**RESULTS**

Differences between stereo-anomalous and stereo-normal groups and cue scaffolding conditions were calculated through means comparison. Analysis of variance was used to establish differences between variables with more than two levels of comparison. One-ANOVAs were used for normal distribution variables and Kruskal-Wallis when distribution was not normal. Two sample comparison was computed using Student t-test if a normal distribution was confirmed and Wilxocon-Mann-Whitney otherwise. Kolmogoro-Simirnov test was used to confirm normal distribution of variables. The relationship between variables was determined through Pearson correlation test with The significance level set to 0.05. The software used was R-Statistics (v3.6.3).

**In-game performance accuracy**

DartBoard game provided 102,252 data points corresponding to the activity of the 20 participants. Examples of the results collected by participant are shown in Figure 1. Each dot represents the in-game accuracy threshold obtained after one hour of game play. Each cue scaffolding condition (game levels 1, 2 and 3) is represented by a different symbol. An exponential fitness curve (that will be explained afterwards in this chapter) is also included.

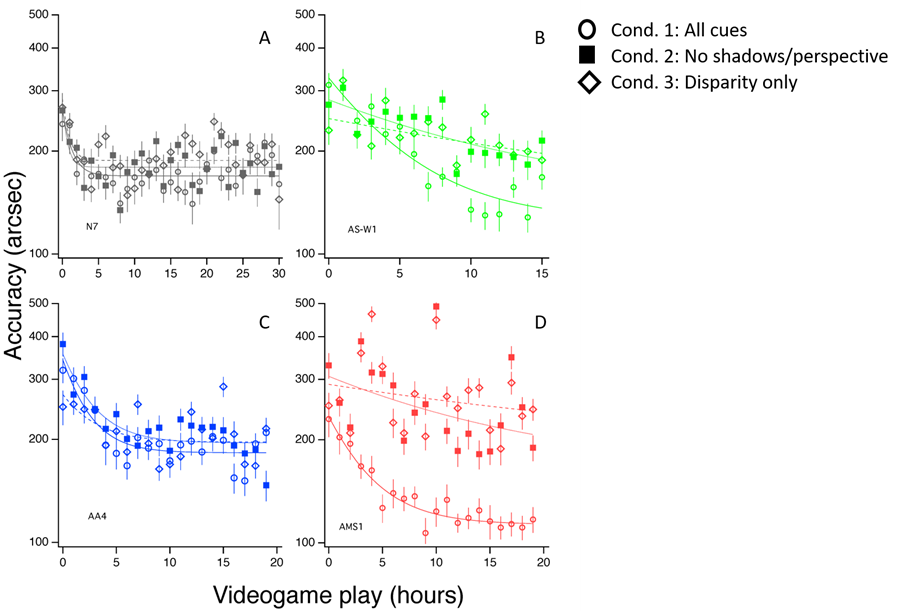


Figure 1. Examples of DartBoard in-game performance accuracy for four participants, one stereo-normal (figure 3-A, participant N7) and three stereo-anomalous: figure 3-B, stereo-weak participant AS-W1; figure 3-C, anisometropic participant AA4; and figure 3-D, micro strabismic participant AMS1. Condition 1 is represented hollow circles, Condition 2 as solid squares and Condition 3 as hollow diamonds.

Kolmogorov-Smirnov test shows that DartBoard game results don’t follow a normal distribution analyzed as a whole, neither considering the six subgroups obtained considering participant group (stereo-anomalous and stereo-normal) and cue scaffolding conditions (level 1, 2 and 3) (p-values = 0.000 in all cases). Therefore, non-parametric test will be used for the following analysis, Kruskal-Wallis for ANOVA and Mann-Whitney for mean comparison.

Means and confidence interval (calculated using Bootstrap method) for each group and condition are shown in Table 1.

Table 1. DartBoard in-game acuity. Means and confidence intervals, calculated using Bootstrap method, for each group and condition considered. Values in seconds of arc.

|  |  |  |
| --- | --- | --- |
| Condition | Stereo-normal | Stereo-anomalous |
| Level 1 | 184 [182,186] | 201 [199,203] |
| Level 2 | 190 [189,192] | 234 [231,237] |
| Level 3 | 191 [189,193] | 240 [238,243] |

ANOVA study using Kruskal-Wallis test shows that stereo-normal results are different from stereo-anomalous (p-values = 0.000 in both groups). The mean comparison between each pair of cue scaffolding conditions for both groups, stereo-normal and stereo-anomalous, (Mann-Whiney test) show that Level 1 and Level 2 results are different for both groups (p-values = 0.000 in both); and also are Level 1 and Level 3 results (p-values = 0.000 in both). When comparing Level 2 and Level 3 results, mean values are different for stereo-anomalous group (p-value = 0.006) but not for stereo-normal group (p-value = 0.534). The mean comparison between each group, stereo-normal and stereo-anomalous, at each cue scaffolding condition (Level 1, 2 or 3), using Mann-Whiney test, show that results are different for both groups at each condition (p-values = 0.000 in the tree conditions).

The results obtained by each patient have been adjusted using a exponential fitness (equation 1). From the fit, we calculated the asymptotic final threshold, time constant, and the ratio of baseline to asymptotic threshold (PPR) for each participant. The asymptotic final threshold represents the best performance achieved through training. The time constant represents the time it took for the exponential to fall 37% (1/e) of the starting value. Thus, it represents the rate of learning. Lastly, the PPR represents the amount of learning as a factor of the starting threshold (higher PPR indicates greater learning).

Equation 1.

PPR analysis using Kruskal-Wallis test show that there are not statistically significant differences due to condition for stereo-normal group (p-value = 0.845), meaning that stereo-normal participants behave similar in the three conditions. Meanwhile, stereo-anomalous group shows statistically significant differences due to condition (p-value = 0.037), that are also (and only) significant when we compare cue scaffolding conditions 1 and 3 (p-value = 0.009), meaning that stereo-anomalous participants behaves different in both conditions. Mann-Whiney test doesn’t show any statistically significant difference between groups in any of the three conditions, meaning that in any of tree cue scaffolding conditions the behavior of both groups is similar.

The other two parameters obtained in the exponential fit, final threshold and time constant, fail to show any statistically significant difference in any analysis.

Threshold analysis using Kruskal-Wallis test show no statistically significant differences due to condition in stereo-normal group (p-value = 0.730), neither in stereo-anomalous group (p-value = 0.665). We don’t find statistically significant differences between groups in any of the three conditions.

We don’t find statistically significant differences due to condition in stereo-normal (p-value = 0.619), neither in stereo-anomalous group (p-value = 0.111). Neither we find statistically significant differences between groups in any of the three conditions.

We analyze the correlation (Pearson) between the PPR and initial threshold (instead of final) as obtained in the exponential fitness (Table 2). We found a moderate or high correlation for stereo-normal participants in all conditions, that is only statistically significant in condition 3. Meanwhile stereo-anomalous participants behave different, showing an inverse or null correlation in conditions 2 and 3. This different behavior is even more evident if we consider all three cue scaffolding conditions together, obtaining a significant correlation and strong correlation in stereo-normal group (0.6510, p-value =0.000) whereas correlation is weak and not statistically significant in stereo-anomalous group (0.0763, p-value=0.694).

Table 2. DartBoard results. Analysis of exponential fitness parameters. PPR and initial threshold correlation using Pearson test for each cue scaffolding condition and all conditions together.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Stereo-normal | | Stereo-anomalous | |
| Condition | Correlation | p-value | Correlation | p-value |
| Level 1 | 0.5616 | 0.091 | 0.3775 | 0.282 |
| Level 2 | 0.6163 | 0.058 | -0.0676 | 0.854 |
| Level 3 | 0.7907 | 0.006 | -0.2618 | 0.465 |
| All levels | 0.6510 | 0.000 | 0.0763 | 0.694 |

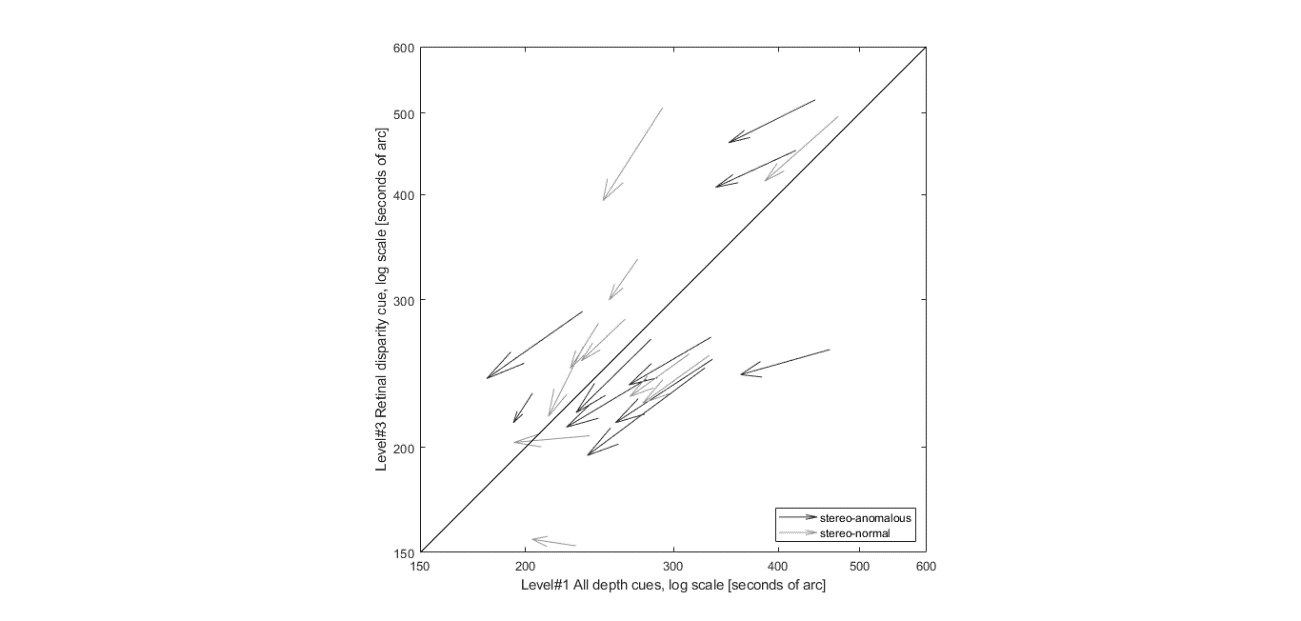
We finally perform a comparison between initial threshold and PPR relative to condition. The idea behind this analysis is to compare the improvements in cue scaffolding condition 1, where all depth cues are available, vs condition 3, where only retinal disparity is available. Those improvements may be different if initial thresholds are different for both conditions. Even the improvement may be depending on initial threshold. A participant who exhibits good in-game acuity when all depth cues are available (level 1) but performs poorly when only retinal disparity is present (level 3), is perhaps likely to improve more in level 3 than in level 1 after treatment. I.e. the lower the initial threshold ratio between level 1 and level 3, the lower the PPR ratio between level 1 and 3. In Figure 2 each participant is represented as an arrow, whose start point is the initial accuracy threshold, in level 1 for horizontal axis and Level 3 for vertical axis. Arrow horizontal length is proportional to the improvement (PPR) in the use of all depth cues when available (Level 1), whereas vertical length is the improvement in the use of retinal disparity alone (Level 3). All arrows point towards a game accuracy limit threshold that is located in the lower left corner of the graph. Stereo-normal participant’s arrows point to that threshold, independently if their start point is below or above the 45 degrees line. Nevertheless, stereo-anomalous participants are represented by an arrow that always shows an angle of improvement lower than 45 degrees, i.e. the improvement attributable to the use of retinal disparity is lower than the improvement attributable to the use of all depth cues together.

Figure 2. DratBoard in-game accuracy initial thresholds and PPR in two cue scaffolding conditions (Level 1 vs Level 3). Each participant is represented as an arrow, whose start point is the initial accuracy threshold, whose horizontal length the improvement in game accuracy (Level 1), and whose vertical length is the improvement in game accuracy (Level 3). Length of the arrows is escalated to facilitate identification of arrow direction, i.e. end point does not fit final threshold. Arrows whose start point is under the diagonal unity line show a better initial performance when all depth cues are present compared with the performance when only retinal disparity is available. Arrows whose angle is lower than 45 degrees show a bigger improvement for near to natural depth perception condition than for stereoacuity alone.

This interpretation of the graph is analyzed statistically. We find that PPR ratio between condition 1 and 3 correlates strongly with initial threshold ratio in both stereo-normal (0.5613) and stereo-anomalous groups (0.3775), but the correlation is statistically significant in stereo-normal group (p-value = 0.091) and not in stereo-anomalous (p-value = 0.282)(Pearson’s test).

The second game, Halloween, provides different results. As previously explained, Halloween in-game measurements result in the proportion of correct responses for each stereo demand at each cue scaffolding condition on each session. Stereo demand takes the values 1000, 800, 600, and 400 arc seconds. In order to measure improvements, we compare the proportion of right answers of the first 4 sessions (4 hours of game play) vs the last 4 sessions, for each participant at each condition. Those results are grouped, taking the mean values, into stereo-normal and stereo-anomalous, i.e. anisometropic, strabismic and stereo-weak (Figure 3). Values above the 45 degrees line indicate an improvement in terms of error reduction.

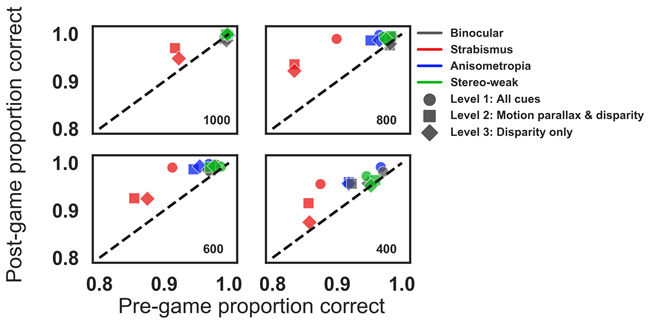
****

Figure 3. Halloween pre to post proportion correct for disparity values of 1000, 800, 600, and 400 arc seconds. Each symbol represents a different clue scaffolding condition (level 1, circles; level 2, squares; and level 3, diamonds) for participants with anisometropia (blue), strabismus (red), stereo-weak (green), and normal binocular (grey). Points above the unity line indicate a reduction of errors in the last six blocks compared to the first six blocks.

**Clinic and** **psychophysical test**

Pre and post results obtained by participants, grouped as stereo-normals and stereo-anomalous, are shown in Figure 4. Figure 4A shows Randot Circles Stereotest® (open symbols) and Random Dot 3 Stereo Acuity Test with Lea Symbols® (filled symbols). Figure 4B is obtained with psychophysical test: PDT (triangle), RDS small (small circle), RDS medium (medium circle) and RDS big (big circle). In both figures, colors indicate binocular condition: anisometropia (blue), strabismus (red), stereo-weak (green), and normal stereo (grey). Data under the diagonal unity line indicate an improvement in stereoacuity

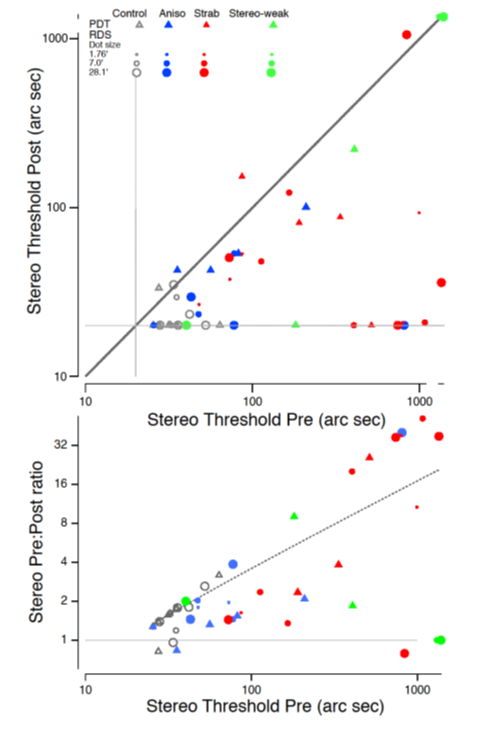
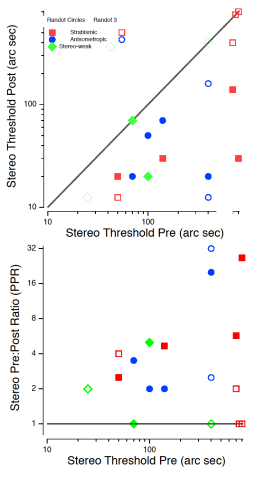


Figure 4. (A) Clinical stereoacuity improvement as a function of initial Randot Circles Stereotest® and Random Dot 3 Stereo Acuity Test with Lea Symbols®. Open symbols are used for Random Dot 3 where filled symbols for Random Circles. (B) Psychophysical stereoacuity improvement as a function of initial stereo threshold. Each symbol represents a psychophysical test: PDT (triangle), RDS small (small circle), RDS medium (medium circle) and RDS big (big circle). In both figures, colors indicate binocular condition: anisometropia (blue), strabismus (red), stereo-weak (green), and normal stereo (grey). Data under the diagonal unity line indicate an improvement in stereoacuity.

Statistical analysis starts analyzing mean differences for each test results (before and after treatment) between stereo-normal and stereo-anomalous groups (Table 3). We find statistically significant differences between groups for all test, but DST results after treatment (p-value = 0.052).

Table 3. Mean stereoacuity results obtained with each test before and after treatment, for stereo-normal and stereo-anomalous patients. Values in seconds of arc.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Stereo-normal | | Stereo-anomalous | |
|  | pre | post | pre | post |
| Randot Circles | 20 | 20 | 129 | 37 |
| Random Dot 3 | 13 | 13 | 339 | 105 |
| PDT | 30 | 22 | 155 | 63 |
| DRS | 25 | 21 | 204 | 62 |
| *DRS small* | 23 | 20 | 105 | 45 |
| *DRS medium* | 22 | 21 | 129 | 50 |
| *DRS big* | 28 | 21 | 263 | 62 |

Clinic tests show a clear improvement in stereoacuity for stereo-anomalous participants, as Table 4 resumes, but that improvement is not detected in stereo-normal group. Meanwhile, psychophysical tests confirm the improvement of stereo-anomalous participants and, importantly, also show improvements of stereo-normal participants.

Table 4. Clinic and Psychophysical test results. Participants where initial value is equal or bigger than after treatment, and participants where initial value is strictly bigger than final value (both stereo-normal group and stereo-anomalous group have N=10)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Stereo-normal | | Stereo-anomalous | |
|  | Initial ≥ Final | Initial > Final | Initial ≥ Final | Initial > Final |
| Randot Circles | 10 | 0 | 10 | 9 |
| Random Dot 3 | 10 | 0 | 10 | 7 |
| PDT | 8 | 7 | 8 | 8 |
| DRS | 9 | 7 | 10 | 9 |

Finally, we analyze if these tests show any correlation between the initial stereoacuity value and the PPR. Due to the small set of data, we perform the correlation considering the whole dataset of participants. All tests show a positive correlation (Table 5), meaning that the higher the initial value (stereoacuity deficit) the higher the PPR (percentage of improvement). Only in the case of the Random Dot 3 test the correlation is not statistically significant.

Table 5. Correlation coefficients and p-values using Pearson test between PPR and initial stereoacuity values, considering all patients.

|  |  |  |
| --- | --- | --- |
|  | All participants | |
|  | Correlation | p-value |
| Randot Circles | 0.7812 | 0.000 |
| Random Dot 3 | 0.3653 | 0.113 |
| PDT | 0.6118 | 0.004 |
| DRS | 0.5363 | 0.015 |