Vision Research 208 (2023) 108222

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Visual motion discrimination experiments reveal small differences between males and females

Omar Bachtoula[a], Sandra Arranz-Paraíso[a], Raúl Luna[a,b], Ignacio Serrano-Pedraza[a,c,*]

[a] Department of Experimental Psychology, Faculty of Psychology, Universidad Complutense de Madrid, Madrid, 28223, Spain
[b] Institute of Optics, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, 28006, Spain
[c] Centre for Behaviour and Evolution, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK

ARTICLE INFO

Keywords:
Sex differences
Motion discrimination
Duration thresholds
Contrast thresholds
Surround suppression

ABSTRACT

Recent results have shown that males have lower duration thresholds for motion direction discrimination than females. Measuring contrast thresholds, a previous study has shown that males have a greater sensitivity to fine details and fast flickering stimuli than females, and that females have a higher sensitivity to low spatial frequencies modulated at low temporal frequencies. Here, we present the data of a contrast-detection motion discrimination experiment and a reanalysis of four different motion discrimination experiments where we compare duration thresholds for males and females using different spatial frequencies, stimulus sizes, contrasts, and temporal frequencies (in two experiments, motion surround suppression was measured). Results from the main experiment and the reanalysis show that, in general, the association between sex and contrast and duration thresholds for motion discrimination is not significant, with males and females showing similar data patterns. Only the reanalysis of one out of four studies revealed different duration thresholds between males and females paired with a strong effect size supporting previous results in the literature, although motion surround suppression was identical between groups. Importantly, most of our results do not show significant differences between males and females in contrast and duration thresholds, suggesting that the sex variable may not be as relevant as previously claimed when testing visual motion discrimination.

1. Introduction

Recent studies have pointed out that sex is a biological variable that should be controlled in neuroscience studies (Cahill, 2006; McCarthy et al., 2017) and particularly in vision research (Abramov et al., 2012a, 2012b; Mathew et al., 2020; S. O. Murray et al., 2018). Previous studies have found sex differences depending on the type of visual function tested. For example, males have better visual acuity (for static and dynamic targets) than females (Burg, 1966; Burg & Hulbert, 1961; McGuinness, 1976). Contrast sensitivity is in general higher for males than females (Abramov et al., 2012a), but it depends on stimulus spatial frequency and orientation (Brabyn & McGuinness, 1979). Females have lower scotopic thresholds and present a longer visual persistence in the dark than males (McGuinness, 1976). Regarding color perception, there are sex differences in many aspects, like color identification (Greene & Gynther, 1995), hue sensations (Abramov et al., 2012b), color saturation in the green-yellow region of color space (I. J. Murray et al., 2012), and color categorization (Fider & Komarova, 2019). In a recent study, Mathew et al., (2020) have shown differences in visuomotor processing between males and females. In particular, a male advantage for visuomotor tracking was found (Mathew et al., 2020). Regarding spatiotemporal contrast sensitivity using sinusoidal gratings, Abramov et al. (2012a) showed that males and females have similar sensitivity to low spatial frequencies, but as spatial frequency increases, sensitivity becomes higher for males. This difference in sensitivity is also greater for temporal frequencies between 1 and 8 Hz. However, there are other visual aspects where no sex differences have been evidenced. For example, no significant differences were found in stereoacuity between males and females (Zaroff et al., 2003).

Regarding motion perception, Brabyn and Mcguinness (1979) measured contrast thresholds for drifting gratings of different spatial frequencies and three orientations at a speed of 10 deg/s. Females had lower contrast thresholds for low spatial frequencies (e.g. lower than 0.8 c/deg), whereas males had lower contrast thresholds for high spatial frequencies (e.g. higher than 8 c/deg), independently of stimulus orientation. No sex differences were found for the middle-range spatial

---

* Corresponding author at: Department of Experimental Psychology, Faculty of Psychology, Universidad Complutense de Madrid, Madrid 28223, Spain.
E-mail address: iserrano@ucm.es (I. Serrano-Pedraza).

https://doi.org/10.1016/j.visres.2023.108222
Received 10 December 2022; Received in revised form 10 March 2023; Accepted 15 March 2023

frequencies. In a recent study, Murray et al. (2018) showed large differences in motion discrimination of suprathreshold stimuli between males and females measuring motion duration thresholds (e.g. minimum time needed to discriminate the correct direction of motion for a defined level of performance). They found significantly shorter duration thresholds for male than for female participants, this difference being larger at high contrasts. This could not be explained by sex differences in speed of visual processing or motor responses. The authors also measured fMRI responses in area MT+ trying to identify the potential neural mechanisms causing the behavioral sex differences. Interestingly, they found that neural responses in area MT+ were associated with behavioral performance (i.e. duration thresholds correlated with MT+ fMRI response magnitude). However, they found no sex differences in the fMRI response amplitude in area MT+. That is, the amplitude of the response in MT+ was identical for males and females.

In summary, regarding motion perception, the results measuring contrast thresholds show that males have a higher sensitivity to fast-moving high spatial frequencies and females have higher sensitivities to slow-moving low spatial frequencies (Abramov et al., 2012a; Brabyn & McGuinness, 1979). For suprathreshold stimuli, males show shorter duration thresholds than females, and these differences are stronger for high contrasts independently of stimulus size (Murray et al., 2018).

In our study, we present the results of a motion direction discrimination experiment (data available in the supplementary S1 Table) measuring contrast thresholds for males and females. We test whether males have a higher sensitivity to fast-moving high spatial frequencies and females to slow-moving low spatial frequencies. This is relevant considering that in the study of Murray et al. (2018), they measured duration thresholds using stimuli with a narrow range of spatial frequencies (1–1.2 c/deg) and speeds (4–4.8 deg/s). Additionally, we present a reanalysis of three published studies from our laboratory where duration thresholds for motion direction discrimination were also obtained. In these reanalyses we compare males and females across a broader range of spatial frequencies, different contrasts, and temporal frequencies. The data are ordered by date of publication. The first set of data (i.e. Cohort 1, S2 Table) is taken from Arranz-Paraíso and Serrano-Pedraza (2018), where they measured duration thresholds in order to estimate a surround suppression index and correlate it with intelligence. The second set of data (i.e. Cohort 2, S3 and S4 Table) is taken from Luna and Serrano-Pedraza (2020), where duration thresholds for different spatial frequencies, temporal frequencies, and contrasts, were measured. This Cohort 2 is divided into two experiments, one with different contrast conditions and another with different temporal frequency conditions. The third set of data (i.e. Cohort 3, S5 Table) is taken from Arranz-Paraíso et al. (2021), where the authors measured duration thresholds and motion surround suppression for different viewing conditions (e.g. monocular and binocular).

The results from our main experiment show that contrast thresholds for males and females are not significantly different for low and high spatial frequencies for both low and high temporal frequencies. The reanalysis of the data from Cohorts 1, 2, and 3, shows that the shape of the results and the strength of surround suppression (obtained in Cohort 1 and 3) are similar for males and females. In general, duration thresholds are slightly lower for males than for females, although these differences are only significant in Cohort 3. Therefore, our results do not provide additional evidence to consider sex as a relevant variable when testing motion direction discrimination.

2. Methods

2.1. Subjects

Participants' visual acuity was tested using the SLOAN ETDRS 2000 letter series acuity chart at 40 cm and 300 cm. Their stereoscopic visual acuity was measured with the Frisby Stereotest (Cohorts 1 and 3) or the Randot Stereotest (Graded circle test) at 40 cm (Main Experiment and Cohort 2). All participants provided written informed consent before conducting the experiments and the Ethics Committee of Universidad Complutense de Madrid (Faculty of Psychology) approved all the experimental procedures performed. These experimental procedures comply with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

In the main experiment, we tested 30 participants (15 females and 15 males) with ages ranging from 20 to 32 years old (mean ± SD, all participants: 22.33 ± 2.25, females: 21.87 ± 0.92, males: 22.8 ± 3.03, Mann-Whitney U test for differences in age between females and males: $U = 108$, $p = 0.862$). Participants were unaware of the purpose of the research. All participants had a visual acuity lower than 0.4 logMAR on both eyes, tested both at 40 cm and 300 cm. Their stereoacuity was lower than 200 arcsec.

For Cohort 1, we analyzed the data of 47 participants (31 females and 16 males) with ages ranging from 18 to 28 years old (mean ± SD, all participants: 20.83 ± 2.49, females: 20.48 ± 2.67, males: 21.50 ± 2, Mann-Whitney U test: $U = 166$, $p = 0.064$). All participants had a visual acuity lower than 0.5 logMAR (in both eyes) and their stereoacuity was lower than 500 arcsec (Arranz-Paraíso & Serrano-Pedraza, 2018).

In the case of Cohort 2, we analyzed the data from two experiments. In the first experiment (Cohort 2a) we had 26 participants (17 females and 9 males) with ages between 18 and 26 years old (mean ± SD, all participants: 20.31 ± 2, females: 19.76 ± 1.44, males: 21.33 ± 2.55, Mann-Whitney U test: $U = 47.5$, $p = 0.115$). Their visual acuity was lower than 0.4 logMAR in all cases and their stereoacuity was lower than 200 arcsec. In the second experiment (Cohort 2b), we analyzed the data of 30 participants (18 females and 12 males) with ages between 18 and 26 years old in the 2 Hz condition (mean ± SD, all participants: 21 ± 2.88, females: 20.11 ± 2.4, males: 22.33 ± 3.11, Mann-Whitney U test: $U = 59$, $p = 0.037$); and 27 participants (16 females and 11 males) aged between 18 and 26 years old in the 8 Hz condition (mean ± SD, all participants: 20.93 ± 2.62, females: 20.31 ± 2.47, males: 21.82 ± 2.68, Mann-Whitney U test: $U = 56.5$, $p = 0.121$). They had a visual acuity lower than 0.4 logMAR and a stereoacuity lower than 200 arcsec (Luna & Serrano-Pedraza, 2020).

Regarding Cohort 3, we analyzed the data of 31 participants (22 females and 9 males) with ages ranging from 18 to 33 years old (mean ± SD, all participants: 22.45 ± 4.45 years, females: 22.91 ± 4.99, males: 21.33 ± 2.65, Mann-Whitney U test: $U = 112$, $p = 0.583$). Their average stereoacuity was 26.77 ± 13.45 arcsec and the average visual acuity across both eyes was 0.011 ± 0.09 logMAR (mean ± SD) for a 40 cm distance and –0.06 ± 0.095 logMAR (mean ± SD) for a 300 cm distance (Arranz-Paraíso et al., 2021).

2.2. Equipment

All data presented here were obtained using similar equipment. The experiments were programmed using Matlab (The MathWorks, Natick, MA, USA) with the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) and under the control of a Mac Pro 3.7 GHz Quad Core Intel Xeon E5 (graphics card AMD FirePro D300 2048 MB). The output from the graphics card was controlled by the DataPixx Lite video processor (VPixx Technologies, Inc., Saint-Bruno, Canada), providing 16-bits of gray-scale resolution. The responses to the experiments were recorded using a ResponsePixx Handheld (VPixx Technologies, Inc., Saint-Bruno, Canada). The luminance of the monitors used was gamma-corrected using a Minolta LS-110 photometer (Konica Minolta Optics, Inc., Osaka, Japan). We asked the participants to place their heads on a chin rest (UHCOTech HeadSpot, Houston, TX, USA) to control the distance to the screen and to avoid head movements.

For Cohorts 1 and 2, we used a 17-in. Eizo Scan T565 with a resolution of 800 × 600 pixels (horizontal × vertical), a vertical frame rate of 148 Hz, and a mean luminance of 49.1 cd/m^2. For our main experiment and Cohort 3 we used a 19-in. Eizo Flex Scan T765 with gamma correction applied using the same photometer. This monitor had a resolution of 1024 × 768 pixels, a vertical frame rate of 120 Hz, and a mean luminance of 40.5 cd/m^2 (our main experiment) and 35.4 cd/m^2 (Cohort 3).

2.3. Stimuli

All stimuli were created with Matlab (The MathWorks, Natick, MA, USA). In our main experiment, the stimuli were images of drifting vertical Gabor patches of 750 × 750 pixels, presented inside a square of 25.63 × 25.63 cm on the center of the screen where the remainder of the screen had the mean luminance. Two spatial frequencies were chosen: 0.6 c/deg and 12 c/deg; and two temporal frequencies: 1 Hz and 8 Hz. For the spatial frequency of 0.6 c/deg the screen was placed at 100 cm from the observer (the image subtended an angle of 14.6 × 14.6 deg). For the 12 c/deg spatial frequency, the distance to the screen was 235 cm (the image subtended an angle of 6.24 × 6.24 deg). The stimuli had a nominal diameter of 2 deg (diameter = 2 × σ_{xy}, where σ_{xy} is the standard deviation of the spatial Gaussian window).

For Cohorts 1, 2, and 3, the stimuli consisted of 512 × 512 pixels images that were presented in a squared area of 19.5 × 19.5 cm centered on the screen. The remainder of the screen had the mean luminance. For Cohort 1, the screen was placed 100 cm from the observers, subtending 20.1 × 20.1 degrees of visual angle. The stimuli were vertical Gabor patches of spatial frequency 1 c/deg and a drifting speed of 2 deg/s. Two nominal diameters were used, 0.7 and 6 deg (diameter = 2 × σ_{xy}, where σ_{xy} is the standard deviation of the spatial Gaussian window). The contrast (92%) was temporally modulated with a temporal Gaussian function, where the standard deviation σ_t was controlled by a Bayesian adaptive staircase (see *Procedure* section). We defined the duration of the stimuli as twice the temporal standard deviation (2 × σ_t).

For Cohort 2, the stimuli were vertical Gabor patches. The nominal diameter of the stimuli was 4 deg (2 × σ_{xy}) and the rest of the screen was set to the mean luminance. We used eight spatial frequencies for the Gabor patches: 0.25, 0.5, 0.75, 1, 1.5, 2, 3, and 6 c/deg. This Cohort was divided into two subcohorts: Cohort 2a, where the Michelson contrasts for the drifting Gabor patches were 10% and 90% and the speed of the Gabor patches was 2 deg/s; and Cohort 2b, where the Michelson contrast was always 90% and we used two drifting temporal frequencies for the Gabor patches: 2 and 8 Hz. For both subcohorts, the contrast was temporally modulated by a Gaussian function, whose standard deviation was controlled by a Bayesian adaptive staircase in order to estimate duration thresholds (defined as twice the temporal standard deviation 2 × σ_t; see *Procedure* section).

For Cohort 3, the stimuli were vertical gratings windowed by a spatial 2D Butterworth function of order 10. The spatial frequency was 1 c/deg and the drifting speed was 2 deg/s. Two diameters of the Butterworth window were selected, 1 and 7 deg. Stimuli were viewed binocularly or monocularly (with an orthoptic patch occluding the eye). Michelson contrast was 85% and was temporally modulated with a Gaussian function with the standard deviation controlled by a Bayesian adaptive staircase. As in Cohort 1 and 2, we measured duration thresholds, where the duration of the stimulus was 2 × σ_t.

Examples of the stimuli used in these experiments can be seen in the original papers (Arranz-Paraíso et al., 2021; Arranz-Paraíso & Serrano-Pedraza, 2018; Luna & Serrano-Pedraza, 2020).

2.4. Procedure

The experiments took place in a dark room, with the participants seating beside a table and their heads placed on a chin rest. In each trial, stimuli drifted leftwards or rightwards randomly. The participants' task was always to indicate the direction of motion (i.e. motion direction discrimination task). Two types of thresholds were measured: contrast thresholds and duration thresholds.

In our main experiment we measured contrast thresholds, that is, the minimum amount of contrast needed to correctly discriminate the direction of motion 82% of the times. Each trial began with the presentation of a black cross in the center of the screen for 500 ms, whose contrast was temporally modulated by a Gaussian function with a standard deviation of 80 ms. After the cross had disappeared, the stimulus was presented during an interval of 1000 ms and its contrast was modulated by a Gaussian temporal function with a standard deviation of 250 ms. Once the stimulus presentation was over, the participant indicated if it had drifted leftwards or rightwards. The next trial did not start until a response was given and no feedback on the correctness of the response was provided. The contrast of the stimulus on each trial was controlled by a Bayesian adaptive staircase (Treutwein, 1995) with the following characteristics: (a) The prior probability-density function describing the distribution of the thresholds was uniform (Emerson, 1986; Pentland, 1980; Serrano-Pedraza et al., 2020) between −4 and 3.8 (log10(contrast)) in steps of 0.001 log units. The initial contrast was −0.1 log10(contrast) (Michelson contrast = 0.794). (b) We used the logistic function as the model function (see a complete description of this function and the Bayesian procedure used here in the Supplementary Information of Serrano-Pedraza et al. (2020) appendices A1 and A2). (c) In each trial, the value of the stimulus contrast was taken from the mean of the posterior probability distribution (in log-decimal values) (King-Smith et al., 1994). (d) The staircase terminated after a fixed number of trials (Anderson, 2003). (e) The estimation of the final contrast threshold was done by taking the mean of the final probability-density function (in log-decimal values) (f) Its spread value was 0.8 (with a 0.01 value for the delta parameter, a lapse rate of 0.01, and a 0.5 guess rate). For each condition we measured three thresholds and each threshold required 30 trials. The final contrast threshold for each condition and participant was the average of those three thresholds.

For Cohorts 1, 2, and 3, we measured duration thresholds, defined as the minimum presentation time of a drifting stimulus needed to correctly discriminate its direction of motion 82% of the times. In these experiments, a black fixation cross was presented in each trial on the center of the screen for 500 ms. Its contrast was temporally modulated with a Gaussian function with a standard deviation of 80 ms. After the presentation of the fixation cross, the stimulus was presented during an interval of 1000 ms. Once the stimulus presentation was over, the participant indicated if it had drifted leftwards or rightwards. To obtain the duration thresholds, we also used a Bayesian staircase procedure with characteristics similar as before: (a) The prior probability-density function describing the distribution of the thresholds was uniform between 0.1 and 3.9 (log10(ms)) in steps of 0.001 log units. The starting duration was 200 ms (σ_t = 100 ms). (b) The logistic function was used as the model function. (c) The spread value was 1 (with a 0.01 value for the delta parameter, a lapse rate of 0.01, and a 0.5 guess rate). (d) A Gaussian temporal window determined the temporal duration of the stimuli. It was used to control the contrast as a function of time. (e) In each trial, the mean of the posterior probability distribution (in log-decimal values) was taken to obtain the value of the temporal standard deviation (σ_t). (f) The staircase terminated after 40 trials. (g) The final estimation of the duration threshold (2 × σ_{t0}) was done by taking the mean of the final probability-density function. For Cohort 2, two duration thresholds per condition were obtained and averaged to compute the definitive duration threshold value. For Cohorts 1 and 3, three duration thresholds per condition were obtained and averaged to compute the final duration threshold value.

2.5. Statistical analyses

In the main experiment, to compare contrast thresholds between females and males, we fitted a linear mixed-effects model (Matlab's function "fitlme") using restricted maximum likelihood and assuming the formula: log10(contrast threshold) ∼ Sex × Spatial frequency × Temporal frequency + (1 | ID). Thus, the dependent variable in this model was the contrast threshold (given in logarithmic units). Sex, Spatial frequency, and Temporal frequency were fixed effects, and the participant's ID number was a random effect.

In all cohorts, we compared logarithmic durations thresholds (log10 (ms)) between females and males by fitting a linear mixed-effects model in the same way as we did in the main experiment. In Cohort 1, we used the formula: log10(duration threshold) ~ Sex × Stimulus size + (1 | ID), where the fixed effects were Sex (male/female) and Size (small/large). In Cohort 2a, we used the formula: log10(duration threshold) ~ Sex × Contrast + (1 | ID), where the fixed effects were Sex (male/female) and Contrast (10% / 90%), while in Cohort 2b, we used log10(duration threshold) ~ Sex × Temporal frequency + (1 | ID), where the fixed effects were Sex (male/female) and Temporal frequency (2 Hz / 8 Hz). Lastly, in Cohort 3, we used the formula: log10(duration threshold) ~ Sex × Stimulus size + (1 | ID), where the fixed effects were Sex (male/female) and Size (small/large), separately for each viewing condition (monocular and binocular).

In order to assess the overall effect of the variables in each study, we performed an ANOVA marginal test using Satterthwaite degrees of freedom for the *F*-test (Type III test of fixed effects) over each of the linear models fitted to the data.

We also calculated the effect size comparing contrast thresholds, and duration thresholds between males and females using the standardized mean difference Cohen's *d*. It is defined as:

$$d = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}},$$

where $\overline{X}_1$ and $\overline{X}_2$ are sample means of each group, S_1 and S_2 are their standard deviations, and n_1 and n_2 are the sample sizes. We will interpret the magnitude of *d* following Cohen's conventions (Cohen, 1988). That is, small (0.2), medium (0.5), and large (0.8).

Additionally, in Cohorts 1 and 3, we compared the means of the Motion Suppression Index (MSI) between males and females. The MSI is defined as the logarithmic ratio between the duration thresholds for the large stimulus and the duration thresholds for the small stimulus ($\text{MSI} = \log_{10}(\text{duration threshold}_{\text{Large}}) - \log_{10}(\text{duration threshold}_{\text{Small}})$). We also tested the normality (Shapiro-Wilk test) and the homogeneity of variance (F test) of the data. When normality and homogeneity of variances were met, we used a Student's *t*-test (if homogeneity of variances was not met, we used the Smith-Welch-Satterthwaite test or unequal variance *t*-test). However, when the data did not meet these assumptions, we took the non-parametric alternative of the Mann-Whitney *U* test. Parametric and non-parametric tests led to the same conclusions; thus, we will only show the parametric analyses.

3. Results

3.1. Main experiment. Effect of spatial and temporal frequency on contrast sensitivity

The results from our main experiment can be seen in Fig. 1. Both groups show lower contrast thresholds for 0.6 c/deg than for 12 c/deg, both for 1 Hz and 8 Hz (see Fig. 1A). In general, contrast thresholds for 8 Hz are lower than for 1 Hz, replicating previous results (Kulikowski & Tolhurst, 1973) (see their Figs. 7 and 8, flicker detection data). For the 0.6 c/deg stimuli, and for both temporal frequencies, contrast thresholds are lower for females. At 12 c/deg and 1 Hz, the contrast thresholds of both groups are almost the same; and for 8 Hz, males have lower contrast thresholds. These differences between the contrast thresholds of males and females are not statistically significant (see the analysis below). Fig. 1A shows similar distributions of contrast thresholds for males and females. Fig. 1B shows the contrast sensitivity (i.e., reciprocal of contrast thresholds in Michelson contrast units) of males and females

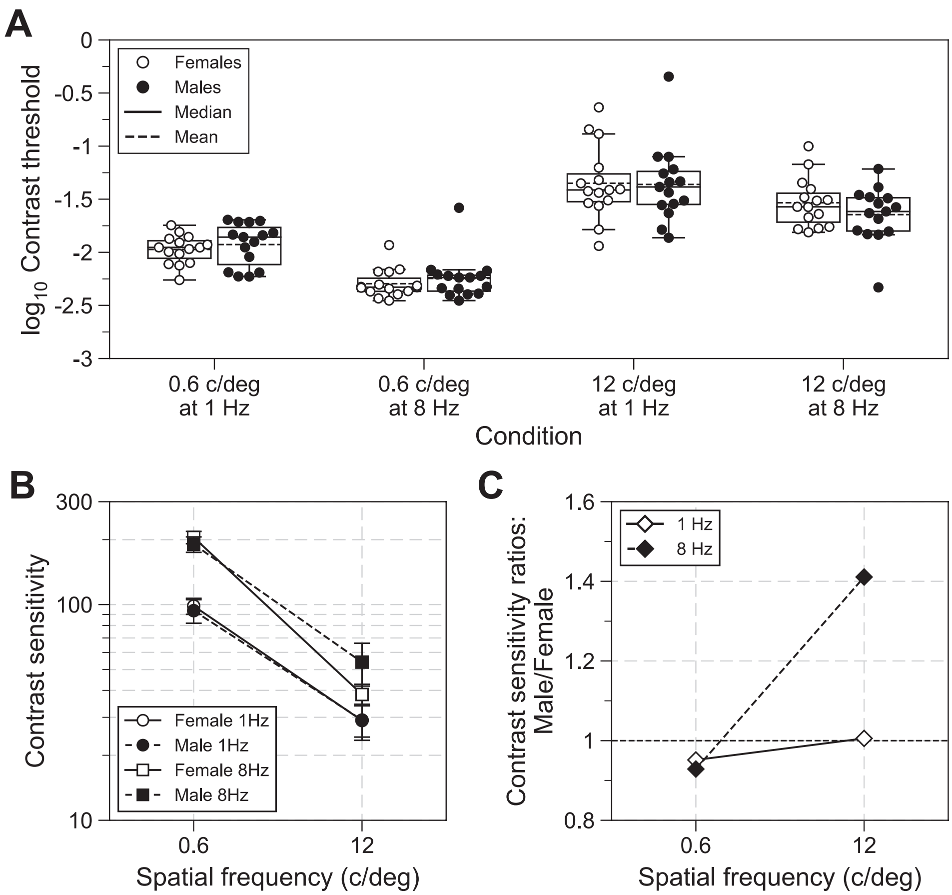

Fig. 1. Results from the main experiment comparing females and males. (A) Beeswarm plots and box plots representing the contrast thresholds (in log-decimal units) for each experimental condition (*N* = 30, 15 females and 15 males). The dashed line inside the boxplot shows the mean and the black line shows the median. The bottom line of the box shows the 25th percentile (Q1) and the upper line the 75th percentile (Q3). The upper whisker corresponds to the largest value that is less than or equal to Q3 + 1.5 × IQR and the lower whisker corresponds to the lowest value that is greater than or equal to Q1 − 1.5 × IQR, where IQR = Q3 − Q1 is the interquartile range. (B) Contrast sensitivity (i.e., reciprocal of contrast thresholds in Michelson contrast units; mean ± SEM) as a function of spatial frequency. (C) Ratios between male and female contrast sensitivity as a function of spatial frequency. Note: Ratios lower than 1 mean higher contrast sensitivity for females and ratios higher than 1 mean higher contrast sensitivity for males.

as a function of spatial frequency for two temporal frequencies. Fig. 1C shows the contrast sensitivity ratios, computed as the ratio between the means of the contrast sensitivity shown in Fig. 1B. The points that stand on the dashed line indicate equal contrast sensitivity (Sensitivity Ratio = 1), the ones below the line indicate higher contrast sensitivity for females and the ones above the line indicate higher contrast sensitivity for males. This plot reflects that for conditions 0.6 c/deg at 1 and 8 Hz, the ratios are 0.95 and 0.93 respectively, and for conditions 12 c/deg at 1 and 8 Hz, the ratios are 1.01 and 1.41 respectively. Only two conditions agree with previous results (Abramov et al., 2012a) (see their Fig. 2). Namely, females for 0.6 c/deg and 1 Hz and males for 12 c/deg and 8 Hz. The other two conditions (0.6 c/deg at 8 Hz and 12 c/deg at 1 Hz) show contrast sensitivity ratios opposite to the study of Abramov et al. (2012a).

The fitted linear mixed-effects model has eight fixed-effects coefficients. The statistically significant ones are the intercept (β_0 = -1.791, 95% CI = [-1.853, −1.73]), the effect of spatial frequency (β_2 = -0.319, 95% CI = [-0.357, −0.282]) and the effect of temporal frequency (β_3 = 0.139, 95% CI = [0.101, 0.176]). The model also has 30 random-effects coefficients or random intercepts (SD = 0.133; 95% CI = [0.087, 0.205]) and random residuals (SD = 0.208, 95% CI = [0.179, 0.242]). The deviance for the model fit is 31.374. The ANOVA marginal test shows a significant main effect of spatial frequency ($F_{1,\ 84}$ = 283.17, p = 1.227 × 10^{-28}) and temporal frequency ($F_{1,\ 84}$ = 53.472, p = 1.415 × 10^{-10}), while sex does not reach statistical significance ($F_{1,\ 28}$ = 0.012, p = 0.916). The interactions between these three variables are not statistically significant either.

The effect size (Cohen's *d*) is small for all conditions, with most of them having values around 0.3, except for 12 c/deg at 8 Hz, which has the strongest effect (see Table 1).

3.2. Results from Cohort 1. Effect of stimulus size on motion discrimination

Fig. 2 shows the results from Cohort 1. Duration thresholds for males and females increase with increasing size, in agreement with previous results (Tadin et al., 2003). Comparing males (black dots) and females (white dots), the latter have higher duration thresholds for both stimulus sizes.

We fitted a linear mixed-effects model with four fixed-effects coefficients. Only the intercept (β_0 = 1.748, 95% CI = [1.714, 1.783]) and the effect of the stimulus size (β_2 = -0.184, 95% CI = [-0.207, –0.162]) are statistically significant. The model also has 47 random-effects coefficients (SD = 0.087; 95% CI = [0.059, 0.128]) and random residuals (SD = 0.104; 95% CI = [0.085, 0.128]). The deviance for the fitted model is −94.598. The ANOVA marginal test shows a significant main effect of stimulus size ($F_{1,\ 45}$ = 264.89, p = 1.781 × 10^{-20}), but sex does not reach statistical significance ($F_{1,\ 45}$ = 2.173, p = 0.147). The interaction between these two variables is not statistically significant either ($F_{1,\ 45}$ = 0.001, p = 0.978). We also computed the Motion Suppression Index (MSI) for males and females. The MSI is not significantly different between males and females (t_{45} = 0.028, p = 0.978).

The effect size (Cohen's *d*) shows a medium effect for the small stimulus size and a small effect for the large stimulus size (see Table 2). The effect size for the MSI is almost zero.

Table 1
Main experiment results. Averaged contrast thresholds in log-decimal values and effect size for each condition.

Condition	Females	Males	Cohen's *d*
	Mean ± SD	Mean ± SD	
0.6 c/deg – 1 Hz	−1.97 ± 0.13	−1.93 ± 0.2	−0.27
0.6 c/deg – 8 Hz	−2.3 ± 0.13	−2.25 ± 0.21	−0.28
12 c/deg – 1 Hz	−1.35 ± 0.35	−1.36 ± 0.36	0.03
12 c/deg – 8 Hz	−1.53 ± 0.23	−1.65 ± 0.26	0.45

N_{Females} = 15, N_{Males} = 15.

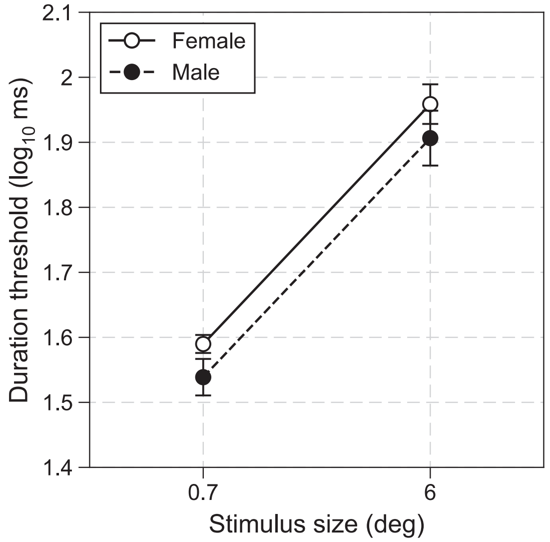

Fig. 2. Results from Cohort 1comparing females and males. The panel shows duration thresholds (in log-decimal units, mean ± SEM) as a function of stimulus size (N = 47, 31 females and 16 males). The spatial frequency of the Gabor patch is 1 c/deg, the contrast 92%, and the speed 2 deg/s.

Table 2
Cohort 1 results. Averaged duration thresholds in $\log_{10}$(ms) and effect size for each condition.

Condition	Females	Males	Cohen's *d*
	Mean ± SD	Mean ± SD	
Small	1.59 ± 0.08	1.54 ± 0.11	0.57
Large	1.96 ± 0.17	1.91 ± 0.17	0.31
MSI	0.37 ± 0.16	0.37 ± 0.12	0.01

N_{Females} = 31, N_{Males} = 16.

3.3. Results from Cohort 2a. Effect of contrast and spatial frequency on motion discrimination

Results from Cohort 2a are plotted in Fig. 3. Fig. 3A and 3B show results for 10% and 90% contrasts, respectively. The results show that duration thresholds for motion discrimination decrease with increasing spatial frequency, indicating a better performance for higher spatial frequencies. In general, females (white dots) show worse performance than males (black dots). For the 10% contrast stimuli, females have higher duration thresholds for all spatial frequencies, with the largest differences occurring in the range of 1 to 6 c/deg (see Fig. 3A). For 90% contrast, females show again higher duration thresholds for most of the spatial frequencies presented (see Fig. 3B). In this case, the largest difference between both groups is between spatial frequencies 0.25 and 1 c/deg. As can be seen in Fig. 3A and 3B, despite the small difference in duration thresholds, the shape of the data is similar for males and females.

Fitting a linear mixed-effects model, it can be seen that, for the low contrast (10%), the model has sixteen fixed-effects coefficients, including the intercept, sex, the effect of spatial frequency, and the interactions between sex and spatial frequency. The main effect of sex and its interaction with spatial frequency is not statistically significant. The model has 26 random-effects coefficients or random intercepts (SD = 0.12; 95% CI = [0.087, 0.165]) and random residuals (SD = 0.119; 95% CI = [0.107, 0.132]). The deviance of the model fit is −160.2. The ANOVA marginal test shows a significant main effect of spatial frequency ($F_{7,\ 168}$ = 6.886, p = 3.347 × 10^{-7}), with sex not reaching

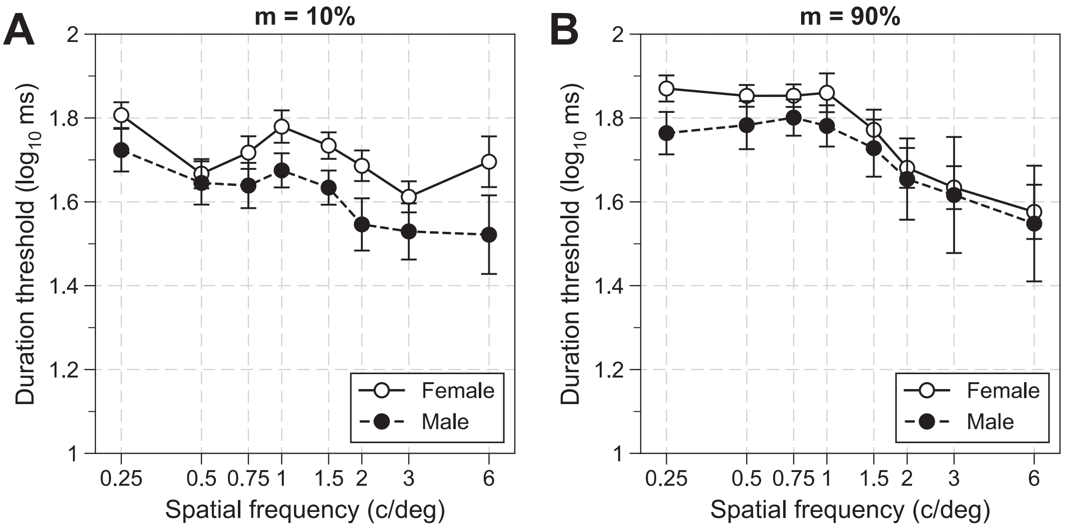

Fig. 3. Results from Cohort 2a comparing females and males. (A) Duration thresholds (in logarithmic units, mean ± SEM) as a function of spatial frequency (c/deg) for 10% contrast stimuli and a drifting speed of 2 deg/s. (B) 90% contrast and a drifting speed of 2 deg/s. In both panels, $N = 26$ participants (17 females and 9 males).

statistical significance ($F_{1,\ 24} = 3.514$, $p = 0.073$), and the interaction between sex and spatial frequency not being statistically significant ($F_{7,\ 168} = 0.841$, $p = 0.555$).

For high contrast (90%), the fitted linear mixed effect model has sixteen fixed-effects coefficients; the same ones it had for low contrast (10%). Once again, the main effect of sex and its interaction with spatial frequency does not reach statistical significance. The model has 26 random-effects coefficients or random intercepts (SD = 0.17; 95% CI = [0.126, 0.231]) and random residuals (SD = 0.13; 95% CI = [0.117, 0.145]). The deviance for the model fit is −113.91. The ANOVA marginal test shows a significant main effect of spatial frequency ($F_{7,\ 168} = 15.093$, $p = 3.161 \times 10^{-15}$), but sex does not reach statistical significance ($F_{1,\ 24} = 0.531$, $p = 0.473$), and the interaction between these two variables is not statistically significant either ($F_{7,\ 168} = 0.321$, $p = 0.944$).

For 10% contrast (see Table 3), the stronger effect sizes are located at high spatial frequencies (1–6 c/deg). On the other hand, for 90% contrast (see Table 4), the stronger effect sizes are found at low spatial frequencies (0.25–1 c/deg).

3.4. Results from Cohort 2b. Effect of spatial and temporal frequency on motion discrimination

Fig. 4 shows the results for Cohort 2b comparing females and males. Fig. 4A shows duration thresholds (in logarithmic units) as a function of spatial frequency for females and males for a temporal frequency of 2 Hz and 90% contrast. In general, performance decreases (i.e. duration thresholds increase) with increasing spatial frequency up to 1 c/deg and then remains constant for higher spatial frequencies. Fig. 4B shows the results for 8 Hz and 90% contrast. Duration thresholds show the same tendency: they increase up to 1 c/deg and then they remain constant for higher spatial frequencies. Duration thresholds for 8 Hz are in general lower than for 2 Hz. For both temporal frequencies and for spatial frequencies lower than 6 c/deg, duration thresholds are slightly higher for females than for males. However, the shape of the data is similar for males and females.

Table 3
Cohort 2a results, contrast 10%. Averaged duration thresholds in $\log_{10}$(ms) and effect size for each condition.

Spatial frequency (c/deg)	Females	Males	Cohen's *d*
	Mean ± SD	**Mean ± SD**	
0.25	1.81 ± 0.13	1.72 ± 0.15	0.61
0.5	1.67 ± 0.14	1.65 ± 0.16	0.15
0.75	1.72 ± 0.16	1.64 ± 0.16	0.49
1	1.78 ± 0.16	1.68 ± 0.12	0.71
1.5	1.73 ± 0.13	1.63 ± 0.12	0.78
2	1.69 ± 0.15	1.55 ± 0.19	0.86
3	1.61 ± 0.15	1.53 ± 0.2	0.49
6	1.7 ± 0.25	1.52 ± 0.28	0.67

$N_{\text{Females}} = 17$, $N_{\text{Males}} = 9$.

Table 4
Cohort 2a results, contrast 90%. Averaged duration thresholds in $\log_{10}$(ms) and effect size for each condition.

Spatial frequency (c/deg)	Females	Males	Cohen's *d*
	Mean ± SD	**Mean ± SD**	
0.25	1.87 ± 0.13	1.76 ± 0.15	0.78
0.5	1.85 ± 0.11	1.78 ± 0.17	0.53
0.75	1.85 ± 0.11	1.8 ± 0.13	0.44
1	1.86 ± 0.19	1.78 ± 0.15	0.45
1.5	1.77 ± 0.2	1.72 ± 0.2	0.22
2	1.68 ± 0.2	1.65 ± 0.29	0.12
3	1.63 ± 0.21	1.62 ± 0.42	0.06
6	1.58 ± 0.27	1.55 ± 0.41	0.09

$N_{\text{Females}} = 17$, $N_{\text{Males}} = 9$.

Fitting a linear mixed-effects model, it can be seen that, for the lower temporal frequency (2 Hz), the model includes sixteen fixed-effects coefficients, consisting of the intercept, sex, the effect of the spatial frequency, and the interactions between sex and spatial frequency. Neither the main effect of sex, nor its interaction with spatial frequency are statistically significant. The model also includes 30 random-effects coefficients or random intercepts (SD = 0.128; 95% CI = [0.096, 0.17]) and random residuals (SD = 0.111; 95% CI = [0.101, 0.123]). The deviance for the model fit is −217.87. The ANOVA marginal test shows a significant main effect of spatial frequency ($F_{7,\ 196} = 51.644$, $p = 3.311 \times 10^{-41}$), sex does not reach statistical significance ($F_{1,\ 28} = 1.474$, $p = 0.235$), and the interaction between these two variables is not statistically significant either ($F_{7,\ 196} = 0.558$, $p = 0.789$).

For the higher temporal frequency (8 Hz), the linear mixed effect model includes sixteen fixed-effects coefficients, the same ones it had for the lower temporal frequency of 2 Hz. The main effect of sex and its interaction with spatial frequency does not reach statistical significance. The model has 27 random-effects coefficients or random intercepts (SD

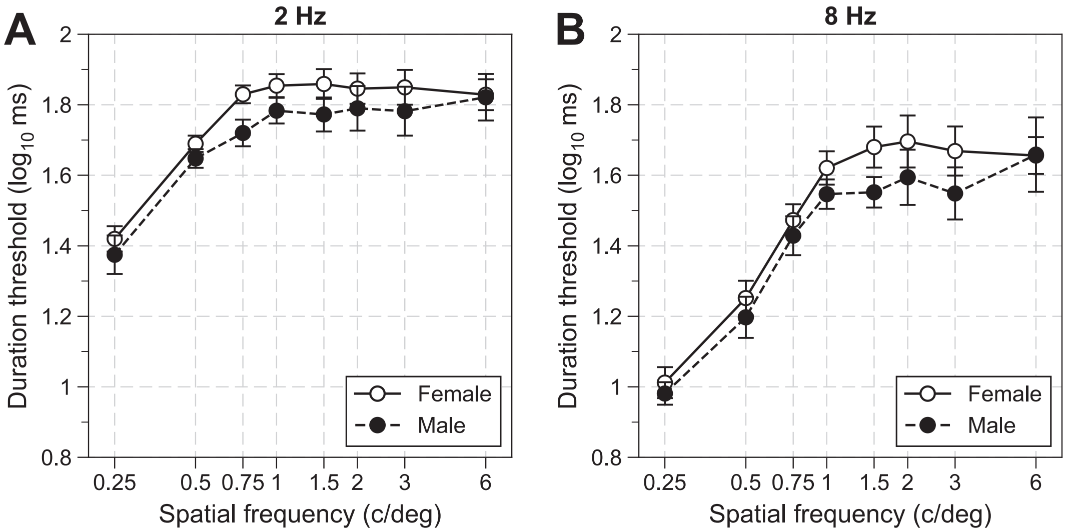

Fig. 4. Results from Cohort 2b comparing females and males. (A) Duration thresholds (in logarithmic units, mean ± SEM) as a function of spatial frequency (c/deg) for a temporal frequency of 2 Hz and 90% contrast ($N = 30$, 18 females and 12 males). (B) Temporal frequency of 8 Hz and 90% contrast ($N = 27$, 16 females and 11 males).

= 0.162; 95% CI = [0.119, 0.221]) and random residuals (SD = 0.149; 95% CI = [0.134, 0.166]). The deviance of the model fit is −73.76. The ANOVA marginal test shows a significant main effect of spatial frequency ($F_{7,\ 175} = 67.499$, $p = 1.995 \times 10^{-46}$), but sex is not statistically significant ($F_{1,\ 25} = 1.064$, $p = 0.312$), and the interaction between these two variables is not statistically significant either ($F_{7,\ 175} = 0.612$, $p = 0.746$).

For 2 Hz (see Table 5), the effect sizes (Cohen's *d*) show a band-pass shape as a function of spatial frequency (i.e. higher Cohen's *d* for medium spatial frequencies and lower Cohen's *d* for low and high spatial frequencies) with a medium effect between 0.5 and 1.5 c/deg and where the strongest effect is obtained for 0.75 c/deg. For 8 Hz (see Table 6), we also have this band-pass shape, but it is displaced to higher spatial frequencies (medium effect between 1 and 3 c/deg). In this case, the strongest effect size is obtained for 1.5 c/deg.

3.5. Results from Cohort 3. Effect of size and viewing condition (monocular vs binocular) in motion discrimination

This Cohort presents data for two viewing conditions (monocular and binocular). Results for the monocular viewing condition can be seen in Fig. 5A. Results for the binocular condition are presented in Fig. 5B. As in Cohort 1 and previous research, duration thresholds for males and females are lower for the smallest stimulus size. In both viewing conditions and for both sizes, duration thresholds for males are lower than for females. However, the Motion Surround Suppression index is similar for females and males (see Fig. 5C).

Table 5
Cohort 2b, temporal frequency of 2 Hz. Averaged duration thresholds in $\log_{10}$(ms) and effect size for each condition.

Spatial frequency (c/deg)	Females	Males	Cohen's *d*
	Mean ± SD	**Mean ± SD**	
0.25	1.42 ± 0.15	1.38 ± 0.19	0.27
0.5	1.69 ± 0.1	1.65 ± 0.09	0.44
0.75	1.83 ± 0.11	1.72 ± 0.13	0.94
1	1.85 ± 0.14	1.78 ± 0.13	0.53
1.5	1.86 ± 0.18	1.77 ± 017	0.5
2	1.85 ± 0.18	1.79 ± 0.22	0.28
3	1.85 ± 0.21	1.78 ± 0.24	0.3
6	1.83 ± 0.19	1.82 ± 0.23	0.04

$N_{\text{Females}} = 18$, $N_{\text{Males}} = 12$.

Table 6
Cohort 2 results, temporal frequency of 8 Hz. Averaged duration thresholds in $\log_{10}$(ms) and effect size for each condition.

Spatial frequency (c/deg)	Females	Males	Cohen's *d*
	Mean ± SD	**Mean ± SD**	
0.25	1.01 ± 0.18	0.98 ± 0.11	0.2
0.5	1.25 ± 0.2	1.2 ± 0.19	0.28
0.75	1.47 ± 0.18	1.43 ± 0.18	0.24
1	1.62 ± 0.19	1.55 ± 0.14	0.43
1.5	1.68 ± 0.23	1.55 ± 0.14	0.64
2	1.7 ± 0.29	1.59 ± 0.26	0.36
3	1.67 ± 0.28	1.55 ± 0.25	0.45
6	1.66 ± 0.21	1.66 ± 0.35	−0.01

$N_{\text{Females}} = 16$, $N_{\text{Males}} = 11$.

We reanalyzed the data for each viewing condition (monocular/binocular) independently. For monocular viewing, the fitted linear mixed-effects model has four fixed-effects coefficients, including the intercept, sex, stimulus size, and the interaction between sex and stimulus size. The intercept ($\beta_0 = 1.519$, 95% CI = [1.477, 1.561]), the effect of sex ($\beta_1 = 0.059$, 95% CI = [0.017, 0.102]) and the effect of the stimulus size ($\beta_2 = -0.07$, 95% CI = [-0.098, −0.041]) are statistically significant. The model has 31 random-effects coefficients or random intercepts (SD = 0.078; 95% CI = [0.047, 0.132]) and random residuals (SD = 0.102; 95% CI = [0.078, 0.131]). The deviance for the model fit is −61.933. The ANOVA marginal test shows a significant main effect of sex ($F_{1,\ 29} = 7.994$, $p = 0.008$) and stimulus size ($F_{1,\ 29} = 24.152$, $p = 3.218 \times 10^{-5}$). Student's *t*-test comparing duration thresholds between males and females shows significant differences for small size ($t_{29} = 2.775$, $p = 0.01$) but not for large size ($t_{29} = 2.167$, $p = 0.039$) (Bonferroni correction, $\alpha = 0.05/2 = 0.025$). The interaction between sex and size is not statistically significant ($F_{1,\ 29} = 0.197$, $p = 0.66$). Despite the significant effect of sex on duration thresholds, no significant differences are appreciated when comparing the Motion Suppression index (MSI) between both groups ($t_{27.7} = 0.646$, $p = 0.524$, Smith-Welch-Satterthwaite test or unequal variance *t*-test).

For binocular viewing, the fitted linear mixed-effects model has the same four fixed-effects coefficients as in the monocular viewing condition. Again, the intercept ($\beta_0 = 1.545$, 95% CI = [1.501, 1.59]), the effect of sex ($\beta_1 = 0.058$, 95% CI = [0.013, 0.103]) and the effect of the stimulus size ($\beta_2 = -0.121$, 95 % CI = [-0.148, −0.094]) are statistically significant. The model has 31 random-effects coefficients or random

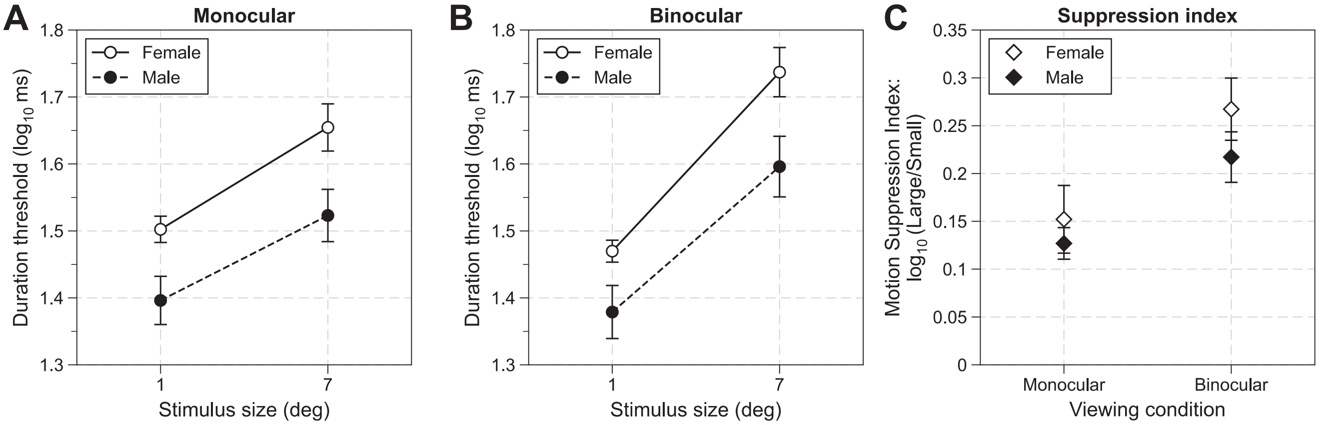

Fig. 5. Results from Cohort 3 comparing females and males. The spatial frequency of the grating is 1 c/deg, contrast is 85%, and speed is 2 deg/s. Panels A and B show duration thresholds (in log-decimal units, mean ± SEM) as a function of stimulus size. (A) Monocular viewing. (B) Binocular viewing. (C) Motion suppression index (mean ± SEM) as a function of viewing condition. For both viewing conditions, $N = 31$ participants (22 females and 9 males).

intercepts (SD = 0.09; 95% CI = [0.059, 0.139]) and random residuals (SD = 0.096, 95% CI = [0.075, 0.125]). The deviance for the model fit is −61.316. The ANOVA marginal test shows a significant main effect of sex ($F_{1,\ 29} = 6.721$, $p = 0.015$) and stimulus size ($F_{1,\ 29} = 80.738$, $p = 7.052 \times 10^{-10}$). Student's *t*-test comparing duration thresholds between males and females shows significant differences for small size ($t_{29} = 2.541$, $p = 0.017$) but not for large size ($t_{29} = 2.182$, $p = 0.037$) (Bonferroni correction, $\alpha = 0.05/2 = 0.025$). The interaction between sex and size is not statistically significant ($F_{1,\ 29} = 0.864$, $p = 0.36$). We also do not find significant differences in the MSI ($t_{29} = 0.93$, $p = 0.36$), thus, females and males present similar motion surround suppression despite the differences in duration thresholds.

The effect size (Cohen's *d*) in the monocular condition is large for both stimulus sizes ($d > 0.8$) and small for the MSI (see Table 7). In the binocular condition, the effect size is large for both stimulus sizes ($d > 0.8$) and small for the MSI (see Table 8).

3.6. Combining results from Cohorts

One way to increase the statistical power of our analysis is by increasing the sample size. There are differences between the experimental parameters of the three Cohorts that make it inappropriate to combine the data, even if we choose the spatial frequency of 1 c/deg and the speed 2 deg/s, which is common in the three Cohorts. For example, one significant difference is the spatial window used in the experiments. We used a Butterworth window in Cohort 3 (C3) and a Gaussian window in Cohort 1(C1) and 2 (C2). The contrast is also lower in C3 than in C1 and C2. These experimental differences are evident when we compare the duration thresholds of the three cohorts for large sizes, 1 c/deg, and speed of 2 deg/s (check Figs. 2, 3, 4, and 5). On the other hand, we cannot combine the results of C2a and C2b because there are repeated participants. However, although the stimulus size is different and duration thresholds depend on the size of the stimulus, here we combine the data for the largest size of C1 (1 c/deg, 2 deg/s, 6 deg, 92% contrast) and the data of C2a (1 c/deg, 2 deg/s, 4 deg, 90% contrast) to achieve a larger statistical power. We believe this analysis will provide us with useful information despite the small differences in stimulus size between cohorts. As well, we combine C1 with C2b (1 c/deg, 2 Hz, 4 deg, 90% contrast).

After combining duration thresholds from C1 and C2a ($N = 73$; 25 males and 48 females), we performed a two-sample *t*-test comparing males (Mean = 1.86 log10(ms), SD = 0.17) and females (Mean = 1.92 log10(ms), SD = 0.18) and the test showed no significant differences $t_{71} = 1.428$, $p = 0.158$. Similarly, combining duration thresholds from C1 and C2b ($N = 77$; 28 males and 49 females) did not show significant differences between males (Mean = 1.85 log10(ms), SD = 0.16) and females (Mean = 1.92 log10(ms), SD = 0.165), $t_{75} = 1.717$, $p = 0.089$. Thus, adding more statistical power to our analysis did not show significant differences between the duration thresholds of males and females.

Table 7
Cohort 3 results, monocular viewing. Averaged duration thresholds in $\log_{10}$(ms) and effect size for each condition.

Condition	Females	Males	Cohen's *d*
	Mean ± SD	Mean ± SD	
Small	1.5 ± 0.09	1.4 ± 0.11	1.1
Large	1.66 ± 0.17	1.52 ± 0.12	0.86
MSI	0.15 ± 0.17	0.13 ± 0.05	0.18

$N_{\text{Females}} = 22$, $N_{\text{Males}} = 9$.

Table 8
Cohort 3 results, binocular viewing. Averaged duration thresholds in $\log_{10}$(ms) and effect size for each condition.

Condition	Females	Males	Cohen's *d*
	Mean ± SD	Mean ± SD	
Small	1.47 ± 0.08	1.38 ± 0.12	1.01
Large	1.74 ± 0.17	1.6 ± 0.14	0.86
MSI	0.27 ± 0.15	0.22 ± 0.08	0.37

$N_{\text{Females}} = 22$, $N_{\text{Males}} = 9$.

4. Discussion

In this work, we have presented the results of one motion discrimination experiment measuring contrast thresholds and a reanalysis of four motion direction discrimination experiments performed in our laboratory, comparing duration thresholds for males and females for a broad range of spatial frequencies, contrasts, and temporal frequencies. Although the results from our main experiment did not show significant differences for any condition, for high spatial and temporal frequencies (e.g. 12 c/deg and 8 Hz), and for low spatial and temporal frequencies (e.g. 0.6 c/deg and 1 Hz), the ratios between males and females contrast sensitivity (see our Fig. 1C) were similar to those reported by Abramov et al., (2012a) (see their Fig. 2). However, for the other two conditions tested (0.6 c/deg at 8 Hz and 12 c/deg at 1 Hz), the contrast sensitivity ratios have shown opposite results to Abramov et al.'s (2012a) study. A few methodological differences with respect to Abramov et al.'s (2012a) study exist that could explain these results. For example, in Abramov et al.'s study they measured contrast thresholds using flickering gratings in an orientation detection task (e.g. vertical vs horizontal). In our case, we measured contrast thresholds using vertical drifting Gabor patches in a motion direction discrimination task (e.g. left vs right). It is well known that contrast thresholds for temporally modulated gratings depend on the detection task. For instance, contrast thresholds for flicker recognition and pattern recognition are different when represented as a function of temporal frequency and depend on spatial frequency (Kulikowski & Tolhurst, 1973) (see their Figs. 6 and 7). Given that our results have shown that contrast sensitivity was lower for low (1 Hz) than high (8 Hz) temporal frequencies for both spatial frequencies tested (0.6 and 12 c/deg), it is reasonable to assume that our participants performed flicker-recognition in all four conditions. However, the results of Abramov et al. (2012a) (see their Fig. 1) show higher contrast sensitivity for 12 c/deg and 1 Hz than for 8 Hz (this is evident for males; see their Fig. 1b). This may suggest that when performing the orientation-recognition task for high spatial frequencies, their participants employed pattern-recognition strategies to discriminate vertical from horizontal orientations. However, it is not clear why this potential difference only affects two conditions and not the four of them. Future sex differences research should explore specifically whether these types of recognition strategies have an effect on motion discrimination experiments.

In a recent study using suprathreshold stimuli, Murray et al. (2018) found that males had shorter duration thresholds than females. These differences were strong at low contrasts and large stimulus sizes, and even stronger at high contrasts independently of stimulus size. In Murray et al.'s study, they measured duration thresholds using stimuli within a narrow range of spatial frequencies (1–1.2 c/deg) and speeds (4–4.8 deg/s). On the contrary, in our reanalysis of four motion direction discrimination experiments, we have compared males and females across a broader range of spatial frequencies with different stimulus sizes, contrasts, and temporal frequencies. The analyses have shown that only the data from our Cohort 3 could replicate the conclusions from Murray et al.'s study, as there were statistically significant differences between the duration thresholds of males and females. Despite the differences between both groups in duration thresholds, there were no differences in motion surround suppression (measured with the Motion Suppression Index). Although Murray et al. (2018) did not report any analysis about surround suppression, their duration thresholds results agree with our conclusions. Their Fig. 1 (panels B, D, E & F) shows a constant increase in duration thresholds (represented in log scale) for the smallest and the largest size, so probably their data also show no differences in surround suppression between males and females. Another difference with Murray et al.'s study is the sample size of both studies. They analyzed data from three Cohorts where two were larger than ours, their Cohort 2 (Rochester cohort from Melnick et al. (2013), $N = 53$, 28 female and 25 male) and their Cohort 3 (Bern cohort from Troche et al. (2018), $N = 177$, 116 female and 61 male). Interestingly, we got a statistically significant result in our Cohort 3 ($N = 33$), similar to Murray et al.'s Cohort 3, with a sample size that is almost six times smaller. The analyses in our other cohorts (e.g. 1, 2a, & 2b) were not significant, as opposed to the results given in Murray et al., but the sample sizes were not that different between both studies. They found significant results with 33 (their Cohort 1), and 53 participants, while we did not reach statistical significance with 47, 26, and 30 participants (our Cohorts 1, 2a, & 2b). In any case, in order to increase the statistical power, we performed new analyses combining the data from Cohort 1 and 2a, and Cohort 1 and 2b for conditions where experimental parameters were similar. Our analysis showed that the differences between the duration thresholds of males and females are still not significant. Thus, these data suggest that the sample size did not play a critical role in the conclusions of this research.

Comparing the effect size with the study from Murray et al. (2018), where they found effect sizes between 0.56 and 1.1, we have shown that one condition from our Cohort 1 and all conditions from our Cohort 3 present similar Cohen's d values. For example, for our Cohort 1, the magnitude of the effect size of motion duration thresholds between males and females was medium for small stimuli (Cohen's $d = 0.57$), however, for large stimuli the effect size was small (Cohen's $d = 0.31$). For our Cohort 3, the effect size was strong in the monocular viewing condition for both stimulus sizes, that is small (Cohen's $d = 1.1$) and large (Cohen's $d = 0.86$); and was also strong in the binocular condition for both small (Cohen's $d = 1.01$) and large (Cohen's $d = 0.86$) sizes. These effect sizes from Cohort 3 are similar to the ones reported by Murray et al. (2018), where they obtained a strong effect for 95 % contrast (Cohen's $d = 0.76$), and for 98% contrast (Cohen's $d = 1.11$).

Results from our Cohort 2 showed duration thresholds for a broad range of spatial frequencies, temporal frequencies, and contrasts. Results from Cohort 2a showed that for 10% and 90% contrast, females had slightly higher duration thresholds (i.e. lower sensitivity for motion direction discrimination) for most of the spatial frequencies presented. For the lower contrast, both groups differed more at high spatial frequencies (1–6 c/deg), and, for the higher contrast, differences were higher at low spatial frequencies (0.25–1 c/deg). However, the analysis revealed that there were no significant differences between the performance of males and females on any of the conditions tested; and both groups showed a similar data shape.

Finally, results from Cohort 2b showed that duration thresholds were in general slightly higher for females than for males for both temporal frequencies (2 and 8 Hz) and for spatial frequencies lower than 6 c/deg. For the lowest temporal frequency, the strongest sex differences were present at low spatial frequencies (peak around 0.75 c/deg), and for high temporal frequencies, the strongest differences were present at higher spatial frequencies (peak around 1.5 c/deg). Nevertheless, once again, the analysis did not reveal significant differences between the duration thresholds of males and females for any of the spatial and temporal frequencies tested.

In summary, we report evidence that shows that males and females perform very similar in visual motion discrimination tasks. Despite the results obtained by males showing a higher sensitivity for motion direction discrimination in most of the studies, the analysis of the data only revealed a statistically significant difference between males and females in the data from Cohort 3; and in all experiments, data patterns were similar for both, females, and males. In general, our results do not support the claim that sex is a strong factor that should be included in the analyses of visual motion perception (Cahill, 2006; McCarthy et al., 2017; S. O. Murray et al., 2018).

CRediT authorship contribution statement

Omar Bachtoula: Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – review & editing. **Sandra Arranz-Paraíso:** Data curation, Investigation, Validation, Visualization, Writing – review & editing. **Raúl Luna:** Data curation, Investigation, Validation, Visualization, Writing – review & editing. **Ignacio Serrano-Pedraza:** Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All relevant data are within the article and its supporting information file

Acknowledgements

Supported by grant PID2021-122245NB-I00 from Ministerio de Ciencia e Innovación (Spain) to ISP. The funding source was not involved in the preparation of this article. RL is supported by a Juan de la Cierva-Formación fellowship (FJC2020-044084-I) funded by MCIN/AEI /10.13039/501100011033 and by the European Union Next Generation EU/PRTR.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.visres.2023.108222.

References

Abramov, I., Gordon, J., Feldman, O., & Chavarga, A. (2012a). Sex & vision I: Spatio-temporal resolution. *Biology of Sex Differences, 3*(1), 20. https://doi.org/10.1186/2042-6410-3-20

Abramov, I., Gordon, J., Feldman, O., & Chavarga, A. (2012b). Sex and vision II: Color appearance of monochromatic lights. *Biology of Sex Differences, 3*(1), 21. https://doi.org/10.1186/2042-6410-3-21

Anderson, A. J. (2003). Utility of a dynamic termination criterion in the ZEST adaptive threshold method. *Vision Research, 43*(2), 165–170. https://doi.org/10.1016/s0042-6989(02)00396-6

Arranz-Paraíso, S., Read, J. C. A., & Serrano-Pedraza, I. (2021). Reduced surround suppression in monocular motion perception. *Journal of Vision, 21*(1), 10. https://doi.org/10.1167/jov.21.1.10

Arranz-Paraíso, S., & Serrano-Pedraza, I. (2018). Testing the link between visual suppression and intelligence. *PloS One, 13*(7), e0200151.

Brabyn, L. B., & McGuinness, D. (1979). Gender differences in response to spatial frequency and stimulus orientation. *Perception & Psychophysics, 26*(4), 319–324. https://doi.org/10.3758/BF03199887

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436.

Burg, A. (1966). Visual acuity as measured by dynamic and static tests: A comparative evaluation. *The Journal of Applied Psychology, 50*(6), 460–466. https://doi.org/10.1037/h0023982

Burg, A., & Hulbert, S. (1961). Dynamic visual acuity as related to age, sex, and static acuity. *Journal of Applied Psychology, 45*(2), 111–116. https://doi.org/10.1037/h0044200

Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews. Neuroscience, 7*(6), 477–484. https://doi.org/10.1038/nrn1909

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Laurence Erlbaum Associates.

Emerson, P. L. (1986). Observations on maximum-likelihood and Bayesian methods of forced-choice sequential threshold estimation. *Perception & Psychophysics, 39*(2), 151–153. https://doi.org/10.3758/bf03211498

Fider, N. A., & Komarova, N. L. (2019). Differences in color categorization manifested by males and females: A quantitative World Color Survey study. *Palgrave Communications, 5*(1), 142. https://doi.org/10.1057/s41599-019-0341-7

Greene, K. S., & Gynther, M. D. (1995). Blue versus periwinkle: Color identification and gender. *Perceptual and Motor Skills, 80*(1), 27–32. https://doi.org/10.2466/pms.1995.80.1.27

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Research, 34*(7), 885–912. https://doi.org/10.1016/0042-6989(94)90039-6

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? In *Perception* (Vol. 36).

Kulikowski, J. J., & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *The Journal of Physiology, 232*(1), 149–162. https://doi.org/10.1113/jphysiol.1973.sp010261

Luna, R., & Serrano-Pedraza, I. (2020). Evidence for different spatiotemporal mechanisms using duration thresholds: An individual differences approach. *Vision Research, 175*, 58–74. https://doi.org/10.1016/j.visres.2020.07.002

Mathew, J., Masson, G. S., & Danion, F. R. (2020). Sex differences in visuomotor tracking. *Scientific Reports, 10*(1), 11863. https://doi.org/10.1038/s41598-020-68069-0

McCarthy, M. M., Woolley, C. S., & Arnold, A. P. (2017). Incorporating sex as a biological variable in neuroscience: What do we gain? *Nature Reviews. Neuroscience, 18*(12), 707–708. https://doi.org/10.1038/nrn.2017.137

McGuinness, D. (1976). Away from a Unisex Psychology: Individual Differences in Visual Sensory and Perceptual Processes. *Perception, 5*(3), 279–294. https://doi.org/10.1068/p050279

Melnick, M. D., Harrison, B. R., Park, S., Bennetto, L., & Tadin, D. (2013). A Strong Interactive Link between Sensory Discriminations and Intelligence. *Current Biology, 23*(11), 1013–1017. https://doi.org/10.1016/j.cub.2013.04.053

Murray, I. J., Parry, N. R. A., McKeefry, D. J., & Panorgias, A. (2012). Sex-related differences in peripheral human color vision: A color matching study. *Journal of Vision, 12*(1). https://doi.org/10.1167/12.1.18

Murray, S. O., Schallmo, M.-P., Kolodny, T., Millin, R., Kale, A., Thomas, P., … Tadin, D. (2018). Sex Differences in Visual Motion Processing. *Current Biology : CB, 28*(17), 2794–2799.e3. https://doi.org/10.1016/j.cub.2018.06.014

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Pentland, A. (1980). Maximum likelihood estimation: The best PEST. *Perception & Psychophysics, 28*(4), 377–379. https://doi.org/10.3758/BF03204398

Serrano-Pedraza, I., Vancleef, K., Herbert, W., Goodship, N., Woodhouse, M., & Read, J. C. A. (2020). Efficient estimation of stereo thresholds: What slope should be assumed for the psychometric function? *PLOS ONE, 15*(1), e0226822.

Tadin, D., Lappin, J. S., Gilroy, L. A., & Blake, R. (2003). Perceptual consequences of centre–surround antagonism in visual motion processing. *Nature, 424*(6946), 312–315. https://doi.org/10.1038/nature01800

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research, 35*(17), 2503–2522. https://doi.org/10.1016/0042-6989(95)00016-X

Troche, S. J., Thomas, P., Tadin, D., & Rammsayer, T. H. (2018). On the relationship between spatial suppression, speed of information processing, and psychometric intelligence. *Intelligence, 67*, 11–18. https://doi.org/10.1016/j.intell.2017.12.002

Zaroff, C. M., Knutelska, M., & Frumkes, T. E. (2003). Variation in Stereoacuity: Normative Description, Fixation Disparity, and the Roles of Aging and Gender. *Investigative Opthalmology & Visual Science, 44*(2), 891. https://doi.org/10.1167/iovs.02-0361