z. L., & Bex, P. J. (2013). Rapid and reliable assessment of the contrast sensitivity function on an iPad. *Investigative Ophthalmology & Visual Science, 54*(12), 7266–7273.
- Douglas, R. M., Alam, N. M., Silver, B. D., McGill, T. J., Tschetter, W. W., & Prusky, G. T. (2005). Independent visual threshold measurements in the two eyes of freely moving rats and mice using a virtual-reality optokinetic system. *Visual Neuroscience, 22*(5), 677–684.
- Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision, 8*(14):18, 1–26, <https://doi.org/10.1167/8.14.18>. [PubMed] [Article]
- Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research, 43*(9), 1035–1045.
- Gibaldi, A., Vanegas, M., Bex, P. J., & Maiello, G. (2017). Evaluation of the Tobii EyeX eye tracking controller and Matlab toolkit for research. *Behavior Research Methods, 49*(3), 923–946.
- Hill, N. J., Mooney, S. W. J., Ryklin, E. B., & Prusky, G. T. (2018). *Shady: A programmer's toolbox and engine for real-time manipulation of visual stimuli*. Manuscript in preparation.
- Kaernbach, C. (1991). Simple adaptive testing with the

- weighted up-down method. *Attention, Perception, & Psychophysics*, 49(3), 227–229.
- Katsumi, O., Denno, S., Arai, M., de Faria, J. D. L., & Hirose, T. (1997). Comparison of preferential looking acuity and pattern reversal visual evoked response acuity in pediatric patients. *Graefes' Archive for Clinical and Experimental Ophthalmology*, 235(11), 684–690.
- Kleiner, R. C., Enger, C., Alexander, M. F., & Fine, S. L. (1988). Contrast sensitivity in age-related macular degeneration. *Archives of Ophthalmology*, 106(1), 55–57.
- Leat, S. J., & Woo, G. C. (1997). The validity of current clinical tests of contrast sensitivity and their ability to predict reading speed in low vision. *Eye*, 11(6), 893–899.
- Leat, S. J., Yadav, N. K., & Irving, E. L. (2009). Development of visual acuity and contrast sensitivity in children. *Journal of Optometry*, 2(1), 19–26.
- Lesmes, L. A., Jeon, S. T., Lu, Z. L., & Doshier, B. A. (2006). Bayesian adaptive estimation of threshold versus contrast external noise functions: The quick TvC method. *Vision Research*, 46(19), 3160–3176.
- Lesmes, L. A., Lu, Z. L., Baek, J., & Albright, T. D. (2010). Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method. *Journal of Vision*, 10(3):17, 1–21, <https://doi.org/10.1167/10.3.17>. [PubMed] [Article]
- Pelli, D. G., & Bex, P. (2013). Measuring contrast sensitivity. *Vision Research*, 90, 10–14.
- Pelli, D. G., Robson, J. G., & Wilkins, A. J. (1988). The design of a new letter chart for measuring contrast sensitivity. *Clinical Vision Sciences*, 2(3), 187–199.
- Pelli, D. G., Rubin, G. S., & Legge, G. E. (1986). Predicting the contrast sensitivity of low vision observers (A). *Journal of the Optical Society of America A*, 3, 56–56.
- Prusky, G. T., Alam, N. M., Beekman, S., & Douglas, R. M. (2004). Rapid quantification of adult and developing mouse spatial vision using a virtual optomotor system. *Investigative Ophthalmology & Visual Science*, 45(12), 4611–4616.
- Regan, D., Raymond, J., Ginsburg, A. P., & Murray, T. J. (1981). Contrast sensitivity, visual acuity and the discrimination of Snellen letters in multiple sclerosis. *Brain*, 104(2), 333–350.
- Riddell, P. M., Ladenheim, B., Mast, J., Catalano, T., Nobile, R., & Hainline, L. (1997). Comparison of measures of visual acuity in infants: Teller acuity cards and sweep visual evoked potentials. *Optometry and Vision Science*, 74(9), 702–707.
- Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision*, 11(5):9, 1–30, <https://doi.org/10.1167/11.5.9>. [PubMed] [Article]
- Suner, M., Prusky, G. T., Carmel, J. B., & Hill, N. J. (2017). Longitudinal quantification of eye-movement impairments after pontine hemorrhage. *Frontiers in Neurology*, 8: 165, 1–8.
- Taylor, H. R. (1978). Applying new design principles to the construction of an illiterate E chart. *American Journal of Optometry and Physiological Optics*, 55(5), 348–351.
- Teller, D. Y., McDonald, M. A., Preston, K., Sebris, S. L., & Dobson, V. (1986). Assessment of visual acuity in infants and children: The acuity card procedure. *Developmental Medicine & Child Neurology*, 28(6), 779–789.
- Vul, E., Bergsma, J., & MacLeod, D. I. (2010). Functional adaptive sequential testing. *Seeing and Perceiving*, 23(5), 483–515.
- Witton, C., Talcott, J. B., & Henning, G. B. (2017). Psychophysical measurements in children: Challenges, pitfalls, and considerations. *PeerJ*, 5:e3231, 1–22.
- Woo, G. C. (1985). Contrast sensitivity function as a diagnostic tool in low vision. *American Journal of Optometry and Physiological Optics*, 62(9), 648–651.

Supplementary Material

- Supplementary Movie S1.** Curveball task.
Supplementary Movie S2. 4AFC task.