# Supplementary material

Katharina Krösl[1, 2] Dominik Bauer[1] Michael Schwärzler[2]
Henry Fuchs[3] Georg Suter[1] Michael Wimmer[1]

Figure S1 shows the use of measurement surfaces in HILITE, which allow us to ensure a minimum brightness for the illumination of escape-route signs, as required by norms or international standards like ISO3864-1.

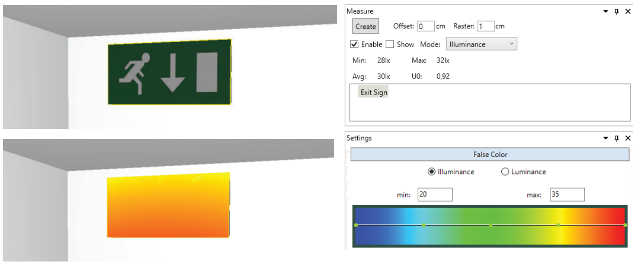

**Fig. S1** Application of the physically accurate light simulation software: A measurement surface (bottom left) with false color visualization (color scale from blue to red: 20 to 35 lux), positioned at the location of the exit sign (top left) ensures that the norm requirements regarding illumination (right) are fulfilled.

As discussed in Sect. 3.2 of our paper, the HTC Vive did not allow for a visual acuity better than 2.5 angular minutes. Figure S2 shows the distribution of measured visual acuity of all participants.

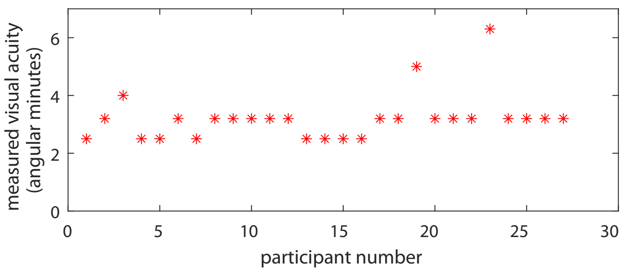


**Fig. S2** Distribution of the visual acuity measured during our eyesight test without blurring the participant's vision.

✉ Katharina Krösl
kkroesl@cg.tuwien.ac.at ; kroesl@vrvis.at

1 TU Wien, Vienna, Austria
2 VRVis Research Center, Vienna, Austria
3 University of North Carolina Chapel Hill, Chapel Hill, USA

| MRD values valid for 85% | | | | | | |
|---|---|---|---|---|---|---|
| | first test run | | second test run | | average of both runs | |
| **sign size and rotation** | **weak blur** | **strong blur** | **weak blur** | **strong blur** | **weak blur** | **strong blur** |
| 15cm, 0° | 6.4m | 4.4m | 6.9m | 4.1m | 6.7m | 4.3m |
| 15cm, 30° | 6.2m | 4.2m | 6.1m | 4.0m | 6.1m | 4.1m |
| 15cm, 60° | 4.6m | 2.8m | 5.0m | 3.0m | 4.8m | 2.9m |
| 30cm, 0° | 13.5m | 8.5m | 13.0m | 8.9m | 13.2m | 8.7m |
| 30cm, 30° | 13.1m | 7.1m | 13.2m | 8.2m | 13.1m | 7.6m |
| 30cm, 60° | 9.4m | 5.8m | 10.5m | 6.2m | 9.9m | 6.0m |

**Table S1** Measured maximum recognition distances that are valid for 85 percent of the participants of our user study.

## Result data

Table S1 shows the maximum MRD values, that are valid for at least 85 percent of our participants.

## Statistical analyses

For our statistical analyses we used a standard t-test or Welch's t-test, which is more reliable for samples that have unequal variances and unequal sample sizes. Using one of these tests we test the null hypothesis that samples come from populations with equal means. The null hypothesis is only rejected at the standard 5% significance level.

### *Learning effect*

Our measured MRDs during the first and second test run show similar distributions (see Fig. S3). We also performed a paired-sample t-test, since we obtained two measurements for each test case with each person (one in the first and one in the second test run). Our null hypothesis tests if samples come from populations with equal means. If the resulting p-values show a statistically significant value (a value below the 5% significance level), the null hypothesis is rejected. The results in Table S2 show that we can not reject the null hypothesis, since all p-values are greater than 0.05. This means that we found no evidance for any significant difference between measurements of our first and second test run and thus no evidance for a learning effect.

### *Comparison of normal sight and corrected sight*

The visualization in Fig. S4 shows the similarities and differences of the compared distributions of normal sight and corrected sight. We also performed an unpaired two-sample t-test,testing the null hypothesis that two data vectors (*normal sight*, *corrected sight*) are from populations with equal means. P-values greater than 0.05 mean that the null hypothesis can

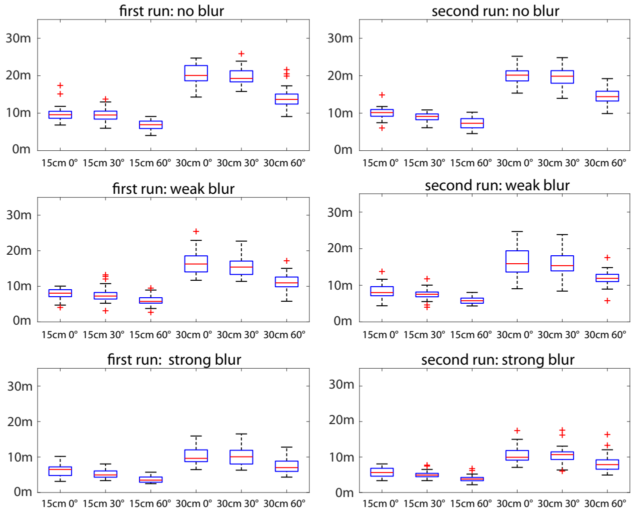


**Fig. S3** A visualization of the data from 2 runs of recognition tests with no blur (corresponding to 2.5 angular minutes, representing a mild vision impairment caused by the resolution of the HTC Vive headset), weak blur (corresponding to 5.0 angular minutes) and strong blur (corresponding to 8.0 angular minutes).

| sign size and rotation | no blur | weak blur | strong blur |
|---|---|---|---|
| 15 cm, 0° | 0.85607 | 0.16896 | 0.19661 |
| 15 cm, 30° | 0.07476 | 0.4002 | 0.85975 |
| 15 cm, 60° | 0.10199 | 0.78392 | 0.37166 |
| 30 cm, 0° | 0.94417 | 0.89359 | 0.26613 |
| 30 cm, 30° | 0.70949 | 0.70691 | 0.0609 |
| 30 cm, 60° | 0.32431 | 0.10355 | 0.06252 |

**Table S2** P-values (rounded to the $5^{th}$ position after decimal point) of Welch's t-test, comparing measurements of the first and second run of tests. Since all p-values are above the standard $\alpha = 0.05$ cutoff value, we can conclude that there is no significant difference between both distributions and therefore no evidence for a learning effect.

not be rejected at the default 5% significance level, or in other words, that there is no evidence for significant differences between MRDs of people with normal sight and people with corrected sight. The results of this statistical analysis yield four p-values below 0.05, as shown in Table S3.

A bonferroni correction, taking all 24 tests into account, would allow us to compare p-values to a cut-off value of $\sim 0.00208$ $(0.05 \div 24)$ and only recognize values below 0.00208 as significant enough to reject the null hypothesis. Applied on our data, this would leave only the one test with weak blur for 30cm high signs at 0° to yield a significant value in the first test run. Please note that even though a bonferroni correction reduces the probability for false positives, it also increases the probability for false negatives, which would obscure our results by hiding potentially significant values that suggest a rejection of the null hypothesis.

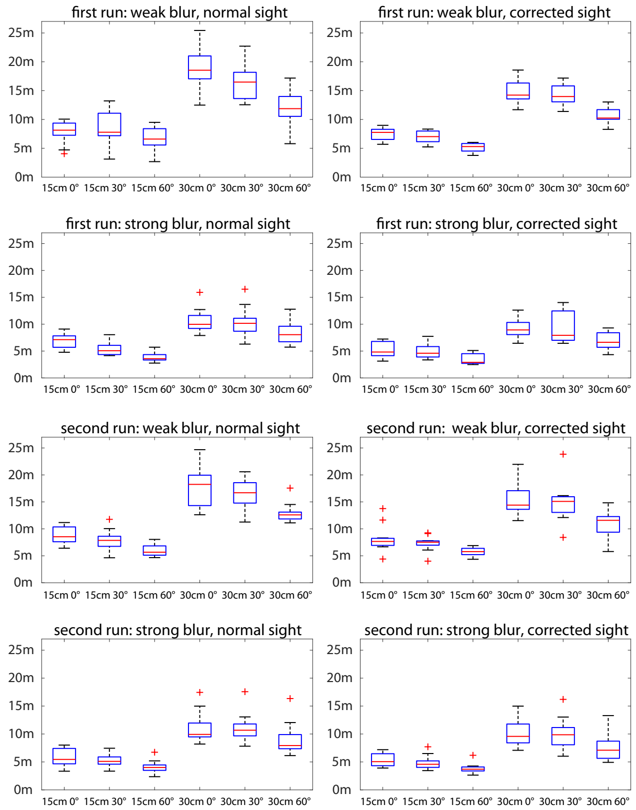


**Fig. S4** A Comparison of measured recognition distances of people with normal sight and people with corrected sight. Looking at results from the first test run, recognition distances measured under a weak blur show significant difference for 15cm tall signs at 60 degrees and 30cm tall signs at 0 and 30 degrees. The data also show a significant difference for recognition distances measured under a strong blur, but only for 15cm tall signs at 0 degrees. The statistical analysis of data from the second run showed no evidence for significant differences.

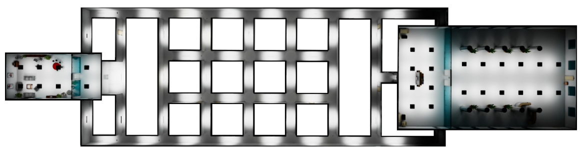

**Fig. S5** Overview of the test environment for interactive walkthroughs.

| | first test run | | second test run | |
|---|---|---|---|---|
| **sign size and rotation** | **weak blur** | **strong blur** | **weak blur** | **strong blur** |
| 15 cm, 0 | 0.46992 | **0.01118** | 0.407 | 0.55751 |
| 15 cm, 30 | 0.06462 | 0.36062 | 0.4249 | 0.53602 |
| 15 cm, 60 | **0.00993** | 0.33835 | 0.64268 | 0.50937 |
| 30 cm, 0 | **0.00122** | 0.13326 | 0.12975 | 0.4025 |
| 30 cm, 30 | **0.03462** | 0.40906 | 0.21391 | 0.33536 |
| 30 cm, 60 | 0.27634 | 0.07809 | 0.05311 | 0.20502 |

**Table S3** P-values (rounded to the 5$^{th}$ position after decimal point) of Welch's t-test, comparing measurements of people with normal sight to people with corrected sight (wearing contact lenses or glasses). 4 (out of 12) t-tests of the first test run show significant differences between the compared distributions (p-values are below the standard $\alpha = 0.05$ cutoff value), while none of the 12 t-tests on data from the second run show any significant p-values to reject the null hypothesis.