

André Montes Rodrigues

**Interactive visual analysis of hermetic and dense  
virtual models: the case of grain structures**

São Paulo

2020



André Montes Rodrigues

**Interactive visual analysis of hermetic and dense virtual  
models: the case of grain structures**

**Original Version**

PhD Thesis presented to Escola Politécnica,  
Universidade de São Paulo, in partial fulfill-  
ment of the requirements for the degree of  
Doctor of Science

Concentration Area: Electronic Systems

Advisor: Marcelo Knorich Zuffo

São Paulo

2020

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

#### Catálogo-na-publicação

Rodrigues, André Montes

Interactive visual analysis of hermetic and dense virtual models: the case of grain structures/ A. M. Rodrigues – versão original – São Paulo, 2020.

243p.

Tese (Doutorado) – Escola Politécnica da Universidade de São Paulo. Departamento de Sistemas Eletrônicos.

1. VISUALIZAÇÃO EM MÚLTIPLAS ESCALAS. 2. VISUALIZAÇÃO DE MICROESTRUTURAS. 3. MODELOS HERMÉTICOS. I. Universidade de São Paulo. Escola Politécnica. Departamento de Sistemas Eletrônicos. II. t.

*This work is dedicated to my dear Pamela and my parents for the unconditional support of my research career and also to all brave Brazilian scientists.*



# Acknowledgements

Special thanks to Ms. Pamela K. S. Bomfim and Prof. Ricardo M. L. Neto for the numerous fruitful scientific discussions and invaluable reviews of this work and related papers. Profs. Marcelo K. Zuffo, Roseli D. Lopes, and Ricardo enabled all the experimental work by allowing about a hundred students to dedicate some time as experiment participants, which demanded some temporary adjustments to the structure of four undergraduate and two graduate courses. I am also grateful for all student participants for their time, dedication, and seriousness.

Finally, this work would not be possible without the financial support of CAPES and material support of Caverna Digital, a laboratory of the Electric Engineering faculty and pioneer in Virtual Reality in Brazil since its foundation two decades ago. Congratulations to the lab and all personnel for the twenty-year birthday! I am very proud to be part of this group for more than seven years and to take part in many glorious achievements. There are still many more to come.





# Abstract

Novel interactive techniques were designed pursuing improvements on the comprehension of grain microstructures and similar hermetic (space-filling) honeycomb structures, based on cognitive principles and in the state of the art of interactive systems. A systematic review was carried out on multiscale visualization, which was considered an ideal paradigm in this work. In a second stage, development shifted to scientific experiments to investigate perceptual performance on fundamental aspects for grain structure comprehension - grain size extremes, regions with localized size heterogeneities, and shape of individual grains. Techniques were consequently adapted for experiments, centered on these three visual search tasks. Human-computer interaction knowledge and preliminary experiments informed the definition of experimental logistics and four dependent variables to attest performance, as well as appropriate statistical methods to analyze mixed factorial experiments results. The main experiment was carried out with two groups of 30 participants, comprising members of the intended target audience. Spatial ability pre-tests indicated homogeneity of spatial skills between the groups. Techniques are Discrete Sections (S), Dynamic Sectioning (DS), and GrainCrawler (GC). S simulates a visualization based on sequences of parallel sections performed over short distances, technically feasible to execute on real materials. DS improves upon S on spatial resolution, and the user can cut through the model interactively in any direction. GC was developed after the model's peculiarities, multiscale visualization potentials, and educational objectives. On GC, grain representation changes according to distance and perceptual objectives, a distinctive feature of the multiscale paradigm. Visualization modes were developed pursuing performance improvements, based on known depth perception principles, leveraging the model's three-dimensional nature. The main goal and hypothesis was that GC would surpass DS and particularly S, and 3D modes would overcome 2D. However, all techniques performed well overall and are deemed useful in improving spatial understanding of this type of structure. They allowed detecting size extremes and heterogeneous regions in a space-filling model containing more than a thousand objects. However, each technique had specific strengths. S excels in the agile detection of large grains and regions, as long as the sections are not widely spaced, and can be applied to real materials. DS allows rotation of the section plane and is ideal for digital models, being the most balanced and flexible technique. Finally, GC excelled in small grain detection and shape analysis, tasks characterized by high complexity and heavy cognitive load. Still, the multiscale paradigm allows combining each technique's strengths according to the intended application's analytical goals.

**Keywords:** Multiscale visualization, microstructure visualization, space-filling models.



# Resumo

## **Análise visual interativa de modelos virtuais herméticos e densos: o caso das estruturas de grãos.**

Duas ferramentas foram desenvolvidas visando melhorias no entendimento de microestruturas de grãos e outras estruturas alveolares similares, baseando-se em princípios cognitivos e no estado da arte em sistemas interativos. Foi realizada uma revisão sistemática em visualização multiescalar, paradigma considerado ideal no contexto deste trabalho. Em um segundo momento o foco passou a experimentos científicos, visando investigar o desempenho perceptivo em pontos considerados fundamentais ao melhor entendimento de estruturas de grãos - extremos de tamanho, regiões com variações localizadas de tamanho, além do fator de forma de grãos individuais. As ferramentas foram conseqüentemente adaptadas para execução de três tarefas experimentais, foi também detalhada a logística experimental e definidas quatro variáveis dependentes para atestar o desempenho, além de métodos estatísticos apropriados à análise dos resultados de experimentos fatoriais mistos. O experimento principal foi realizado com dois grupos de 30 participantes integrantes do público alvo, considerados homogêneos em termos de habilidades espaciais de acordo com os pré-testes realizados. As ferramentas são Seções Discretas (S), Seccionamento Dinâmico (DS) e GrainCrawler (GC). S simula uma visualização baseada em seqüências de seções paralelas realizadas em curtas distâncias, o que é tecnicamente viável em materiais reais. DS permite ao usuário navegar como se estivesse cortando o modelo. Já o GC segue o paradigma multiescalar, onde a representação dos grãos muda conforme a distância e objetivos perceptivos. Modos de visualização visaram melhorar o desempenho das ferramentas, se valendo de princípios como a visão em profundidade, tendo em vista a natureza tridimensional do modelo. A meta e hipótese principal era que GC superasse DS e principalmente S e que os modos 3D superassem os 2D. No entanto, as ferramentas desempenharam bem no geral e podem melhorar a compreensão especial deste tipo de estrutura. Todas permitiram detectar extremos de tamanho e regiões heterogêneas em um modelo hermético com mais de mil objetos, mas cada uma se destacou em pontos específicos. S pode ser aplicada a seções pouco espaçadas de materiais reais, além de ser ágil na detecção de grãos grandes. DS é ideal para modelos digitais, sendo a técnica mais equilibrada e flexível pois permite rotação do plano de corte. Já o GC destacou-se na detecção de grãos pequenos e na análise de forma, tarefas complexas e cognitivamente cansativas. De qualquer forma, o paradigma multiescalar permite mesclar as vantagens de cada ferramenta conforme as necessidades e objetivos analíticos da aplicação pretendida.

**Palavras-chave:** Visualização em múltiplas escalas, visualização de microestruturas, modelos herméticos.



# List of Figures

Figure 1 – Typical sections of real microstructures. Grain boundaries are chemically attacked to become visible. . . . .	1
Figure 2 – Visualization experiments with a digitally reconstructed microstructure: isosurface rendering (left), rendered video snapshot (center), printed model (right). Extracted from Lanzagorta et al. (1998). . . . .	2
Figure 3 – Grain structure multiscale analytics: shape of a single grain (left), neighborhood of this grain (center), heterogeneity or clustering around the grain (right). . . . .	6
Figure 4 – Purdue spatial ability test. The first sequence shows a 180° rotation around the vertical axis. The same operation applied to the second figure results in the second image after the arrow. Adapted figure (BODNER; GUAY, 1997). . . . .	10
Figure 5 – Example of Multiscale Visualization. Extracted from (MCCRAE et al., 2009). . . . .	13
Figure 6 – Multiscale visualization in materials: atoms and crystal planes, plane stacking, grain boundaries, grains, macroscopic behavior. . . . .	14
Figure 7 – Frustum culling (left) and Level of Detail (right) allows reducing graphic processing loads. . . . .	14
Figure 8 – Articles statistics. . . . .	18
Figure 9 – Articles statistics (cont.) . . . . .	19
Figure 10 – Articles statistics (cont.) . . . . .	20
Figure 11 – Navigation and representation issues. . . . .	21
Figure 12 – Structure and cognitive overload issues. . . . .	22
Figure 13 – Grain structure models: Dream 3d, low and high resolution models (left), extracted from (GROEBER; JACKSON, 2014). Tessellated model generated by Neper (center right), extracted from Quey, Dawson e Barbe (2011). Static section visualization on Graphite (right), screenshot extracted by the author. . . . .	25
Figure 14 – Prototyping visualization methods. Top: filtering geometry by grain isolation; orientation cube used in some CAD software; marking features of interest in space with icons; rendering method to improve depth perception. Center: communicating crystallographic orientation with ‘voxelized’ boundaries; alternative geometric representations and visual codes allows sample-wide analytics. Bottom: image sequence exemplifying guided navigation along a grain edge (highlighted by the circle). . . . .	31

Figure 15 – Left: model generated by Rodrigues et al. (2018). Center/right: each grain of the imported model is assigned a random clear color. . . . .	33
Figure 16 – Typical keyboard mapping for 6DOF navigation: WASD keys for translations, arrows for rotations. . . . .	34
Figure 17 – Mobile navigation. Typical touch operations for translations (top row): initial position, moving sideways, up and down; Intuitive movements for rotations (bottom row): initial position (center), rotating left around the green grain in the corner (left) and right around the same corner (right). . . . .	34
Figure 18 – Rendering modes. Top row: flat 2D, 2.5D and depth enhanced 2.5D. Bottom row: 3D anaglyph and 3D split screen. . . . .	35
Figure 19 – Depth enhanced 2.5D in three quality levels, from low to high. . . . .	36
Figure 20 – System implementation diagram. . . . .	37
Figure 21 – Advanced operations: marking (left) and picking (center left) individual grains; Isolating neighbors (center right); Fracture (right). . . . .	38
Figure 22 – Advanced operations: volume clipping. Three infinite planes filter geometry and define a new volume from the cubic sample. . . . .	38
Figure 23 – Adaptive interface. Desktop version (left): note the minimal menu along the page bottom. Mobile version (right): note the collapsible menu. . . . .	39
Figure 24 – Realtime sectioning implementation. Side view (left): camera near plane (dashed line) coincides with the sectioning plane. The camera is attached to the hinge point (red arrow) and is controlled by navigation commands. Flat 2D rendering (center) creates the illusion of a solid section by ignoring the shading component. Standard shading (right) : the user can see through the sectioning plane. . . . .	40
Figure 25 – Ground reference. Flat shading (left). Standard 2.5D (center). The ground looks more realistic with the shading effect created by ambient occlusion rendering in 2.5D depth enhanced mode (right). . . . .	42
Figure 26 – ‘Cubemap’ design possibilities. Crosshair section plane, and beveled cube (top left). Section perimeter highlight (bottom left). Transparent octant (top right). Opaque octant (bottom right). . . . .	43
Figure 27 – Face identifications in flat and 2.5D depth enhanced modes. . . . .	43
Figure 28 – Tentative configuration of the visualization area. . . . .	44
Figure 29 – GrainCrawler’s multiscale features. HR analysis (top row): distant view and focus view; Standard view for size analysis (bottom left) and proximal view for shape analysis (bottom right). . . . .	45
Figure 30 – Examples of virtual heads up displays. Panels can ease menu organization by grouping related functions in a consistent visual fashion (left). Frames may provide spatial references (right). Source: <a href="http://www.free3d.com">www.free3d.com</a> . . . . .	48

Figure 31 – System variations resulting from relevant combinations of interaction (from abstract to natural) and visualization modes (from flat 2D to split 3D). 2D interaction relies on abstract commands in devices such as keyboards and mouses. Rotations and touch commands (2.5D) with hands are considered more intuitive, while walking and head rotations are familiar natural movements (3D). Visualization, in turn, progresses from 2D, depth-enhanced 2D, to high FOV stereoscopic vision. Combinations highlighted in grey boxes are within the scope of this work. . . . .	54
Figure 32 – Nine possible shapes of a grain, given the combinations of three predefined 2D shapes (Square, Rectangle, Long rectangle) for each of the two main sides. Only the four highlighted shapes were found in the models.	58
Figure 33 – Example of a large grain within a cluster: note the ‘flower’ pattern around the central grain, formed by other large neighbors. . . . .	59
Figure 34 – A small grain appears briefly during translation. . . . .	59
Figure 35 – Rotation around equiaxial grain. . . . .	59
Figure 36 – Rotation around stretched (two images at the left) and highly stretched grains(two images at the right). . . . .	59
Figure 37 – Small grain heterogeneous region: note the clustering of small and medium-sized grains around the black one. . . . .	60
Figure 38 – A large grain visualized in GrainCrawler (highlighted in red). . . . .	60
Figure 39 – A small grain visualized in GrainCrawler (highlighted in red). Note the higher concentration of edges around the small grain. In this task the participant is instructed to pick 6 to 10 grains among the smaller using only translation movements, i.e, rotation is not allowed. . . . .	61
Figure 40 – GrainCrawler: visualizing a heterogeneous region comprising big grains. Note the wide space between adjacent grain centers. . . . .	61
Figure 41 – GrainCrawler: visualizing a heterogeneous region comprising small grains. Note the short distance between adjacent grain centers. . . . .	62
Figure 42 – GrainCrawler: visualizing each of the four shapes (equiaxial, discoid, brick and stretched disc). . . . .	62
Figure 43 – Detecting neighbors with a 180° clockwise rotation in flat 2D visualization. At least four angles are necessary to detect the 9 neighbors of the central pink grain and color is clearly a good mnemonic aid. Note the mirroring between first and last images. . . . .	62
Figure 44 – Although two angles might suffice to detect all neighbors in 2.5D modes, resolution limitations (left, zoom box) and shading properties of ambient occlusion rendering (right, zoom box) might mask small faces. . . . .	63
Figure 45 – Left: overall performance X total time. Right: tutorial performance (time) X overall performance. . . . .	65

Figure 46 – Tutorial time X performance on each task on T1 (left) and T2 (right).	66
Figure 47 – Tutorial time X performance on each task on T3 (left) and T4 (right).	66
Figure 48 – Overall performance X task time on T1 (left) and T2 (right).	67
Figure 49 – Overall performance X task time on T3 (left) and T4 (right).	67
Figure 50 – Correlation between performance and time in T1 (left) and T2 (right).	67
Figure 51 – Correlation between performance and time in T3 (left) and T4 (right).	68
Figure 52 – Visualizing a grain (white dot) with 2D and 2.5D modes. 1 <sup>st</sup> row: entering the grain, clearly detectable with enhanced 2.5D (right). 2 <sup>nd</sup> row: larger section of this grain, depth differences are noted. 3 <sup>rd</sup> row: the end of the grain is detectable with 2.5D (center) and unambiguously detected with enhanced 2.5D (right).	73
Figure 53 – Cluster detection. Flat 2D: the user cannot tell the difference between small grains or the beginning of bigger ones (left). 2.5D allows seeing the end and guessing that a cluster might be present (center). Depth-enhanced 2.5D shows clearly that adjacent grains have similar depths (right).	74
Figure 54 – Color palettes obtained with the HSL color model. The brighter palette (top) was achieved with 95% saturation and 85% on lightness while the bottom darker palette adopted 100% and 20% respectively.	75
Figure 55 – Color in 2.5D depth enhanced mode. Although the depth perception seems inferior using multiple colors, note that the perception of darker spots (‘holes’) is well preserved compared to a single sepia tone.	75
Figure 56 – Size task: boxplots of Time, Top30 and Factor responses per color mode and visualization mode. A significant difference was found in color mode on Top 30 (highlighted in grey).	77
Figure 57 – HR: boxplots of Time, Top30 and Factor responses per color mode and visualization mode. No significant difference was found.	78
Figure 58 – Size task: percentage of choices within five size ranking strata (smaller of the sample, top 5 smaller, top 20, top 70, top 250).	79
Figure 59 – Advanced HMD systems: Oculus Rift CV1, HTC Vive and Microsoft Hololens.	82
Figure 60 – Examples of 3D geometric transformations on the model. Standard model, front face (left images); 180° rotation (center); 90° rotation and mirroring on the vertical plane (right). On the first row only geometric changes are shown, while on the second row colors were altered as well.	86
Figure 61 – User study summary. Note that S is tested only in Flat 2D and GC do not use 2d flat rendering.	96
Figure 62 – Boxplots of MRT (top) and SBSOD (bottom) scores.	100
Figure 63 – Size tasks: Boxplots of TLX per pairs (top) and color modes (bottom).	102



Figure 64 – Shape task: Boxplot of TLX showing techniques (two left), pairs (central seven) and color modes (two right). . . . .	103
Figure 65 – HR tasks: Boxplot of TLX showing techniques (two left), pairs (central nine) and color modes (two right). . . . .	105
Figure 66 – Boxplots of scores A and B for Size tasks, by execution order. . . . .	107
Figure 67 – Boxplots of scores A and B for HR tasks, by execution order. . . . .	108
Figure 68 – Score A. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	141
Figure 69 – Score A counts by stratum. Top: techniques. Center: DS pairs. Bottom: GC pairs. . . . .	143
Figure 70 – Score B. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	145
Figure 71 – Score B = 10 counts (%). Top: techniques. Center: DS and GC. Bottom: color modes. . . . .	147
Figure 72 – Score C. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	149
Figure 73 – Score C means by score A stratum. Top: techniques. Center: DS. Bottom: GC. . . . .	151
Figure 74 – Score D. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	153
Figure 75 – Score D. Top: counts D = 1 per technique (%). Center: per DS and GC pairs. Bottom: per score A stratum. . . . .	155
Figure 76 – Score A. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	157
Figure 77 – Score A counts by stratum. Top: techniques. Center: DS pairs. Bottom: GC pairs. . . . .	159
Figure 78 – Score B. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	161
Figure 79 – Score B = 10 counts (%) by score A stratum. Top: techniques. Center: DS and GC pairs. Bottom: color modes. . . . .	163
Figure 80 – Score C. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	165

Figure 81 – Score C means by score A stratum. Top: techniques. Center: DS. Bottom: GC. . . . .	167
Figure 82 – Score D. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	169
Figure 83 – Score D. Top: counts D = 1 per technique (%). Center: per DS and GC pairs. Bottom: per score A stratum. . . . .	171
Figure 84 – Score A. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	173
Figure 85 – Score A counts by stratum. Top: techniques. Center: DS pairs. Bottom: GC pairs. . . . .	175
Figure 86 – Score B. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	177
Figure 87 – Score B = 10 counts (%) by score A stratum. Top: techniques. Center: DS and GC pairs. Bottom: color modes. . . . .	179
Figure 88 – Score C. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	181
Figure 89 – Score C means by score A stratum. Top: techniques. Center: DS pairs. Bottom: GC pairs. . . . .	183
Figure 90 – Score D. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	185
Figure 91 – Score D. Top: counts D = 1 per technique (%). Center: per DS and GC pairs. Bottom: per score A stratum. . . . .	187
Figure 92 – Score A. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	189
Figure 93 – Score A counts by stratum. Top: techniques. Center: DS pairs. Bottom: GC pairs. . . . .	191
Figure 94 – Score B. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	192
Figure 95 – Score B = 10 counts (%) by score A stratum. Top: techniques. Center: DS and GC pairs. Bottom: color modes. . . . .	195

Figure 96 – Score C. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	196
Figure 97 – Score C means by score A stratum. Top: techniques. Center: DS pairs. Bottom: GC pairs. . . . .	198
Figure 98 – Score D. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode. . . . .	200
Figure 99 – Score D. Top: counts D = 1 per technique (%). Center: per DS and GC pairs. Bottom: per score A stratum. . . . .	202
Figure 100–Shape task. Top: histograms of Time, Movement and Oscillations responses. . . . .	204

# List of Tables

Table 1 – HMD comparison table (09/2017).He:Head, Ha:Hand, W:walk, G:gestures, B:buttons . . . . .	82
Table 2 – Mean differences of MRT scores per experimental group . . . . .	99
Table 3 – Mean differences of SBSOD scores per experimental group. . . . .	99
Table 4 – TLX analysis: significant differences (%) per technique on Size tasks . .	101
Table 5 – TLX analysis: significant differences (%) per Color mode on Size tasks .	101
Table 6 – TLX analysis: significant differences (%) between pairs on Size tasks . .	103
Table 7 – TLX analysis: significant differences (%) per technique on Shape tasks .	104
Table 8 – TLX analysis: significant differences (%) per Color mode on Shape tasks	104
Table 9 – TLX analysis: significant differences (%) between pairs on Shape tasks .	104
Table 10 – TLX analysis: significant differences (%) per technique on HR tasks . .	104
Table 11 – TLX analysis: significant differences (%) per Color mode on HR tasks .	104
Table 12 – TLX analysis: significant differences (%) between pairs on HR tasks . .	105
Table 13 – Sequence of execution order per group. Each letter refers to a system variation . . . . .	106
Table 14 – Shape factor task: performance per order group (% of correct answers). Mean = 31, S.d.= 6 . . . . .	106
Table 15 – Results summary of Sb task: ranks per score. . . . .	109
Table 16 – Results summary of Ss task: ranks per score . . . . .	109
Table 17 – Results summary of HRb: ranks per score . . . . .	109
Table 18 – Results summary of HRs task: ranks per score . . . . .	110
Table 19 – Results summary of shape task: ranks per response . . . . .	110
Table 20 – Sb: group size and outliers on score A . . . . .	140
Table 21 – Sb: significant differences (%) per technique on score A . . . . .	140
Table 22 – Sb: significant differences (%) per Color mode on score A . . . . .	142
Table 23 – Sb: significant differences (%) between pairs on score A . . . . .	142
Table 24 – Sb: correlations of other variables with score A . . . . .	142
Table 25 – Sb: group size and outliers on score B . . . . .	144
Table 26 – Sb: significant differences (%) per technique on score B . . . . .	144
Table 27 – Sb: significant differences (%) per Color mode on score B . . . . .	144
Table 28 – Sb: significant differences (%) between pairs on score B . . . . .	146
Table 29 – Sb: correlations of other variables with score B . . . . .	146
Table 30 – Sb: group size and outliers on score C . . . . .	148
Table 31 – Sb:significant differences (%) per technique on score C . . . . .	148
Table 32 – Sb: significant differences (%) per Color mode on score C . . . . .	150
Table 33 – Sb: significant differences (%) between pairs on score C . . . . .	150

Table 34 – Sb: correlations of other variables with score C . . . . .	150
Table 35 – Sb: group size and outliers on score D . . . . .	152
Table 36 – Sb: significant differences (%) per technique on score D . . . . .	152
Table 37 – Sb: significant differences (%) per Color mode on score D . . . . .	154
Table 38 – Sb: significant differences (%) between pairs on score D . . . . .	154
Table 39 – Sb: correlations of other variables with score D . . . . .	156
Table 40 – Ss: group size and outliers on score A . . . . .	156
Table 41 – Ss: significant differences (%) per technique on score A . . . . .	156
Table 42 – Ss: significant differences (%) per Color mode on score A . . . . .	158
Table 43 – Ss: significant differences (%) between pairs on score A . . . . .	158
Table 44 – Ss: correlations of other variables with score A . . . . .	158
Table 45 – Ss: group size and outliers on score B . . . . .	160
Table 46 – Ss: significant differences (%) per technique on score B . . . . .	160
Table 47 – Ss: significant differences (%) per Color mode on score B . . . . .	162
Table 48 – Ss: significant differences (%) between pairs on score B . . . . .	162
Table 49 – Ss: group size and outliers on score C . . . . .	164
Table 50 – Ss: significant differences (%) per technique on score C . . . . .	164
Table 51 – Ss: significant differences (%) per Color mode on score C . . . . .	166
Table 52 – Ss: significant differences (%) between pairs on score C . . . . .	166
Table 53 – Ss: correlations of other variables with score C . . . . .	168
Table 54 – Ss: group size and outliers on score D . . . . .	168
Table 55 – Ss: significant differences (%) per technique on score D . . . . .	170
Table 56 – Ss: significant differences (%) per Color mode on score D . . . . .	170
Table 57 – Ss: significant differences (%) between pairs on score D . . . . .	170
Table 58 – Sb: correlations of other variables with score D . . . . .	172
Table 59 – HRb: group size and outliers on score A . . . . .	172
Table 60 – HRb: significant differences (%) per technique on score A . . . . .	174
Table 61 – HRb: significant differences (%) per Color mode, score A . . . . .	174
Table 62 – HRb: significant differences (%) between pairs, score A . . . . .	174
Table 63 – HRb: correlations of other variables with score A . . . . .	176
Table 64 – HRb: group size and outliers on score B . . . . .	176
Table 65 – HRb: significant differences (%) per technique, score B . . . . .	178
Table 66 – HRb: significant differences (%) per Color mode, score B . . . . .	178
Table 67 – HRb: significant differences (%) between pairs, score B . . . . .	178
Table 68 – HRb: correlations of other variables with score B . . . . .	178
Table 69 – HRb: Group size and outliers on score C . . . . .	180
Table 70 – HRb: significant differences (%) per technique, score C . . . . .	180
Table 71 – HRb: significant differences (%) per Color mode, score C . . . . .	180
Table 72 – HRb: significant differences (%) between pairs, score C . . . . .	182

Table 73 – HRb: correlations of other variables with score C . . . . .	182
Table 74 – HRb: group size and outliers on score D . . . . .	184
Table 75 – HRb: significant differences (%) per technique, score D . . . . .	184
Table 76 – HRb: significant differences (%) per Color mode, score D . . . . .	184
Table 77 – HRb: significant differences (%) between pairs, score D . . . . .	184
Table 78 – HRb: correlations of other variables with score D . . . . .	186
Table 79 – HRs: group size and outliers on score A . . . . .	188
Table 80 – HRs: significant differences (%) per technique, score A . . . . .	188
Table 81 – HRs: significant differences (%) per Color mode, score A . . . . .	188
Table 82 – HRs: significant differences (%) between pairs, score A . . . . .	188
Table 83 – HRs: correlations of other variables with score A. . . . .	190
Table 84 – HRs: group size and outliers on score B . . . . .	193
Table 85 – HRs: significant differences (%) per technique, score B . . . . .	193
Table 86 – HRs: significant differences (%) per Color mode on score B . . . . .	193
Table 87 – HRs: significant differences (%) between pairs, score B . . . . .	193
Table 88 – HRs: correlations of other variables with score B. . . . .	194
Table 89 – HRs: group size and outliers on score C . . . . .	194
Table 90 – HRs: significant differences (%) per technique, score C . . . . .	197
Table 91 – HRs: significant differences (%) per Color mode on score C . . . . .	197
Table 92 – HRs: significant differences (%) between pairs, score C . . . . .	197
Table 93 – HRs: correlations of other variables with score C. . . . .	199
Table 94 – HRs: group size and outliers on score D . . . . .	199
Table 95 – HRs: significant differences (%) per technique on score D . . . . .	201
Table 96 – HRs: significant differences (%) per Color mode on score D . . . . .	201
Table 97 – HRs: significant differences (%) between pairs, score D . . . . .	201
Table 98 – HRs: correlations of other variables with score D. . . . .	203
Table 99 – Shape analysis: group size and outliers on time response . . . . .	205
Table 100–Shape analysis: significant differences per technique on time response . .	205
Table 101–Shape analysis: significant differences per Color mode on time response .	205
Table 102–Shape analysis: significant differences between pairs on time response . .	205
Table 103–Shape task: correlations of other variables with time response . . . . .	206

# List of abbreviations and acronyms

VR	Virtual Reality
HMD	Head-Mounted Display
3D	Three-dimensional
2D	Two-dimensional or 2D visualization mode
DS	Dynamic Sectioning (experimental technique) or 2D visualization mode within Dynamic Sectioning technique
S	Discrete Sectioning - experimental technique
GC	GrainCrawler - experimental technique
Techniques	Experimental interactive visualization techniques (S, DS or GC)
2.5D	Visualization mode 2.5D in conventional 2D monitors
3Da	3D visualization mode using anaglyph goggles
3Ds	3D visualization mode using split-screen stereoscopic devices
Mode	Visualization mode, either 2D, 2.5D, 3Da or 3Ds
DSd	Technique DS in 2.5D visualization mode
DSa	Technique DS in 3Da visualization mode
DSs	Technique DS in 3Ds visualization mode
GCd	Technique GC in 2.5D visualization mode
GCa	Technique GC in 3Da visualization mode
GCs	Technique GC in 3Ds visualization mode
Pair	Combination of a technique with a visualization mode: S, DS, DSd, DSa, DSs, GCd, GCa, GCs
Sb	Size experimental task - detection of bigger grains
Ss	Size experimental task - detection of smaller grains
HR	Heterogeneous regions (as an experimental task)

HRb	Heterogeneous regions experimental task - detection of bigger grain clustering
HRs	Heterogeneous regions experimental task - detection of smaller grain clustering
ANOVA	Analysis of Variance
GLMM	Generalized Linear Mixed Model
LMM	Linear Mixed Model
ART	Aligned Rank Transform
WS	Within-subjects
BS	Between-subjects
DOF	Degree of Freedom
HCI	Human-Computer Interaction
TLX	NASA Task Load Index
SBSOD	Santa Barbara Sense of Direction



# Contents

<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
<b>1.1</b>	<b>Objective</b>	<b>2</b>
<b>1.2</b>	<b>Motivation</b>	<b>2</b>
<b>1.3</b>	<b>Contributions</b>	<b>3</b>
<b>1.4</b>	<b>Manuscript Organization</b>	<b>4</b>
<b>2</b>	<b>LITERATURE REVIEW</b>	<b>5</b>
<b>2.1</b>	<b>Relevant topics for microstructure learning</b>	<b>5</b>
<b>2.2</b>	<b>Human-computer interaction (HCI) foundations</b>	<b>7</b>
2.2.1	Spatial ability and navigational skills	9
2.2.2	Visual search theories	11
<b>2.3</b>	<b>Multiscale visualization systems: a systematic review</b>	<b>13</b>
2.3.1	Review protocol	14
2.3.1.1	Research questions	14
2.3.1.2	Search strategy	15
2.3.1.3	Study selection criteria and procedures	15
2.3.1.4	Results extraction and compilation strategy	16
2.3.2	Execution	16
2.3.3	Results	16
2.3.3.1	Articles statistics	16
2.3.3.2	Issues and solutions	17
2.3.3.3	Navigation	17
2.3.3.4	Structure	21
2.3.3.5	Cognitive overload	22
2.3.3.6	Representation	22
2.3.3.7	Interaction	23
2.3.3.8	Cognitive issues	23
<b>2.4</b>	<b>Current approaches and tools for microstructure visualization</b>	<b>24</b>
<b>2.5</b>	<b>Available development platforms for 3D interactive software</b>	<b>25</b>
<b>2.6</b>	<b>Literature review summary</b>	<b>27</b>
<b>3</b>	<b>DEVELOPMENT</b>	<b>29</b>
<b>3.1</b>	<b>System design: features in a multiscale paradigm</b>	<b>29</b>
<b>3.2</b>	<b>Target audience</b>	<b>32</b>
<b>3.3</b>	<b>System features</b>	<b>32</b>
3.3.1	System implementation description	33

3.3.2	Basic capabilities . . . . .	33
3.3.3	Advanced capabilities . . . . .	36
3.3.4	Interface design . . . . .	38
<b>3.4</b>	<b>Realtime interactive sectioning technique . . . . .</b>	<b>39</b>
<b>3.5</b>	<b>GrainCrawler technique . . . . .</b>	<b>41</b>
3.5.1	Spatial reference features . . . . .	42
3.5.2	Multiscale features . . . . .	42
3.5.3	Prospective features . . . . .	44
3.5.3.1	Guided animations . . . . .	44
3.5.3.2	Gamification possibilities to boost motivation in serious interactive software . . . . .	46
3.5.3.2.1	Relevant game elements . . . . .	47
3.5.3.2.2	Game mechanics ideas . . . . .	48
3.5.3.2.3	Prospective mechanics for experimental setups . . . . .	49
<b>3.6</b>	<b>Development summary . . . . .</b>	<b>51</b>
<b>4</b>	<b>EVALUATION . . . . .</b>	<b>53</b>
<b>4.1</b>	<b>Conceptual framework: the interaction-visualization immersive continuum . . . . .</b>	<b>53</b>
<b>4.2</b>	<b>Selection of dependent variables and evaluation strategy . . . . .</b>	<b>55</b>
<b>4.3</b>	<b>Preliminary task design . . . . .</b>	<b>56</b>
4.3.1	Size tasks . . . . .	56
4.3.2	Shape task . . . . .	56
4.3.3	Heterogeneous regions tasks . . . . .	57
4.3.4	Possibilities for grain neighborhood analysis . . . . .	57
<b>4.4</b>	<b>Preliminary user study . . . . .</b>	<b>63</b>
4.4.1	Execution . . . . .	64
4.4.2	Desktop platform results . . . . .	64
4.4.2.1	Response distributions . . . . .	64
4.4.2.2	Results . . . . .	65
4.4.2.3	Results analysis and impressions . . . . .	68
4.4.2.4	Student feedback . . . . .	69
4.4.2.5	Preliminary user study final remarks . . . . .	70
<b>4.5</b>	<b>Second preliminary experiment . . . . .</b>	<b>72</b>
4.5.1	Discussing and detailing 2.5D modes . . . . .	72
4.5.2	Color . . . . .	74
4.5.3	Experiment design - 2.5D modes and color . . . . .	76
4.5.4	Results . . . . .	76
<b>4.6</b>	<b>Main study design . . . . .</b>	<b>80</b>
4.6.1	3D anaglyph issues . . . . .	81
4.6.2	Influences of HMD parameters on 3D split-screen visualization . . . . .	81

4.6.3	List of independent variables . . . . .	83
4.6.4	Discussion of experimental design possibilities . . . . .	83
4.6.5	Details of experimental techniques . . . . .	84
4.6.5.1	Benchmark technique: discrete sections . . . . .	84
4.6.5.2	Technique 2: interactive sectioning . . . . .	85
4.6.5.3	Technique 3: GrainCrawler . . . . .	87
4.6.6	Consolidated experimental procedure . . . . .	88
4.6.6.1	Pre and post tests . . . . .	88
4.6.6.2	Main experiment . . . . .	89
4.6.6.3	Dependent variables and auxiliary measurements . . . . .	90
4.6.7	Possible confounding factors . . . . .	91
4.6.8	Tutorials details . . . . .	91
4.6.8.1	Navigation . . . . .	91
4.6.8.2	Feature detection sensitivity . . . . .	92
4.6.9	Data analysis strategy and procedures . . . . .	92
4.6.10	Hypothesis statement . . . . .	94
<b>4.7</b>	<b>Alternative: Single-subject and small-N baseline study designs . . .</b>	<b>94</b>
<b>4.8</b>	<b>Evaluation summary . . . . .</b>	<b>96</b>
<b>5</b>	<b>RESULTS . . . . .</b>	<b>99</b>
<b>5.1</b>	<b>Pre-tests on experimental groups . . . . .</b>	<b>99</b>
5.1.1	MRT . . . . .	99
5.1.2	SBSOD . . . . .	99
<b>5.2</b>	<b>TLX analysis . . . . .</b>	<b>101</b>
5.2.1	Size tasks . . . . .	101
5.2.2	Shape task . . . . .	101
5.2.3	Heterogeneous Regions tasks . . . . .	104
5.2.4	Discomfort on 3D conditions and VR sickness . . . . .	104
<b>5.3</b>	<b>Task order influence assessment . . . . .</b>	<b>105</b>
<b>5.4</b>	<b>Results summary . . . . .</b>	<b>106</b>
5.4.1	Size - Big . . . . .	106
5.4.2	Size - Small . . . . .	106
5.4.3	HRb - Bigger grain clustering . . . . .	109
5.4.4	HRs - Smaller grain clustering . . . . .	110
5.4.5	Shape task . . . . .	110
5.4.6	Final comments on results . . . . .	110
<b>6</b>	<b>DISCUSSION . . . . .</b>	<b>113</b>
<b>6.1</b>	<b>Discussion summary . . . . .</b>	<b>118</b>

<b>7</b>	<b>CONCLUSIONS</b> . . . . .	<b>121</b>
<b>7.1</b>	<b>Limitations and future work</b> . . . . .	<b>122</b>

	<b>BIBLIOGRAPHY</b> . . . . .	<b>125</b>
--	-------------------------------	------------

## **APPENDIX** **137**

	<b>APPENDIX A – DETAILED RESULTS</b> . . . . .	<b>139</b>
--	------------------------------------------------	------------

<b>A.1</b>	<b>Size task analysis: bigger grains (Sb)</b> . . . . .	<b>140</b>
A.1.1	Score A . . . . .	140
A.1.1.1	Distributions and outliers . . . . .	140
A.1.1.2	Boxplots, ANOVAs and comparisons . . . . .	140
A.1.1.3	Strata analysis . . . . .	140
A.1.1.4	Correlations . . . . .	142
A.1.1.4.1	Score A summary . . . . .	142
A.1.2	Score B . . . . .	144
A.1.2.1	Distributions and outliers . . . . .	144
A.1.2.2	Boxplots, ANOVAs, and comparisons . . . . .	144
A.1.2.3	Analysis of score B = 10 . . . . .	146
A.1.2.4	Correlations . . . . .	146
A.1.2.4.1	Score B summary . . . . .	146
A.1.3	Score C . . . . .	148
A.1.3.1	Distributions and outliers . . . . .	148
A.1.3.2	Boxplots, ANOVAs, and comparisons . . . . .	148
A.1.3.3	Strata analysis . . . . .	148
A.1.3.4	Correlations . . . . .	150
A.1.3.4.1	Score C summary . . . . .	152
A.1.4	Score D . . . . .	152
A.1.4.1	Distributions and outliers . . . . .	152
A.1.4.2	Boxplots, ANOVAs, and comparisons . . . . .	152
A.1.4.3	Strata analysis . . . . .	154
A.1.4.4	Correlations . . . . .	154
A.1.4.4.1	Score D summary . . . . .	154
<b>A.2</b>	<b>Size task analysis: smaller grains (Ss)</b> . . . . .	<b>156</b>
A.2.1	Score A . . . . .	156
A.2.1.1	Distributions and outliers . . . . .	156
A.2.1.2	Boxplots, ANOVAs, and comparisons . . . . .	156
A.2.1.3	Strata Analysis . . . . .	158

A.2.1.4	Correlations . . . . .	158
A.2.1.4.1	Score A summary . . . . .	160
A.2.2	Score B . . . . .	160
A.2.2.1	Distributions and outliers . . . . .	160
A.2.2.2	Boxplots, ANOVAs and comparisons . . . . .	160
A.2.2.3	Strata analysis . . . . .	162
A.2.2.4	Correlations . . . . .	162
A.2.2.4.1	Score B summary . . . . .	162
A.2.3	Score C . . . . .	164
A.2.3.1	Distributions and outliers . . . . .	164
A.2.3.2	Boxplots, ANOVAs, and comparisons . . . . .	164
A.2.3.3	Strata analysis . . . . .	166
A.2.3.4	Correlations . . . . .	166
A.2.3.4.1	Score C summary . . . . .	168
A.2.4	Score D . . . . .	168
A.2.4.1	Distributions and outliers . . . . .	168
A.2.4.2	Boxplots, ANOVAs, and comparisons . . . . .	170
A.2.4.3	Strata analysis . . . . .	170
A.2.4.4	Correlations . . . . .	172
A.2.4.4.1	Score D summary . . . . .	172
<b>A.3</b>	<b>Heterogeneous regions task: bigger grains clustering . . . . .</b>	<b>172</b>
A.3.1	Score A . . . . .	172
A.3.1.1	Distributions and outliers . . . . .	172
A.3.1.2	Boxplots, ANOVAs, and comparisons . . . . .	172
A.3.1.3	Strata analysis . . . . .	174
A.3.1.4	Correlations . . . . .	174
A.3.1.4.1	Score A summary . . . . .	176
A.3.2	Score B . . . . .	176
A.3.2.1	Distributions and outliers . . . . .	176
A.3.2.2	Boxplots, ANOVAs, and comparisons . . . . .	176
A.3.2.3	Strata analysis . . . . .	176
A.3.2.4	Correlations . . . . .	178
A.3.2.4.1	Score B summary . . . . .	178
A.3.3	Score C . . . . .	180
A.3.3.1	Distributions and outliers . . . . .	180
A.3.3.2	Boxplots, ANOVAs, and comparisons . . . . .	180
A.3.3.3	Strata analysis . . . . .	180
A.3.3.4	Correlations . . . . .	182
A.3.3.4.1	Score C summary . . . . .	182

A.3.4	Score D . . . . .	184
A.3.4.1	Distributions and outliers . . . . .	184
A.3.4.2	Boxplots, ANOVAs, and comparisons . . . . .	184
A.3.4.3	Strata analysis . . . . .	186
A.3.4.4	Correlations . . . . .	186
A.3.4.4.1	Score D summary . . . . .	186
<b>A.4</b>	<b>Heterogeneous regions task: smaller grains clustering . . . . .</b>	<b>186</b>
A.4.1	Score A . . . . .	186
A.4.1.1	Distributions and outliers . . . . .	186
A.4.1.2	Boxplots, ANOVAs, and comparisons . . . . .	188
A.4.1.3	Strata analysis . . . . .	190
A.4.1.4	Correlations . . . . .	190
A.4.1.4.1	Score A summary . . . . .	190
A.4.2	Score B . . . . .	190
A.4.2.1	Distributions and outliers . . . . .	190
A.4.2.2	Boxplots, ANOVAs, and comparisons . . . . .	193
A.4.2.3	Strata analysis . . . . .	193
A.4.2.4	Correlations . . . . .	194
A.4.2.4.1	Score B summary . . . . .	194
A.4.3	Score C . . . . .	194
A.4.3.1	Distributions and outliers . . . . .	194
A.4.3.2	Boxplots, ANOVAs, and comparisons . . . . .	197
A.4.3.3	Strata analysis . . . . .	197
A.4.3.4	Correlations . . . . .	199
A.4.3.4.1	Score C summary . . . . .	199
A.4.4	Score D . . . . .	199
A.4.4.1	Distributions and outliers . . . . .	199
A.4.4.2	Boxplots, ANOVAs, and comparisons . . . . .	199
A.4.4.3	Strata analysis . . . . .	201
A.4.4.4	Correlations . . . . .	201
A.4.4.4.1	Score D summary . . . . .	203
<b>A.5</b>	<b>Shape task analysis . . . . .</b>	<b>203</b>
A.5.1	Main score: correct hits . . . . .	203
A.5.2	Time response . . . . .	203
A.5.2.1	Distributions and outliers . . . . .	203
A.5.2.2	Boxplots, ANOVAs, and comparisons . . . . .	205
A.5.2.3	Correlations . . . . .	205
A.5.2.3.1	Shape task summary . . . . .	206

<b>ANNEX</b>	<b>207</b>
<b>ANNEX A – ETHICS COMMITTEE APPROVAL . . . . .</b>	<b>209</b>
<b>ANNEX B – PARTICIPANT CONSENT FORM . . . . .</b>	<b>215</b>
<b>ANNEX C – MRT FORM . . . . .</b>	<b>219</b>
<b>ANNEX D – SBSOD FORM . . . . .</b>	<b>237</b>
<b>ANNEX E – TLX FORM . . . . .</b>	<b>241</b>





# 1 Introduction

In the materials realm, grain structures are traditionally visualized with 2D sections (Figure 1), aiming to characterize grain properties by way of stereological methods (UNDERWOOD, 1973). However, the limitations of such methods are being questioned recently both for visualization and materials characterization (FISCHER et al., 2012). Those techniques are also used in other fields such as petrography, histology, and neuroanatomy. For the sake of simplicity, the term ‘grain’ is used on the remainder of this text to refer to the cells of any honeycomb-shaped structure. An explicit limitation of single sections is that it cannot reveal much about the geometry of the sectioned grains or spatial relationships. In fact, the exclusive use of sections usually leads to misconceptions about grain shape and size among students. Interactive sectioning can help hinder such misconceptions (LANZAGORTA et al., 1998), allowing infinite sections to be visualized continuously in a brief period, an ideal solution on a cognitive visualization perspective. The original video sequencing technique introduced by Lanzagorta et al. (1998) was an early experiment on continuous sectioning, but it was not further investigated in the literature (figure 2).

However, this sequencing technique was restricted to a single straight ‘camera’ path, as the sections were obtained from real materials. The solution to this problem is 3D digitally reconstructed samples, as exemplified by Ullah et al. (2013). The method of choice for visualization adopts selective volumetric rendering, filtering features of interest such as grains and phases. Volume rendering specifies color and opacity levels for every voxel or ‘volumetric pixel’. Static sections are obtained by surface rendering, extracting isosurfaces of the region of interest (KAUFMAN; MUELLER, 2005). Besides the possibility of satisfactory real-time rendering allowed by recent experimental libraries such as Gigavoxels (CRASSIN et al., 2009), and information filtering capabilities, volume rendering is usually adopted

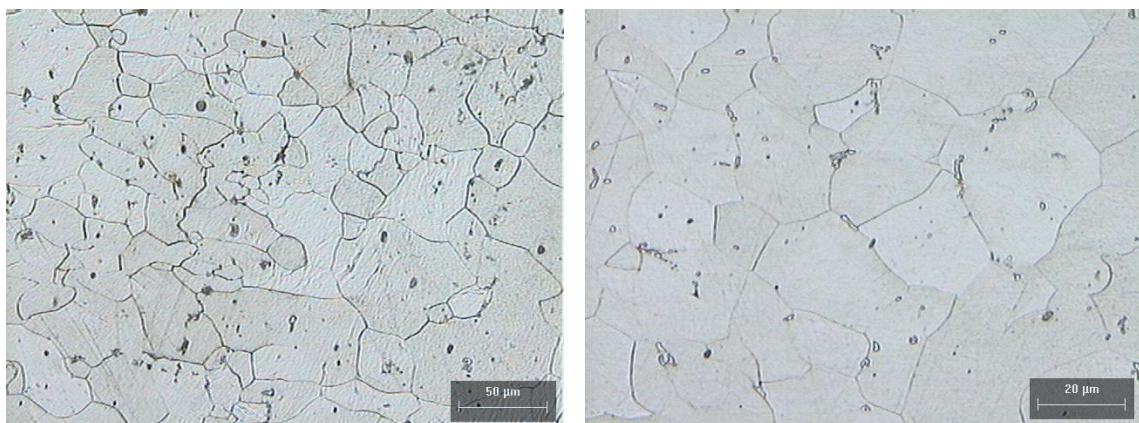


Figure 1 – Typical sections of real microstructures. Grain boundaries are chemically attacked to become visible.

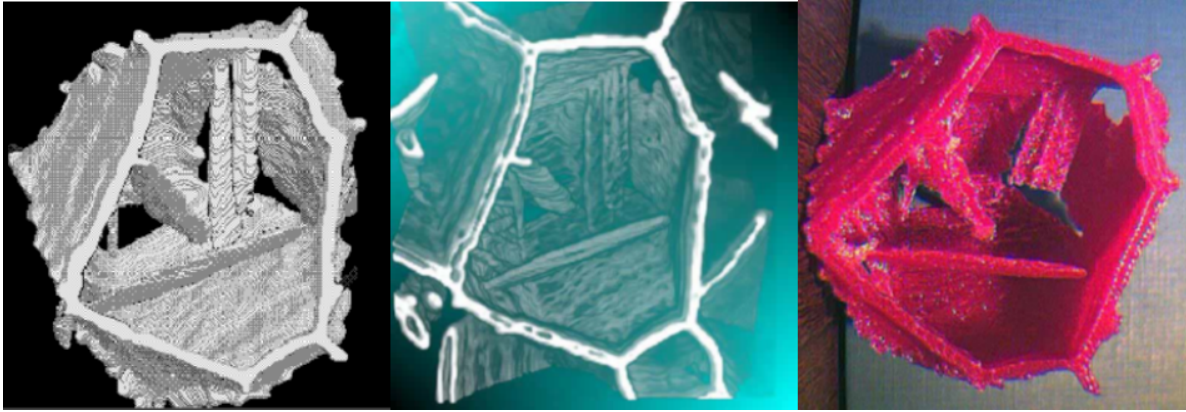


Figure 2 – Visualization experiments with a digitally reconstructed microstructure: isosurface rendering (left), rendered video snapshot (center), printed model (right). Extracted from [Lanzagorta et al. \(1998\)](#).

within a static visualization paradigm that lacks domain-specific interaction techniques. There are several technical interaction requirements not satisfied by existing software and relevant materials' science concepts not adequately supported by such techniques.

It should be clear here what is meant by the term 'visualization'. According to [Rhyne et al. \(1999\)](#), the most known aspect of visualization is rendering, where most fundamental problems were already solved. However, the authors state that 'sensory design' is the greatest challenge of visualization, involving human perception or cognitive processing. Here the effort lies in representing abstract concepts, not realistic objects, to facilitate the understanding of complex phenomena. The goal is to define and emphasize the relevant features for a specific application while minimizing unimportant details and to ease building a mental hierarchy of meaningful relationships ([RHYNE et al., 1999](#)). Accordingly, visualization is considered here as a domain-specific method that takes advantage of available filtering and rendering possibilities, orchestrated by an interaction rationale, in this case, aiming at novel interaction paradigms for grain structures.

## 1.1 Objective

This work presents the development and assessment of novel interactive visualization methods for space-filling honeycomb models, specifically virtual grain structures, aiming to improve the learning experience by allowing superior spatial comprehension. An important goal is making those methods accessible to teachers, instructors, and the target audience.

## 1.2 Motivation

Recent research ([FISCHER et al., 2012](#)) discussed limitations of stereological methods typically used to assess grain properties, supporting the development of 3D methods.

Early experiments with 3D interactive visualization methods (LANZAGORTA et al., 1998) allowed researchers to discover features of cementite precipitates in high-carbon steel, asserting limitations of 2D methods and the potential of 3D interactive technologies.

Hornbogen (1986) proposed the first comprehensive classification framework for microstructures based on geometric patterns, in a pioneering effort to define the essence of microstructures. Recent work by DeHoff (2012) demonstrates that the endeavor to define and understand microstructures clearly is far from finished, stating that microstructures are inherently 3D, space-filling, not-regular and not-random geometric entities, and their description should be framed in terms of geometric concepts, reinforcing Hornbogen's stance. Therefore, it is paramount to invest in spatial exploration techniques that allow experiencing the inherent 3D geometric nature of microstructures and tackling complex visualization challenges such as the space-filling character.

Research (HUK, 2006; BYRNE, 1996) suggests that direct visualization and spatial interaction with the concerned phenomena reduce abstractions, providing better support for understanding complex systems than passive media such as text or images. It is known that spatial interaction aids deciphering complex systems by stimulating conceptual connections – cognitive maps emerge from the detection of functional relationships and causal correlation (ALIAS, 2000a). Immersion and interactivity can also boost enthusiasm and attenuate learning barriers (HUK, 2006; BYRNE, 1996). In this sense, the adoption of interactive microstructural digital models can help to unveil grain size and shape, dislocation density, volumetric fraction, spatial distribution of phases, and microcracks on a cognitive level. Engineering students are expected to benefit more from interactive models, given that they usually have good spatial ability (HUK, 2006; ALIAS, 2000a). However, specific visualization mechanisms for learning purposes are still absent within materials research, especially in the microstructural scale, and valid and proper microstructural models are neither readily available nor straightforward to obtain.

## 1.3 Contributions

This work indicates promising directions and contributes novel solutions for interactive visual analysis of grain structures aimed at educational purposes. The adopted polygonal 3D grain microstructures were built and validated in previous work (RODRIGUES et al., 2017), featuring data structures that facilitate the development of visualization prototypes in game engines and other frameworks. Domain-specific interactive experimental tools informed by learning and cognitive research are presented. Although this work is restricted to a grain model, the intention is to add other microstructural features (e.g., fibers, pores, second phases) and to investigate proper interaction techniques for those elements.

It is important to stress the relevance of such interactive techniques for exploring

similar space-filling 3D tessellations in general, such as foams, rock aggregates (geology), powders and biological tissues. Thus, the approaches, results and findings are not exclusive to materials science education, being valuable to other fields as well. However, the case study depicted here is about grain microstructures and will include the relevant specifics.

Therefore, the main contributions of this research projects are:

- Discussion of possible learning objectives for materials science topics, prospective visualization approaches to improve microstructure learning and analysis of existing visualization tools.
- Systematic Review of multiscale visualization systems.
- Cross-platform WebGL visualization system, supporting interaction and several geometric operations on space-filling digital models in section and multiscale modes.
- Evaluations and comparisons of three interactive techniques to visualize space-filling hermetic virtual models.
- Scientific contributions to HCI, in particular Visual Search tasks, with results of user studies comprising about 100 participants.

## 1.4 Manuscript Organization

The remainder of this manuscript is organized into six chapters.

A Literature Review discusses relevant materials science learning topics, available visualization software for educational purposes, prospective development platforms, a comprehensive systematic review about multiscale visualization systems, finishing with foundations of Human-computer interaction.

The Development chapter describes the design principles, main software elements, implementation process, and features of three interactive techniques.

The Evaluation chapter describes the design process of the main experiment, the necessary modifications made to the experimental versions of the three techniques, and adopted data analysis procedures, including statistical methods. The results and implications of two preliminary experiments are also presented in this chapter.

The Results chapter presents pre and post-tests conducted on experimental groups, an analysis of task order influence and a summary of results for each experimental task. Details are presented on Appendix A, including means, boxplots, pairwise comparisons of scores, and auxiliary analyses such as correlations between pertinent variables.

The Discussion chapter confronts findings with hypotheses, objectives and the most relevant literature.

## 2 Literature Review

### 2.1 Relevant topics for microstructure learning

Aside from the general applicability of this research for honeycomb structures, the case study presented here mainly concerns grain microstructures, and the goal is improving the learning of related materials science topics. However, there are numerous topics and evaluating learning improvements is a complex and time-consuming process. Still, it is possible to assess improvements on fundamental parameters to learning. Therefore, an elementary assumption of this study is that improving the cognition of relevant microstructural features will likely improve learning. Given that microstructural models are not widely available for educational applications either in physical or digital forms and considering inherent limitations of 2D images and diagrams, as reinforced by [Rodrigues et al. \(2017\)](#) and [Bardella, Rodrigues e Leal Neto \(2017\)](#), we can assume that students have difficulties visualizing the materiality and volumetric character of such structures and their features.

Spatial comprehension is a parameter commonly associated with success on efficient learning, especially on topics that rely on geometry. In fact, spatial abilities are associated with success in engineering, architecture, design, physics, chemistry, and biology ([BODNER; GUAY, 1997](#); [SOBOYEJO, 2002](#)). Fortunately, [Hornbogen \(1986\)](#) and [DeHoff \(2012\)](#) conceived microstructural formulations centered on geometry, relating material properties to the microstructural topology and the geometric character of features. Thus it is assumed that superior spatial comprehension of microstructural features will probably lead to better learning. Furthermore, since no accessible 3D microstructural model is available for educational applications, it is reasonable to start with the most straightforward volumetric features related to material properties, such as the grain structure. This assumption seems to apply to other honeycomb structures as well.

The grain is the fundamental segmentation of material space, defined by a phase shift or a change in crystallographic orientation. A monocrystalline material like feldspar or beryl has no grains since the crystal structure is continuous, i.e., a single-crystal is defined by the absence of grain boundaries. Usually, the characteristic crystal angles are visible on the surface of such natural materials. On the other hand, all other crystalline defects are likely to be present. Real perfect monocrystalline materials rarely occur in nature and must be produced under controlled conditions. Some specific applications require single-crystals, particularly in optics and electronics, where material properties derive solely from crystal properties, including the crystal structure and concentration, types, and spatial distributions of crystalline defects ([HERTZBERG; VINCI; HERTZBERG, 2012](#)).

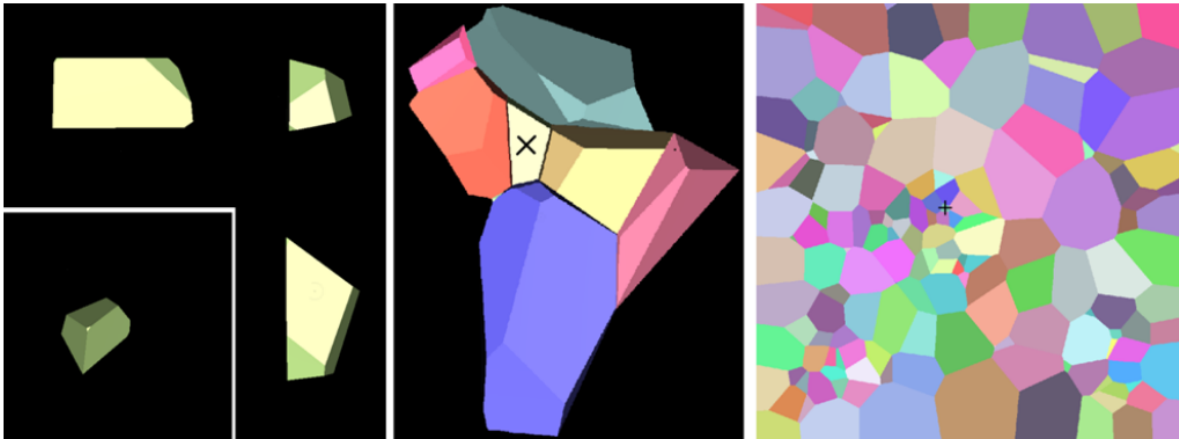


Figure 3 – Grain structure multiscale analytics: shape of a single grain (left), neighborhood of this grain (center), heterogeneity or clustering around the grain (right).

However, most engineering materials are polycrystalline, and relevant properties such as mechanical strength are related to distributions (grain size and shape) and other defects such as dislocations and impurities. A plausible interpretation of [Hornbogen \(1986\)](#) and [DeHoff \(2012\)](#) microstructural formulations ultimately implies a correlation of grain size distribution to material properties.

Industrial materials are usually processed until accepted homogeneity standards are achieved, including relevant crystalline defects and, most importantly, the average grain size. Firstly, a comparative visualization of grain size and shape between different materials would represent a rich educational experience. Secondly, the perception of grain size variations along a material sample is of particular interest, as they imply a localized divergence of properties. Intentional variations are enforced, for instance, in modern Functionally Graded Materials (FGM), where localized differences (such as electrical conductivity or resistance) are desirable for the specific application. Unintended variations, on the other hand, should be avoided, detected and studied. Such variations may result in weaknesses and catastrophic failures in structural materials ([HERTZBERG; VINCI; HERTZBERG, 2012](#); [ABBASCHIAN; REED-HILL, 2008](#)).

Property variations related to grain size occurs in regions concentrating large or small grains, where the grain size deviates from the sample's average. Therefore, considering the formulation of a grain structure visual analytics, a good starting point is the detection of heterogeneous regions. It is possible to conceptualize a scale progression, where a student captures characteristics of elementary features such as grains, their relation to neighbors and then proceeds to understand the impact of variations of those characteristics on materials properties (Figure 3).

The summary of [Hornbogen \(1986\)](#) and [DeHoff \(2012\)](#) microstructural formulations interpretation is that microstructures have an inherent space-filling 3D nature and most polycrystalline material's properties derive from fundamental geometric features, namely

grain size, shape, number of neighbors, local heterogeneities, extreme values, and distribution patterns. The relation to geometry may also hold for many relevant properties of other honeycomb structures. Therefore, it seems reasonable that a student should, in the first place, capture characteristics of elementary features such as grains, their relation to neighbors and then, following a scale progression, proceed to understand the impact of variations of those characteristics on materials properties (figure 3). It is also vital, as indicated by Lanzagorta et al. (1998), to offer educational applications user-friendly interfaces, further improving visualization experiences by engaging the user with aesthetics, visual design (RHYNE et al., 1999) and exciting interactions (BYRNE, 1996).

## 2.2 Human-computer interaction (HCI) foundations

HCI visual search theories are appropriate for informing the design and evaluation of the proposed visualization techniques, considering the desired learning improvements and the focus on the detection of specific features. This section will first introduce HCI fundamentals and then proceed to detail relevant aspects of visual search.

The main goal of Human-computer interaction (HCI) is to assess how humans use and interact with computer programs, focusing on interface and tools performance, as expressed by subjective and objective measures of task time and scores. HCI adopts the human information-processing approach, assuming that human behavior can be analyzed in terms of subsystems and their interrelation (PROCTOR; VU, 2009), i.e., a computer metaphor. The classic Atkinson-Shiffrin knowledge acquisition or cognitive model and the derived elaborations (Wickens, Rasmussen, Norman, Chapanis) are models of this ‘human computer’. Such models consider that the brain has limited processing power, unstable short-term memory, and a stable vast long-term memory that stores relevant events (episodic), language and meaning, beliefs, procedures, skills and habits (NUNES, 2014).

The available models of information processing or cognitive architectures (ACT, Soar, EPIC, MIDAS, MIACE) divide the brain into specialized modules or circuits: perceptual sensors (vision, hearing, tactile) with attention sensitivity; processing, reasoning, response selection or decision making units; hierarchical memory and response mechanisms. Almost all models assume that task execution relies heavily on limited working memory or buffer, disputed between perception and action activities. Only the EPIC model assumes independence of response resources or action circuits (Ocular, Vocal, Manual) not mediated by working memory and with dedicated specialized processors, allowing them to execute their actions in parallel. However, they will still have to dispute a single cognitive processor somewhere in the process.

Therefore, the human information-processing approach is based on the idea that

human performance is a function of several processing stages, a fundamental aspect of the computer metaphor (PROCTOR; VU, 2009). Consequently, all cognitive architectures (or information-processing models) have bottlenecks both in perception and decision-making. The limited amount of processing and memory resources, along with the perceptual and executive bottlenecks, are the theoretical basis for the cognitive load theory (CHANDLER; SWELLER, 1991), an essential concept in HCI.

Learning or performing new skills demands processing resources and rely solely on working or short memories, an onerous process that usually creates competition for mental resources and leads to cognitive overload (PROCTOR; VU, 2009). Complex tasks that call for different skills combinations or heavy parallel use of resources (working memory, attention, decision making, response capacity) are also likely to create overload. In short, either learning or performing complex skills in new contexts will probably lead to some cognitive overload. Ergo, as expected by common sense, the application of robust consolidated skills or skill combinations in known situations tends to be comfortable and faster - people deal better with known tasks and situations. The more it deviates from what is known, the slower and demanding the task will be, and negative impacts in performance follow, objectively expressed in time increments, score decrements, and subjectively in mental stress, typical measures in HCI user studies. Thus, the adopted information processing model correlates initial performance in a new task with inherent individual capabilities, previous specialized experience, specific or similar task experience, and task-induced cognitive overload.

The training specificity concept that applies to all human learning postulates that efficiency in specific tasks increases with experience time, i.e., performance progress from an initial cognitively demanding phase to automatic processing (PROCTOR; VU, 2009). In other words, the more experienced in a given task, the less energy will be spent. The power function proposed by Newell e Rosenbloom (1993) became widely accepted, establishing a relative learning rate that decreases hyperbolically as a function of practice trials and, consequently, the most significant benefits occur early.

Neural circuit is a relevant foundational theoretical neuroscience concept to back up the information-processing model and the derived cognitive architectures. This formulation considers the brain as several independent but connected specialized circuits. Each circuit has inherent biological capacities or potentials derived from the individual gene pool, trained by experience on specific tasks that require its unique capabilities, such as speaking, hearing, vision, and motor control (PURVES, 2012). However, a task usually asks for a combination of different circuits or has elements that compete for a single one. When a single circuit or resource is disputed by two or more tasks or diverse elements of a single task, overload occurs. However, the training specificity still applies and a complex skill can also be learned, trained and optimized with experience. In this sense, individual skills such



as specific hand movements are learned separately and can be performed automatically in parallel with other resources, i.e., each independent circuit learns and stores its own procedures. Examples are sports (strategy planning and multiple joint movements), music performance (reading scores, singing and playing), and driving, to name a few. Therefore, a complex task that uses several resources demands proficiency of the concerned circuits and effective coordination (task specificity). The overall process efficiency correlates with total mental effort.

Task loading is not the only predictor of performance – low loading is not intrinsically good or high loading intrinsically bad. Computer-based tasks that impose either too much or too little demand will likely be appraised as stressful (SZALMA; HANCOCK, 2009). A core factor here is motivation, i.e., whether that task or method is meaningful or compelling - any activity that demands cognitive resources and mental effort must have a compatible cost-beneficial value to the individual, which is a function of direct value (meaningfulness or usefulness) or apparent relation to personally relevant matters (curiosity). Motivation or engagement acts mainly as a driver to attention and modulator for the perception of cognitive workload – a motivated user will be willing to devote more resources and maintain high output for longer. The role of motivation is to set and maintain one or more meaningful goals during task execution, either intrinsic (personal achievement or satisfaction) or extrinsic (obligations). As long as the user has an active personal goal, attention will be directed towards the task and a high cognitive workload can be achieved and maintained, along with performance levels (SZALMA; HANCOCK, 2009). In the absence of a goal, high outputs will feel highly stressful and will not be maintained for long. Interestingly, some cognitive architectures assume that the user wants to perform the task while others include an attention factor. Attention can also reduce cognitive load by filtering irrelevant information. Finally, further experience on demanding tasks will reduce the perceived load and enhance the stamina under high workload, allowing a resilient level of performance (SZALMA; HANCOCK, 2009).

### 2.2.1 Spatial ability and navigational skills

Spatial ability is one of the most relevant skills in 3D applications, which can be divided into spatial orientation and spatial visualization (BODNER; GUAY, 1997). Specifically, spatial abilities indicate individuals' capacities in searching the visual field, capturing objects shapes and positions, forming mental representations of those features and manipulating them mentally (CARROLL, 1993). Relevant task experience is proficiency in navigating 3D spaces with abstract keyboard commands (eye-hand coordination) while performing visual search and analysis (spatial ability). This experience is usually found in 3D modelers and players of computer games - research suggests that gaming experience is associated with enhanced spatial skills, eye-hand coordination and decision-making

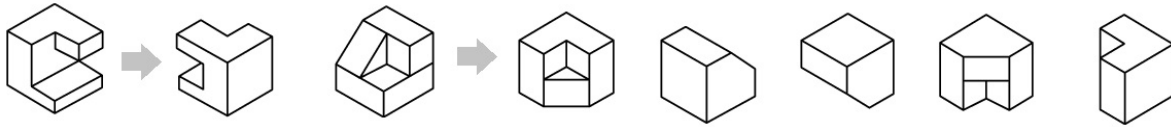


Figure 4 – Purdue spatial ability test. The first sequence shows a 180° rotation around the vertical axis. The same operation applied to the second figure results in the second image after the arrow. Adapted figure (BODNER; GUAY, 1997).

(HIRVASOJA, 2004). Spatial ability is usually assessed with mental rotation tests (Figure 4), questions about previous experience in drawing and modeling, urban navigation proficiency, and activities that require spatial awareness such as field sports. Task experience is assessed based on previous experience in similar tasks such as 3D modeling, digital 3D games such as first-person shooters and third-person multiplayer games, and experience with virtual reality applications or interactive systems. Those tasks require memorizing abstract navigation mappings and the subsequent transference to motor learning when the navigation becomes effortless and automatic.

Consider the example of experienced 3D first-person shooter (FPS) gamers. They should perform well on 3D analytics tasks, since modern 3D games have sophisticated and vibrant graphics, requiring a high level of visual recognition capabilities, spatial awareness, and navigation proficiency, all at blistering speeds. Multiplayer FPS or online role-playing games (MMORPG) also demands simultaneous speaking and strategy planning. That is why FPS or MMORPG gamers should be a pinnacle of performance in 3D interactive analytic tasks, as they have more training in individual relevant resources, high efficiency while combining skills and in alternating cognitive resources when they are disputed. One should also note that gamers are usually highly motivated to play and that their high skill levels probably will not be challenged by simpler 3D analytics tasks. Thus, the inclusion of some experienced 3D gamers in user studies is an interesting strategy to assess overall task difficulty, for they should be able to perform exceptionally well compared to non-gamers. However, tasks should either instigate their curiosity or be challenging enough, motivating such skilled participants to use their resources fully. If gamers can be benchmark references and are likely to perform well, at the other end of the spectrum lies unskilled and inexperienced participants with low spatial abilities. The performance difference between them should be substantial but not extreme; otherwise, non-gamers will mentally break when exposed to the tasks' difficulty level. If an experienced gamer cannot perform well on a given task, this means that the design has many flaws and should be reconsidered.

Thus, lower performance is expected as primary resources such as spatial ability and navigational motor skills are lacking. As users have less motivation, attention and training in those skills, task performance decreases and response time increases, accompanied by higher stress levels or cognitive loads. As suggested by the systematic review, multiscale

visualization interactive systems demand high levels of spatial awareness. Cognitive overload was also broadly reported, indicating that several resources are intensively used and disputed, which probably results from demanding requirements of visual analytics, spatial ability and navigation skills. The complexity of this scenario suggests that gamers or pilots (aircraft or drones) are expected to perform better, particularly when graduated in engineering and design courses, which require good spatial ability (BODNER; GUAY, 1997). Nevertheless, abstract navigation mechanisms should leverage existing abstractions or give way to more intuitive natural approaches.

One could suppose from this discussion that specific skills and their combination can be infinitely improved by training; however, this is limited by individual biological capacities such as availability of neural resources and plasticity (ability to shift or transfer resources). Therefore, significant individual variations should be expected, considering the combination of biological limits and training (PURVES, 2012). Several studies report gender or age-related variations of proficiency on specific skills, including spatial abilities, where, according to BODNER e GUAY (1997), there is a statistically significant difference in performance on spatial skill in favor of males.

In any case, as usually recommended by HCI user studies, navigation should be well trained and assessed before testing, along with proper survey or testing of previous experience, which is necessary to properly restrict the influence of such confounding factors when comparing task performance with different methods or tool alternatives.

### 2.2.2 Visual search theories

Signal detection tasks are typically adopted for studying human information processing, where the participant must detect a given signal within a sample, which is particularly useful when the task can be depicted in terms of binary discriminations (PROCTOR; VU, 2009). Aside from the numerous ways of measuring responses, by using the hit rate (signal present and detected) and false alarm rate (signal detected but not present), one can evaluate the effect of a variable on detectability or bias. In short, measures of bias reflect the participant's tendency to say that a signal is present regardless of the truth. The participant adopts some criteria which define how much evidence is needed to decide if a signal is present. Efficiency is assessed by measuring reaction time or accuracy as a function of the number of items in the display (WOLFE, 2000).

Technically, signal-detection theory assumes that every trial's response is a function of encoding and decision operations. The decision is whether the event is from the distribution of signal or the distribution of noise - both are assumed to be normal and of equal variance, which allows deriving a measure of detectability and response bias (PROCTOR; VU, 2009).

Visual search can be considered a type of signal detection task, where participants must distinguish a target amidst distractors. Target properties provide the signal, and the distractors properties provide the noise (WOLFE, 2000). Thus, the detection of grain size and heterogeneous regions is fundamentally a visual search task. When the target can be distinguished by a single feature such as color, the number of distractors present will have a negligible impact on error rate and reaction time measures, according to the Feature Integration Theory (TREISMAN; GELADE, 1980). However, in conjunctive search, both factors increase rapidly as more features must be combined to perform the distinction, for attention is required to scan the search field meticulously for a target or until reaching the end of the field. Although Feature Integration Theory inspired a large body of research (PROCTOR; VU, 2009), many cases could not be properly explained, and numerous variations were created to model and explain how visual attention works.

Nowadays, researchers agree that visual attention can be purposely guided. Preattentive processes gather objects prior to attention, segregating items from the background. When this filtering stage is absent, attention is deployed without guidance on all visible items, and performance will be strongly dependent on the set size. Items that can be detected by preattentive processing are noticeable. Between this eye-catching state and a complete absence of guidance, preattentive processing provides partial orientation. Preattentive features are properties processed prior to attention, such as size, scale, and topology. Some features are considered exceptional in their ability to group distractors or guide attention, such as motion, stereopsis and color. In conjunctive search, subjects must find targets by combining two or more of those basic features (PROCTOR; VU, 2009; WOLFE, 2000).

Therefore, finding effective ways to highlight features of interest is paramount for detection performance. The intention is to draw attention to the intrinsic characteristics of the objects. Color, in this study, must simply not interfere. Size is of particular interest, as search for bigger items is highly efficient (BILSKY; WOLFE, 1995), and the smaller are easier to find if compared to medium-sized targets (WOLFE, 2000). If the stimuli are of adequate size, then it should be possible to identify targets easily. On the other hand, if items need to be foveated to discriminate targets from distractors, performance will decrease substantially, and as foveal load increases, peripheral processing decreases in the form of attentional tunnel vision (IKEDA; TAKEUCHI, 1975; WILLIAMS, 1985; WILLIAMS, 1989). The hierarchical structure of multiscale models can also be effective in attracting attention (WOLFE, 1994). As for 3D viewing and rendering, stereopsis can support efficient search (NAKAYAMA; SILVERMAN, 1986), when targets and distractors differ in the tilt out of the frontal plane (HOLLIDAY; BRADDICK, 1991). Some experiments also indicate restriction of attention to surfaces defined by stereo disparity (HE; NAKAYAMA, 1994a; HE; NAKAYAMA, 1994b; HE; NAKAYAMA, 1992) and to locations cued in depth (ANDERSEN, 1990; DOWNING; PINKER, 1985). Thus, the use of immersive stereo and multiscale visualization seems very promising.



Figure 5 – Example of Multiscale Visualization. Extracted from (MCCRAE et al., 2009).

## 2.3 Multiscale visualization systems: a systematic review

Given the space-filling nature of a microstructure model and prospective preattentive features suggested by visual search theories, interactive multiscale visualization was deemed a promising approach. The term encompasses methods to deal with models that have details on several spatial or even temporal scales, exceeding eye resolution (JOY, 2009). A classic example is a digital map such as Google Maps, which reveals more details as we increase the zoom level (Figure 5).

However, the number of objects in multiscale models can easily exceed conventional visualization methods' capabilities, challenging traditional 3D modeling paradigms. In this sense, for some applications, the only viable approach to modeling is automation with procedural (SMELIK et al., 2014) or tiling techniques (DOŠKÁŘ; NOVÁK, 2016) or to avoid modeling at all by adopting alternative or statistic representations of groups of objects.

In the case of grain structures, two scales are of particular interest, the grain scale and the atomic scale (Figure 6). An automated process to model valid synthetic grain structures were devised by Rodrigues et al. (2017), focusing on achieving proper data structures for multiscale visualization in interactive engines. Modeling of atoms in a given crystal structure, including atomic defects, can be done procedurally, as geometric rules of such structures are known and clear. Conventional visualization would demand vast amounts of storage and graphics capacity. The solution is to control the computational load by balancing quantity and quality, based on the premise that it is senseless to display more objects than we can perceive, offering an automatic and intuitive filtering rationale. Conventional techniques are frustum culling (limiting what is visible), LOD (varying level of detail of objects), and scene partitioning schemes (on-demand loading of scene sectors) (JOY, 2009), commonly available in modern game engines (Figure 7).

Visualization and interaction in multiscale environments also pose cognitive challenges. Typical issues are navigation on crowded and occluded environments, localization within a vast world, acknowledgment of data in other scales, and scale transitions. Following the approach proposed by Kitchenham e Charters (2007), a thorough systematic review was planned and performed to survey the most relevant visualization and interaction challenges

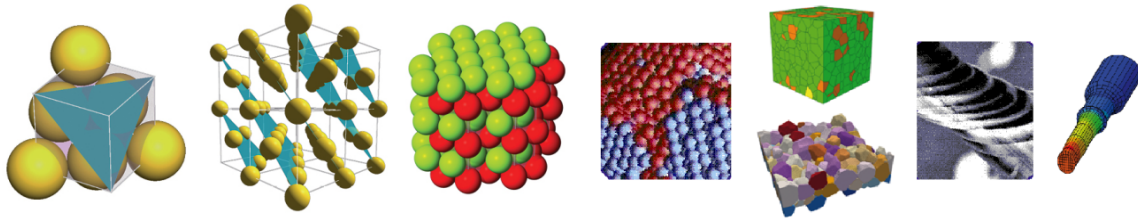


Figure 6 – Multiscale visualization in materials: atoms and crystal planes, plane stacking, grain boundaries, grains, macroscopic behavior.

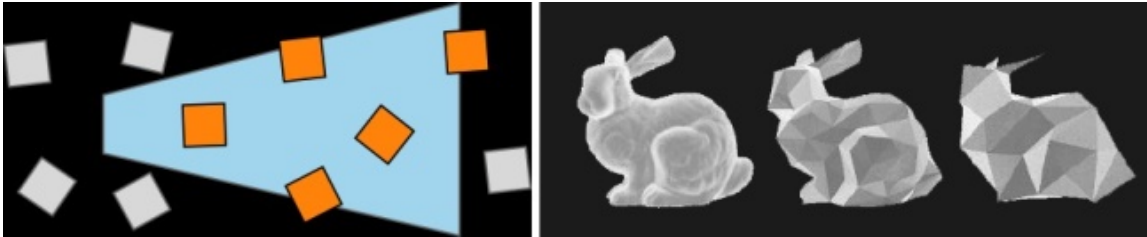


Figure 7 – Frustum culling (left) and Level of Detail (right) allows reducing graphic processing loads.

and available prospective solutions, along with cognitive foundations, in the pursuit of learning enhancements in such environments.

### 2.3.1 Review protocol

A protocol defines precisely the objectives, specific research questions, reference search strategies, selection criteria, and data extraction methods, allowing reproducibility and extension of the performed review. The main objective was identifying the main visualization challenges and implementation solutions in three-dimensional visualization systems.

Search privileged known and prestigious scientific databases for peer-reviewed articles in the English language, namely IEEE, ACM, ISI, Springer, Scirus, and Scopus. Only articles after 2000 were considered, as software research and development demands recent work.

#### 2.3.1.1 Research questions

- Primary question: what are the visualization challenges and solutions available for multi-scale three-dimensional systems ?
- Secondary question: what are the cognitive foundations ?
- Intervention: challenges and solutions.
- Control: review ([MCFARLANE et al., 2012](#)) on the subject.

- Population: research work that implements or evaluates three-dimensional multi-scale systems.
- Results: educational resources that may increase knowledge extraction from three-dimensional models at multiple scales.
- Intended application: visualization systems for complex material models (materials science), containing elements in two or more scales.

#### 2.3.1.2 Search strategy

- Sources
  - indexed electronic databases (IEEE, ACM, ISI and Springer);
  - electronic search engines (Scirus, Scopus);
  - existing reviews and seminal articles.
- Language: English, for being internationally accepted as the primary language for writing scientific papers. High-quality Brazilian papers are expected to be found in English.
- Keywords
  - terms for ‘multiple scales’: multiscale OR multi-scale OR multilevel OR multi-dimensional OR multiresolution;
  - terms for ‘systems’: system OR application OR software OR program OR (computer AND program) or (system AND software);
  - terms for ‘three dimensions’: 3D OR three-dimensional;
  - ‘visualization’ terms: visualization OR environments;
  - terms for ‘interaction’: interactive OR exploration OR navigation.
- Time-lapse: Restricted to papers published after year 2000 as software system references should be relatively recent.

#### 2.3.1.3 Study selection criteria and procedures

- Inclusion
  - peer-reviewed articles (journals or proceedings);
  - three-dimensional system containing elements in two or more scales or that addresses applicable techniques in three dimensions;
  - optional: includes evaluations or discussions on cognitive issues.

- Exclusion: Articles before the year 2000.
- Selection procedure: Search strings were created by combining synonyms of the chosen keywords. These strings were tested and submitted to related search engines. The relevance of a paper will be assessed by applying the inclusion criteria, reading the title, abstract and introduction, as needed.

#### 2.3.1.4 Results extraction and compilation strategy

A full reading of selected works was carried out by filling a form for each article, containing the following: basic information (title, authors, publication date), software and hardware platforms adopted, application domain, number of scales present, visualization challenges, implemented solutions, any underlying cognitive concepts and evaluation results. Data synthesis was discursive (list) and graphical (frequencies). All the challenges encountered were organized in classes.

### 2.3.2 Execution

The review process took two months in late 2014, and a one-week update revision was executed in late 2019 to check for any breakthroughs or relevant novelties. After entering the chosen keywords in all considered databases, applying inclusion and exclusion criteria, removing duplicates and reading abstracts and introductions as needed, 51 articles were selected from 380 and read in full. Individual forms were filled for each article, containing basic information, software and hardware platforms used, application domain, number of scales present, main challenges found and proposed solutions, evaluation results, and cognitive concepts involved (if present).

### 2.3.3 Results

#### 2.3.3.1 Articles statistics

Figure 8 (top) shows that most articles were relatively recent when the review was conducted. From 32 articles found in the update revision, only four were considered relevant; however, they did not bring new elements to the original analysis. Two articles from Miao et al. are worth noting as they are similar to this work, the first (MIAO et al., 2019) consisting of a literature review, while the second shows the application of multiscale methods in the domain of molecular modeling and visualization (MIAO et al., 2018). A closer analysis of their work revealed some of the challenges found in this review. Therefore, as the present review is more comprehensive, including issues, challenges and solutions found by Miao et al. (2019), the original systematic review was considered up to date considering the specificities of this work.



Most systems adopted custom desktop platforms (Figure 8, center) developed in Open GL and C++ (Figure 8, bottom), focusing on generic, biomedical and data visualization applications (Figure 9, top) with 3D geometry and sometimes bundled information objects (Figure 9, center), displaying 2, 3 or multiple scales (Figure 9, bottom). About half the articles did not mention Level of Detail schemes and from the rest, half mentioned or clearly adopted (Figure 10, top right). The remaining articles adopted a similar rationale with some form of alternative representation. Finally, 41% conducted some sort of user testing, although only half of those adopted formal methods (Figure 10, top left).

### 2.3.3.2 Issues and solutions

Issues were divided into eight categories, namely, navigation, structure, cognitive overload, representation, interaction, precision and integration, customization and selection (Figure 10, bottom). The first five categories were further subdivided and are detailed below.

### 2.3.3.3 Navigation

Concerning navigation (Figure 11), the user must first be able to situate himself in the world (context) and capture its structure (spatial structuring) without getting lost (disorientation). Then, one should be able to find the desired information (wayfinding) with fluidity (navigability) and finally, move between the scale levels. Navigation challenges were the most frequent in the articles (31 %).

- Spatial awareness or context: global maps and landmarks (MCFARLANE et al., 2012; KOPPER et al., 2006; LI; FU; HANSON, 2006; DENNIS; HEALEY, 2002; ZHANG, 2009; ZHANG, 2005);
- Capturing the spatial structure: scale, fixed and general references (TRINDADE; RAPOSO, 2014; VICECONTI et al., 2011; GLUECK et al., 2009);
- Wayfinding (being able to find things): assisted navigation, hierarchical trees (PASEWALDT; TRAPP; DÖLLNER, 2011; KOPPER et al., 2006; VICECONTI et al., 2011; BACIM; BOWMAN; PINHO, 2009; GUIARD et al., 2004; HILDEBRANDT; TIMM, 2014);
- Navigation fluidity: spatial references, obstacle assistance, assisted speed adjustments (BOSCA; BONINO, 2006; KOPPER et al., 2006; MCCRAE et al., 2009; ZHANG, 2008; ZHANG, 2009; GODDARD; HUANG; FERRIN, 2005);
- Scale transitions: slow animations (SESTER; BRENNER, 2009; MCFARLANE et al., 2008; TAGKOPOULOS, 2011; KOPPER et al., 2006; BACIM; BOWMAN; PINHO, 2009; ZHANG, 2009);

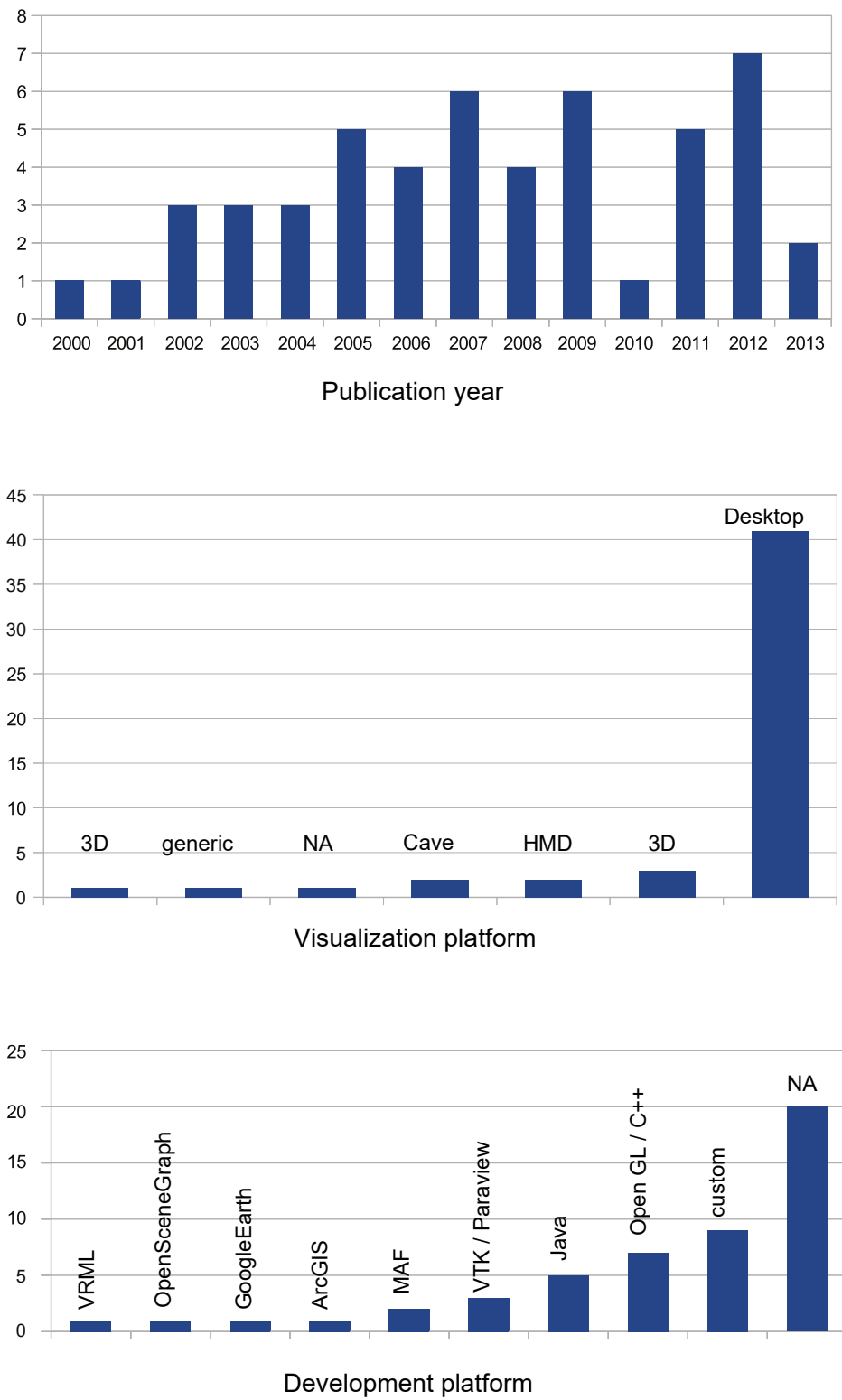


Figure 8 – Articles statistics.

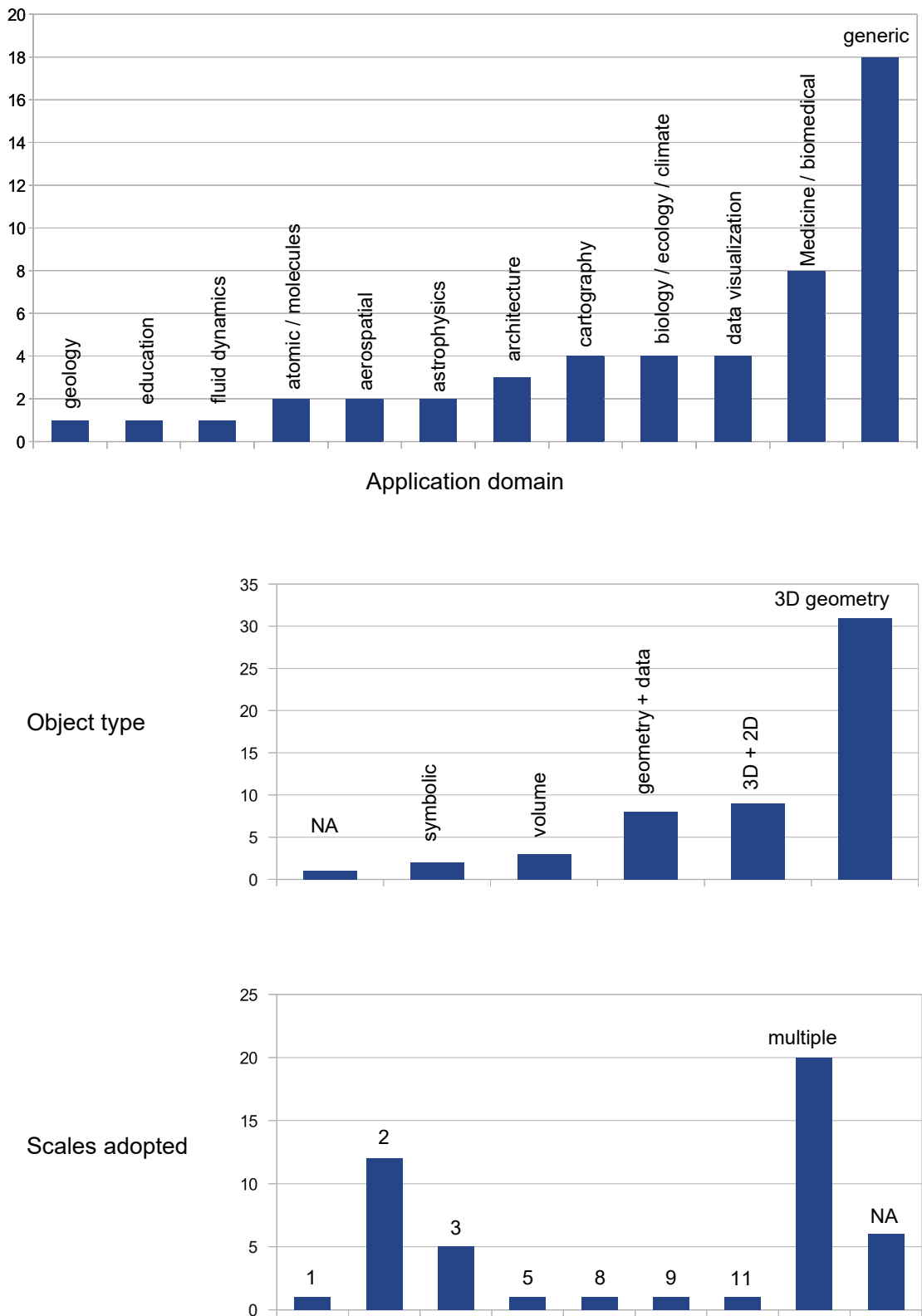
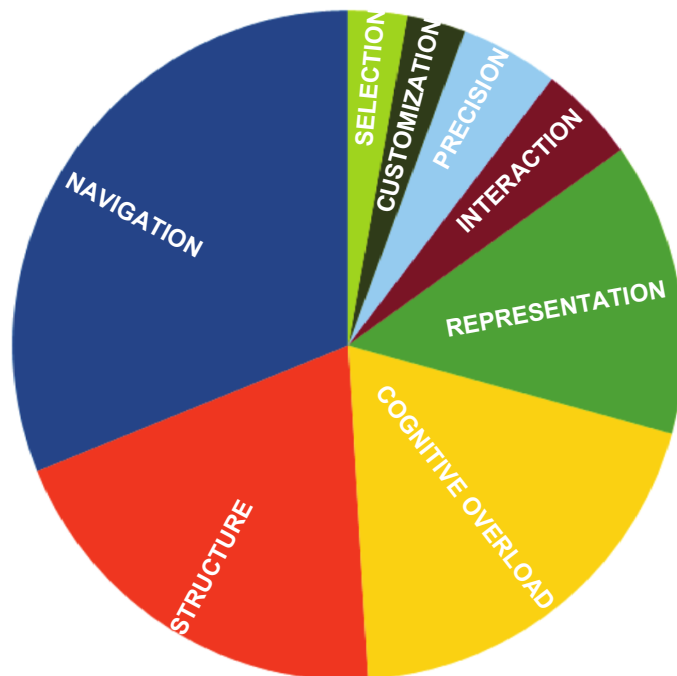
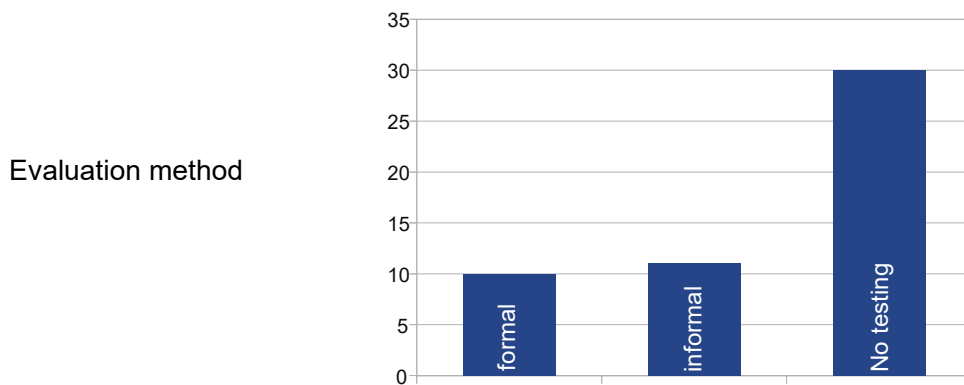
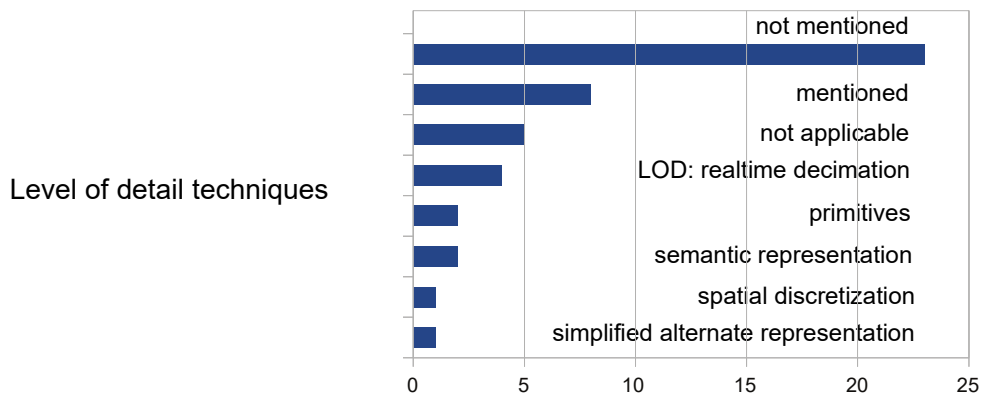


Figure 9 – Articles statistics (cont.)



Main issues found

Figure 10 – Articles statistics (cont.)

- Disorientation (MCCRAE et al., 2009; TRINDADE; RAPOSO, 2014; MCCRAE et al., 2010a; KOPPER et al., 2006; HILDEBRANDT; TIMM, 2014; GLUECK et al., 2009).

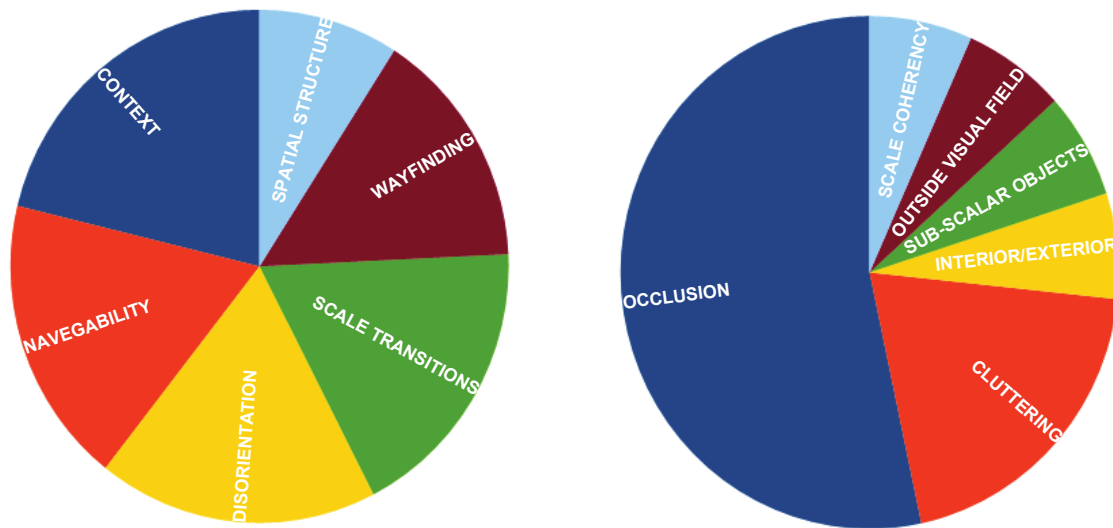


Figure 11 – Navigation and representation issues.

#### 2.3.3.4 Structure

As for structure issues (Figure 11), in order to extract knowledge, the user must be able to build mental schemes that represent relationships between objects. They occupied the second place, with 20 % of frequency.

- Relationships among entities (differences and similarities): graphic or symbolic connections, guided animations, hierarchical trees (KREUSELER, 2000a; LAW, 2002; HENRY; POLYS, 2010; INSLEY; GRINBERG; PAPKA, 2011; LUBOSCHIK et al., 2012; ZHANG, 2005; ZHANG, 2008);
- Pattern recognition: guided exploration (HENRY; POLYS, 2010; GRIEBEL et al., 2004);
- Spatial comprehension: landmark structure or networks (SHARMA et al., 2003; CORDEIL et al., 2013; ZHANG, 2008; CHANG et al., 2007; BOWMAN et al., 2003; GLUECK et al., 2009);
- Spatial continuity: robust and preserved topology and spatial relationships (especially while zooming), speed control (HSU; MA; CORREA, 2011; PINDAT et al., 2012; KOPPER et al., 2006; LI; FU; HANSON, 2006; COCKBURN; KARLSON; BEDERSON, 2009);

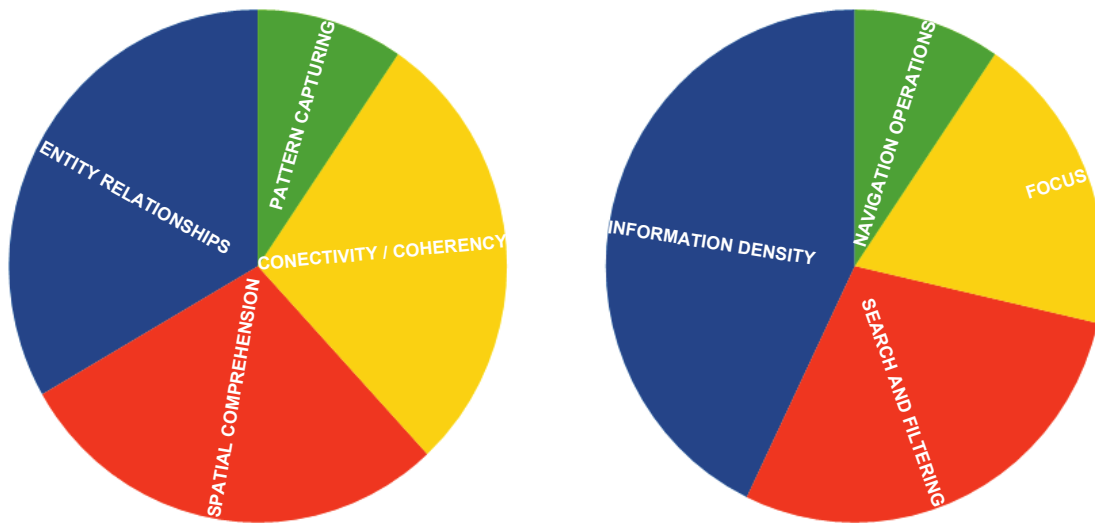


Figure 12 – Structure and cognitive overload issues.

### 2.3.3.5 Cognitive overload

The cognitive overload problem (Figure 12) is central to multiscale systems, as the brain has limited resources to process information. They tied for second place, also with 20 % frequency.

- Item or object crowding: level of detail techniques, objects of interest, similar object grouping (BOSCA; BONINO, 2006; MCFARLANE et al., 2012; CORDEIL et al., 2013; PIETRIGA; APPERT; BEAUDOUIN-LAFON, 2007; BACIM; BOWMAN; PINHO, 2009; VANACKEN; GROSSMAN; CONINX, 2007; LAVIOLA et al., 2001);
- Navigation on complex structures: gestural navigation or direct manipulation, assisted navigation (HOUTGAST; PFEIFFER; POST, 2005; HILDEBRANDT; TIMM, 2014);
- Search for specific information: filters, grouping, hierarchical trees (STOLTE; TANG; HANRAHAN, 2003; GYULASSY et al., 2009; LUBOSCHIK et al., 2012; ZHANG et al., 2012; PIETRIGA; APPERT; BEAUDOUIN-LAFON, 2007);
- Focus: global and detail viewports (ZHANG et al., 2012; ZHANG; FURNAS, 2005; GUIARD et al., 2004; GLUECK et al., 2009).

### 2.3.3.6 Representation

Representation (Figure 12) should allow ample access to the information in a visually consistent way. They came in fourth, with 14 % attendance.

- Interior/exterior distinctions (GLUECK et al., 2009);

- Subscale data exhibition and distinction: annotations, lenses, alternative symbolic representations (icons) (MCFARLANE et al., 2012);
- Relevant information out of the visual reach: graphic indicators (MCCRAE et al., 2010a);
- Scale coherency (ZHANG, 2008);
- Cluttering: animations, preset viewpoints, focus, alternative symbolic representations (icons) (MCFARLANE et al., 2008; MCCRAE et al., 2010a; ARTERO; OLIVEIRA, 2004);
- Occlusion: visualizing hidden or internal objects (PASEWALDT; TRAPP; DÖLLNER, 2011; BRYSON, 2005; MCFARLANE et al., 2008; TRINDADE; RAPOSO, 2014; MCCRAE et al., 2010a; BOWMAN et al., 2003) (ARTERO; OLIVEIRA, 2004).

#### 2.3.3.7 Interaction

Despite being apparently relevant, interaction problems appeared only with 4.7 % frequency:

- Interaction method should change according to scale (VALIATI; PIMENTA; FREITAS, 2006);
- Selection in high occlusion and cluttering situations: filtering and hierarchical trees;
- Stereoscopy and natural manipulation techniques has a limited range for large or complex scenes/models (HOUTGAST; PFEIFFER; POST, 2005).

#### 2.3.3.8 Cognitive issues

Cognitive theories and studies support most of the proposed solutions. Cognitive overload is a typical issue in multiscale environments, demanding filtering, visual coding of information, and a clear hierarchy. Navigation and orientation problems were also extensively investigated by cognitive science, demonstrating that references, structures and spatial landmarks are paramount for proper mental map building although it is not enough - navigational experience or proficiency is equally important. Besides, the brain leverages several environmental tips usually neglected in virtual scenes. Finally, the spatial context is also created hierarchically as the environment is continuously explored, by combining maps in different scales. As for interaction, virtual systems should greatly benefit from natural manipulation techniques and dimensional coherency. Concerning pattern recognition and knowledge building, structure is the key concept. Highly structured visual reference systems and techniques that allow preserving visual and structural continuity between all relevant elements (animations, for example) should enhance learning.

## 2.4 Current approaches and tools for microstructure visualization

Although very restricted in scope, [Lanzagorta et al. \(1998\)](#) is the only initiative to date with clear intention to develop specific interactive visualization tools to improve understanding of microstructural characteristics. This early experiment with 3D visualization demonstrated benefits of realtime interaction, interactive 3D and stereoscopic visualization, restating the limitations of static visualization approaches. This attempt revealed development opportunities and challenges of interactive visualization of microstructures, such as lack of integration of computational solutions and difficulties for end-users to operate visualization systems and interact with models. However, the interaction aspects of this pioneering research were not further developed. Subsequent authors used visualization to satisfy specific research needs rather than developing specific interactive techniques for visual analysis of grain structures.

There are several relatively accessible tools capable of building microstructural models from user-defined parameters (Figure 13), such as free software Dream 3D ([GROEBER; JACKSON, 2014](#)) and Neper ([QUEY; DAWSON; BARBE, 2011](#)), or proprietary options such as Simpleware, Micress and Digimat. However, those applications are research-oriented, still lack integrated domain-specific visual analysis and interaction capabilities, and the user interface is not friendly enough for educational use. Even if one masters using such complex software, the algorithms do not scale well to larger samples (more than 500 grains) or higher voxel resolutions, resulting in small models and poor visual quality. Thus, the mentioned programs are inappropriate for educational applications.

Another relevant aspect is that the resultant models' data structure is not adequate for use in 3D interactive software. A 3D object is usually represented by their boundaries (B-Rep) in most CAD applications. This type of data structure is also expected in powerful interactive development engines (or game engines) such as Unity, Unreal, Cryengine, and web-based graphics libraries such as WebGL and X3D. Given this structure, interactive visualizations are simple to create with the resources available on such systems: rendering configurations for every object (colors, image textures, and transparency), ready-to-use first or third-person controls, collision detection, object clicking and selection, raycasting and customizable user interfaces. Therefore, given the rationale of game engines, models of grain structures should be structured so that each grain is a single polyhedral object, allowing a high degree of control.

Another explicit limitation of existing software is the character of interaction. Despite the visualization capabilities of software such as Graphite (Figure 13 - right) to perform realtime rotation of the cutting plane, this feature is limited to rotation around a central pivot and is clearly intended to select the angle of the static section ([Alice - Geometry and Light, 2020](#)). Users will have a hard time using this feature to inspect a portion located far away from the object center (pivot point).



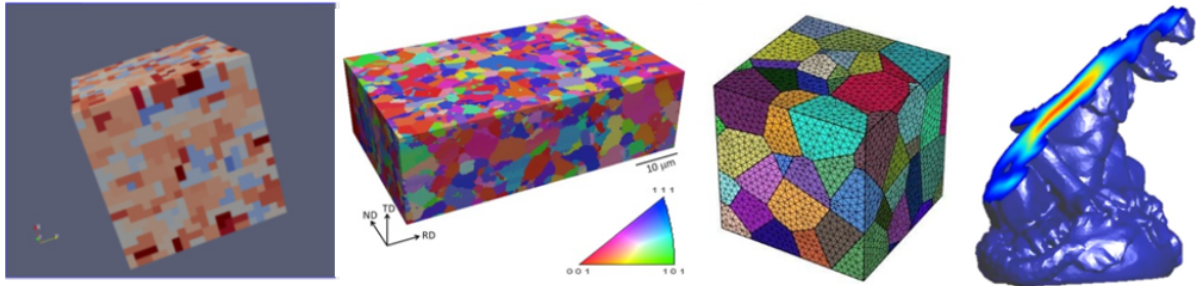


Figure 13 – Grain structure models: Dream 3d, low and high resolution models (left), extracted from (GROEBER; JACKSON, 2014). Tessellated model generated by Neper (center right), extracted from Quey, Dawson e Barbe (2011). Static section visualization on Graphite (right), screenshot extracted by the author.

Therefore, simple, open-source, automatic and integrated tools must be created to empower teachers, students, 3D content creators, and even researchers, providing them with meaningful visual analysis and interaction capabilities to allow advancement of teaching methodologies and comprehension of complex subjects such as grain structures. Models produced within 3D content creation CAD applications or game engines are easily captured and transformed by artists and ready to use within such applications or interchanged between them.

## 2.5 Available development platforms for 3D interactive software

As of late 2015, when system development started, the literature (PATTRASITIDECHA, 2014; WAERNÉR, 2012; PETRIDIS et al., 2012) highlighted powerful and flexible desktop 3D development platforms: Unreal/UDK, Blender, Unity 3D, CryEngine, Source and Ogre3D. Despite the possible graphical capabilities and development productivity features, standard desktop applications are not the most accessible or portable options. They demand purchase, downloads, and installation in a desktop computer or a laptop that fits the minimum graphics performance requirements. In this sense, 3D web technologies are better alternatives for this project, considering the relevance of accessibility and portability requirements.

The main available options for 3D web development at the time (late 2015) were JOGL, LWJGL, WebGL and Stage 3D. However, in 2011 the Java platform was already compromised by several incompatibilities and instabilities, and all derived technologies went into decay (Java 3D, JOGL e LWJGL). After a technological evaluation, WebGL was elected as the best alternative, despite the risks related to the maturity of the technology at that time. In 2015 WebGL was confirmed as an HTML5 standard by W3C, reinforcing the choice as the project's platform, and nowadays is currently supported by most web browsers. Chrome, Firefox, Opera, and Safari are notoriously more robust regarding HTML5 support as Internet Explorer 11 and Edge runs WebGL only in an experimental mode. The advent

of HTML5 and Web GL may also diminish obstacles to achieve genuine interactive web applications for mobile devices due to better use of native hardware graphic capabilities. Therefore it is paramount to invest in optimized techniques to allow application access even in middle-tier mobile devices. In this sense, cloud services can be explored to alleviate local hardware resource use (BREIVOLD et al., 2014).

WebGL was based initially on OpenGL ES 2.0 and built upon the Canvas element of HTML5, providing a 3D rendering graphics API accessed through Javascript language (PARISI, 2012). The Canvas provides several methods to build 2D or 3D graphics using only native browser resources. Although WebGL is a low-level programming language, several libraries were already available in 2015, such as SceneJS, GLGE, Copperlicht, CubicVR 3D, X3DOM and the highly popular ThreeJS, expediting development with high-level programming (BARBOSA, 2013). SceneJS never advanced from the beta phase. GLGE supported only basic OpenGL functionalities, missing physics capabilities, and shader classes. CubicVR 3D lacked advanced effects, requiring shader programming. CopperLicht was well documented and allowed animations, collisions, math modules, lighting options, and optimized binary mesh storage. X3DOM was an experimental framework after the X3D specification for web content, a clear advantage for developers familiarized with this standard, although some functionalities still lacked maturity. The web page detailed the API reasonably and several tutorials and demonstrations were available.

After a comprehensive analysis, the ThreeJS library was adopted for this project, being the most popular and proven API for WebGL development. The main element is the renderer, that offers a standard scene with a main camera as a start. Scene objects can be created by code from available primitive geometry or imported in several standard mesh formats such as OBJ, COLLADA, and STL. Although other libraries existed or were launched after 2016, such as BabylonJS and Blend4Web, the choice was based on feature availability, code structure, performance, robustness and, more importantly, on documentation, adoption and community support. Nowadays, BabylonJS is a reliable option for WebGL development; however, the project has currently only 524 followers and 1884 forks on GitHub (09/2019), more than all other options but still way below ThreeJS's numbers back in 2016. Today, ThreeJS has 2454 followers and an impressive 20700 forks (a three-fold increase since the beginning of 2016), showing that it is a developer pick for WebGL development. Consequently, all basic bugs were tackled, several improvements and optimizations were made, and developer support is widely available across online forums hosted in sites such as StackOverflow.

Therefore, the WebGL library evaluation performed at the beginning of development still holds, reinforcing the choice for ThreeJS.

## 2.6 Literature review summary

The goal of this study is to improve the learning of related materials science topics. A fundamental assumption is that improving the spatial comprehension of relevant microstructural features will likely improve learning. It is reasonable to start with features that are most related to material properties, such as the grain structure. Microstructural formulations correlate grain size distribution to material properties. Thus, the perception of variations of grain size along a material sample is of particular interest, as they imply a localized divergence of properties.

HCI visual search theories are appropriate to inform both the design and evaluation of the proposed visualization techniques, considering the desired learning improvements, and the focus on the detection of specific features. In visual search, participants must distinguish a target amidst distractors, which can be a demanding task. Cognitive architectures of the human information-processing approach expose bottlenecks both in perception and decision making, the theoretical basis for the cognitive load. Assuming the limited amount of processing resources, either learning or performing complex skills in new contexts is likely to lead to cognitive overload. Attention can reduce cognitive load by filtering irrelevant information. Thus, finding effective ways to highlight features of interest is paramount for detection performance. Preattentive processes gather objects prior to attention, and some features were considered promising in the context of this study for their ability to group distractors or guide attention, namely size, depth cues, stereopsis, hierarchical structure, and color. Still, primary resources such as spatial ability and navigational motor skills are necessary for minimal performance. Motivation also plays a role, acting as a driver to attention and modulator for the perception of cognitive workload.

Given the space-filling nature of the microstructure model and prospective preattentive features, interactive multiscale visualization was deemed a promising approach. A thorough systematic review was performed to identify visualization challenges and available solutions, along with cognitive foundations. About 380 articles were found and 51 selected. From eight categories, four were considered more relevant: navigation, structure, cognitive overload, and representation. Concerning navigation, the user must first be able to situate within the virtual world and capture its structure. Then, one should be able to find the desired information with fluidity and transition between scale levels smoothly. As for structure issues, the user must be able to build mental schemes that represent relationships between objects, in order to extract knowledge. Given the brain's limited resources to process information, cognitive overload is a central problem to multiscale systems. Representations should allow access to the information in a visually consistent way. In summary, solutions can be broadly categorized into two groups: information filtering or highlighting in a cluttered visually dense environment; coherent hierarchical spatial structure with visual landmarks.

However, existing visualization software is heavily research-oriented, still lack integrated domain-specific visual analysis and interaction capabilities, does not support multiscale models, and the user interface is not friendly enough for educational use. Data structures of models are also not adequate for use in 3D interactive software. In fact, no initiative up to date intended to develop specific interactive visualization tools to improve understanding of microstructural characteristics. 3D web technology was preferred over desktop applications for system development, prioritizing accessibility and portability criteria. Given the Java platform decay, WebGL was elected and nowadays is currently supported by most web browsers. Several libraries are available to expedite the development, and ThreeJS was chosen based on feature availability, code structure, performance, robustness and, more importantly, on documentation, adoption, and strong community support.

## 3 Development

### 3.1 System design: features in a multiscale paradigm

In an ideal scenario, an interactive visualization system for microstructural models should feature multiple scales and include at least the atomic and grain scales. The systematic review provided the following best practices in this scenario, along with detailed implementations for each solution (examples in Figure 14):

- Assistance and guidance using animations, navigation, speed control;
- Global/local maps, viewports and zoom windows or lenses;
- Hierarchical trees, filtering and analytics;
- Alternative graphic/visual representations;
- Additional information on demand (pop ups);
- Simultaneous exhibition of objects in different scales;
- Natural visualization and interaction, and adoption of environmental clues.

The most critical issues are occlusion and cluttering due to the space-filling character of honeycomb structures, a problem faced by biomedical applications as well. Typical solutions are filtering around regions or objects of interest, by removing objects or performing sections, for instance. Thus, one should be able to select and hide grains individually or in structured groups, by isolating a grain and its neighbors, for instance (Figure 14, top left). Another solution is transparency and wireframe objects (Figure 14, middle left).

Grains are less diversified and distinctive than biological structures, demanding extensive referencing landmark structures to help the user locating itself and mapping the sample (Figure 4). Typically, microstructural samples present precise boundaries, adopting geometries such as plates or blocks, which allows for basic spatial awareness, building the ‘surroundings’ of the sample. Furthermore, a ‘ground’ element can introduce a top/bottom reference, and identifying the front face completes the situation map as a reference for the remaining sides (back, left and right). The user would always see the ground reference even in sectioning operations. An orientation cube can be provided to reinforce this situation map (Figure 14, top center, left), following the direction of the sample in real-time. The cube can be promoted to a global map by locating the user within the sample with a 3D crosshair. Other references can be introduced in strategic regions of the sample, such as

the center. For that, specific grains can be colored or identified, and a graphic grid can divide the cubic sample in octants. Finally, as a mnemonic aid, the user could be provided with annotation tools to mark grains or features of interest (Figure 14, top center, right).

Good spatial or volumetric perception is paramount for proper identification of shape and size of features, as well as their topological relationships. Adoption of perspective, stereoscopic vision and environmental cues such as directional lighting and realistic shadows improves depth perception (Figure 14, top right). In this sense, an ideal sectioning method should properly depict lighting behavior in the remaining portion of sectioned grains, dimming light proportionately as a function of grain depth and shape.

Within the grain scale, an important feature derived from the atomic scale is the crystallographic orientation, which can be depicted by a regular lattice, resulting in stair-stepped grain boundaries (Figure 14, middle left). Grids (the lattice) are commonly used in illustrations to depict typical crystalline defects. A seamless scale transition can be applied when the user wants to inspect a specific defect within a given grain's crystal structure. Atoms might gradually appear in grid positions as the user approaches the crystal lattice. All representation changes should have tailored thresholds to avoid estrangement and confusion, where automatic speed changes may help to diminish such effects.

Additional information can be provided for individual grain characteristics (shape factor, volume, neighbors) along with visual analytics for each entry. Bounding boxes, spheres or ellipsoids can depict the shape factor, neighbors can be isolated, and defects might appear as colored dots, showing concentration gradients. Similar grains can be highlighted per statistic type upon user request. One instance or all additional information might appear along with its representation as the user chooses a given grain to inspect or as symbols/icons in aggregate views (Figure 14, middle center), allowing comparisons among different grains. The grain's geometry may assume a semi-transparent rendering whenever the user is visualizing information (Figure 14, middle right).

Several articles in the systematic review indicate assisted navigation mechanisms and animations as solutions to some cognitive issues of multiscale models. However, one should note that proper assistance mechanisms are domain-specific and build on established knowledge about the model's nature and structure. Still, the geometrically-based formulations by [Hornbogen \(1986\)](#) and [DeHoff \(2012\)](#) favors some mechanisms as a way to assist visualizing a grain structure model: navigation assistance mechanisms such as speed controls when entering the sample in sectioning mode and in scale transitions; navigation restriction along grain boundaries (Figure 14, bottom) and in lattice directions; automatic speed control whenever a change is close ahead (edges, triple and quadruple points in boundaries, defects in crystal lattice) and in representation transitions that might occur in zooming operations or at user requests for individual information on grains and defects; zoom sensitive orbital navigation to inspect individual grains and neighbors as

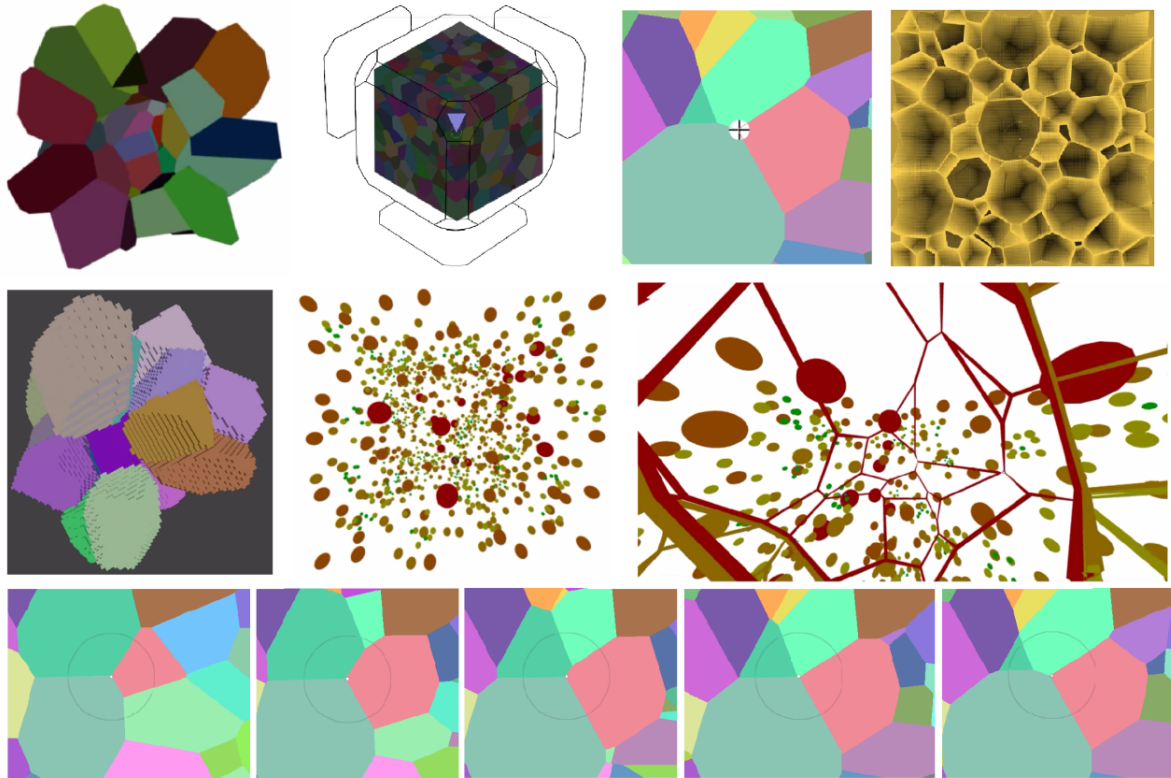


Figure 14 – Prototyping visualization methods. Top: filtering geometry by grain isolation; orientation cube used in some CAD software; marking features of interest in space with icons; rendering method to improve depth perception. Center: communicating crystallographic orientation with ‘voxelized’ boundaries; alternative geometric representations and visual codes allows sample-wide analytics. Bottom: image sequence exemplifying guided navigation along a grain edge (highlighted by the circle).

well as heterogeneous regions; natural interactions and stereoscopic vision with virtual reality technology, diminishing navigation and manipulation abstractions (HOUTGAST; PFEIFFER; POST, 2005; GLUECK et al., 2009). As possibilities are vast and the field is open, probably the most valuable assistance approach in an educational perspective is leveraging expert knowledge on an individual basis, offering tools for instructors or teachers to build multiple meaningful sequences to support explaining microstructural concepts. This approach merges custom expert-based assistance with animation and will be further discussed.

Finally, one or more scales might appear together using viewports or visual distortion rendering techniques, showing simultaneously the explored region, graphic representations of relevant grain statistics, neighborhood relationships, and the crystal orientation. A useful application of this technique would be showing a smaller scale detail of a crystal defect inside the lattice of a grain, for instance. This technique could help to build deeper insights by way of better cognitive connections, resulting probably in superior analytic capabilities and enhanced learning.

## 3.2 Target audience

Lecturing experience in materials science and engineering classes in undergraduate and graduate courses revealed limitations of didactic materials support to the comprehension of fundamental microstructure topics, given the inherent limitations of traditional bi-dimensional representations of grain structures and its defects (RODRIGUES *et al.*, 2017). Literature suggests that the persistence of erroneous interpretations of such fundamental topics among students will negatively impact comprehension of material's relationships between mechanical properties, structure, and processing (KRAUSE *et al.*, 2003) and possibly lead to adverse consequences to their future professional practices, particularly in engineering classes (BODNER; GUAY, 1997; SOBOYEJO, 2002). Although this work will focus on a target audience of engineering students, other areas where materials science is relevant in production or academic environments will be included as well, especially physics and chemistry.

The proposed methods should be relevant to the academic and professional performance of the target audience, given the potentials of interactive microstructural digital models to improve cognition of features such as grain size and shape, dislocation density, volumetric fraction, spatial distribution of phases and microcracks. On the other hand, since engineering students are expected to have good spatial abilities (HUK, 2006; ALIAS, 2000a) and given the preference of Sao Paulo University students for active, sensorial, and visual learning styles (KURI; SILVA; PEREIRA, 2006), it is hypothesized that the proposed methods should perform well for the target audience, as discussed on section 2.2.

## 3.3 System features

As described in section 1.2, this work is restricted to propose novel visualization methods to improve the spatial perception of honeycomb structures for learning purposes. However, the ultimate goal is to obtain instructive interaction methods for complex microstructures, comprising typical phases and defects. As argued previously and supported by the systematic review, the multiscale paradigm is a promising approach to enhance the learning of microstructures, given the nature of such models. Since the grain structure (individual cells) is considered the most fundamental feature of microstructures and no specific interaction method exists yet, the scope of this work was restricted to experimental methods to interact specifically with grains, i.e, a single scale. Despite having only the most relevant scale at this moment, the platform must be planned considering the desired characteristics described previously. Therefore, identified challenges and proposed solutions for the grain scale will be considered in the current system specification. Finally, the focus on the individual cell makes the results more relevant to the analysis of simpler honeycomb structures.



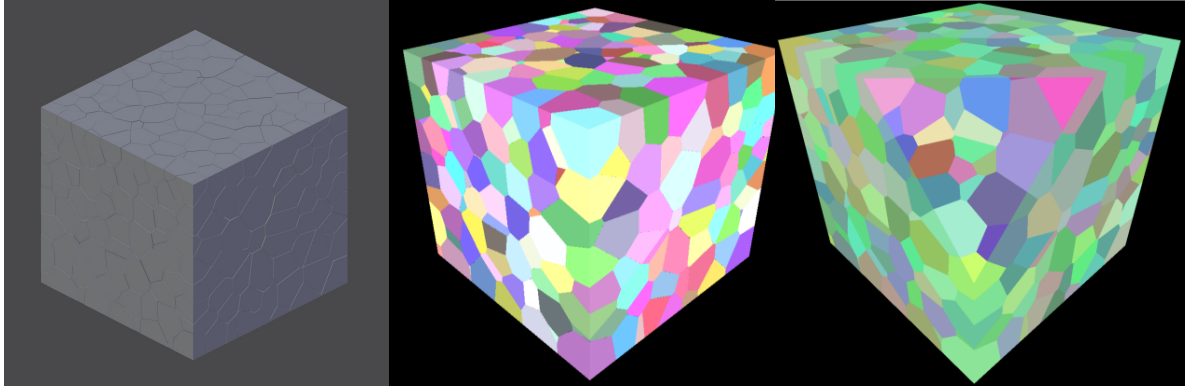


Figure 15 – Left: model generated by [Rodrigues et al. \(2018\)](#). Center/right: each grain of the imported model is assigned a random clear color.

### 3.3.1 System implementation description

### 3.3.2 Basic capabilities

As already described, the ThreeJS library ([RICARDO CABELLO, 2020](#)) was adopted in this project for WebGL development. The system was implemented as an HTML file and a JavaScript module, which currently adds up to more than 2500 lines of code. The HTML file calls dependent libraries such as Three JS and some specialized modules (3D stereo modes, shaders, geometry loaders, interactive controls) along with known productive web development libraries such as JQuery and Bootstrap (menu handling).

The grain models used in this work were synthesized according to the procedure described in ([RODRIGUES et al., 2017](#)) and exported in the COLLADA format as a collection of independent objects, each with its own center and self-contained geometry (Figure 15, left). Before exporting, each grain's geometry was optimized to reduce the total model size and increase rendering performance, reducing polygons with a decimation algorithm. Currently, the system is capable of loading the popular OBJ 3D file format and COLLADA, used due to better support for structured object assemblies (preserving individual centers and positions) and its known animation storage capabilities. Upon loading, a random color is applied to each grain (Figure 15, center and right).

After loading the model, module 'FlyControls.js' implements navigation, providing six DOF keyboard controls (Figure 16), while the 'DeviceOrientationController.js' allows using mobile device sensors and touchscreen gestures. The mobile control navigation mode uses gyroscope sensors for rotation operations and touchscreen for translations and focus (zoom) with pinch gestures, leveraging both natural and known interactions (Figure 17). A peculiar camera configuration (see next section) and a mixed Fly/Orbital navigation were adopted to achieve the desired interactive sectioning method.

Available rendering modes allow balancing image quality and rendering performance and introducing relevant depth enhancing effects. 2D modes are flat render, 2.5D and

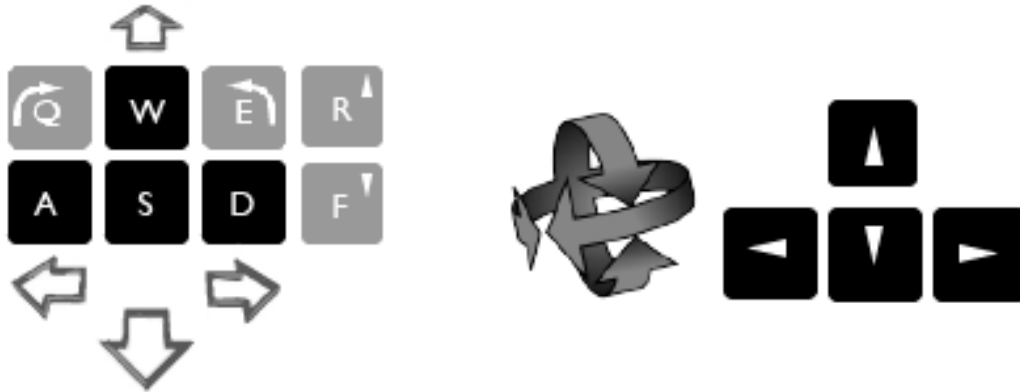


Figure 16 – Typical keyboard mapping for 6DOF navigation: WASD keys for translations, arrows for rotations.



Figure 17 – Mobile navigation. Typical touch operations for translations (top row): initial position, moving sideways, up and down; Intuitive movements for rotations (bottom row): initial position (center), rotating left around the green grain in the corner (left) and right around the same corner (right).

Ambient Occlusion (Figure 18), while 3D is available as Anaglyph or Split Stereo (Figure 18, bottom right). Rendering modes rely basically on typical shaders and post-processing effects available as ThreeJS classes. The conventional flat-shaded 2D section mode relies on the economic ‘MeshBasicMaterial’ and the 2.5D mode on the known and efficient opaque Lambert shader, both allowing high framerates even on mobile devices. 2.5D depth enhanced modes rely on the Ambient Occlusion (AO) post-effect offered as a ThreeJS class (‘SSAOShader’). The effect level and quality can be adjusted according to hardware capabilities and to allow either interactive framerates or extreme high-quality rendering for screenshots or slow interaction. Three quality modes were configured: low-quality focus mid-range mobile devices or low-end computers; normal quality fits most computers and high-end mobiles; the high-quality mode targets high-performance workstation computers or conventional computers for screenshots only (Figure 19).

3D is available through ‘StereoEffect’ and ‘AnaglyphEffect’ ThreeJS classes, offering Split Screen and Anaglyph stereo modes. Split-screen is a side-by-side stereoscopic mode

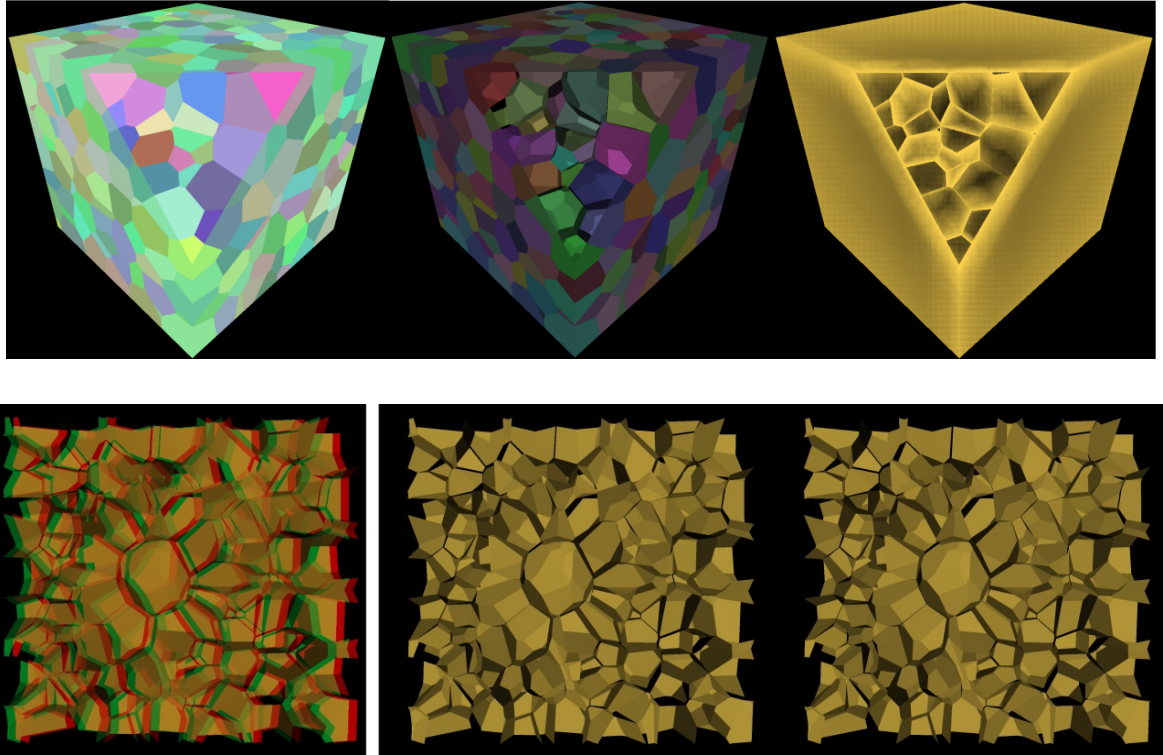


Figure 18 – Rendering modes. Top row: flat 2D, 2.5D and depth enhanced 2.5D. Bottom row: 3D anaglyph and 3D split screen.

that works out-of-the-box on conventional 3D displays and mobile devices equipped with VR adapters. Specialized HMDs such as Oculus Rift, HTC Vive and GearVR are currently not supported, although the split-screen mode is technically compatible. The anaglyph 3D is accessible and should work on any modern screen with the adoption of good quality red-blue anaglyph glasses. However, it is a known fact that the anaglyph technique is not universally comfortable and that some users cannot perceive the 3D effect, mostly due to ghosting issues and heterogeneous filtering quality across the color spectra (SANFTMANN; WEISKOPF, 2011). In order to deal with such limitations, a yellow sepia tone was configured after several iterations in different screens, achieving a good depth sense even when mild ghosting is present. All AO rendering levels are available in both 3D modes. The flat rendering is not available in 3D because the image is always 2D in this mode. Only anaglyph and standard split-screen 3D modes run on mobile devices. All AO 3D modes demand powerful computers to achieve interactive framerates. However, low-quality 3D AO might run in interactive framerates on conventional computers and high-end smartphones.

Object picking is possible with mouse clicks or by touch. ThreeJS offers raycasting functions for picking, allowing detecting the closest object to the clicked point. Since the camera near plane coincides with the screen space, given known viewport sizes and the camera position in the scene, it is straightforward to obtain the click in the 3D world with a few vector operations. The closest grain is then found and isolated by casting a ray in

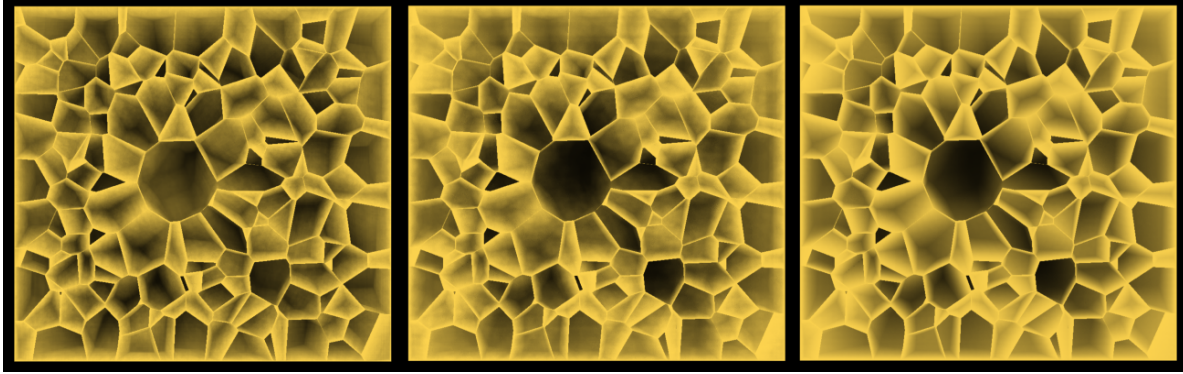


Figure 19 – Depth enhanced 2.5D in three quality levels, from low to high.

any direction (Figure 21, center left). Picking operations are required in this project's sequence, for user studies and the implementation of advanced features. An annotation method was created in order to allow users to identify selected grains or clicked points by placing a crosshair icon (Figure 21, left). This was accomplished by adopting ThreeJS' Sprite objects, which behaves as a 2D icon in 3D space, maintaining parallel orientation in relation to the camera and, if desired, constant visual size.

Another requirement for user studies is robust data recording. For every click, the system captures elapsed time, the point in scene space, the clicked grain, the amount of movement since the last click, a snapshot (camera position vector and rotation quaternion), visibility status of bigger and smaller grains including exposure time, additional support information such as task type and user ID. A ThreeJS Clock object provides time measurement. An online Google form was created to simplify data recording, avoiding the use and onerous configuration process of server databases. Although the form is focused on user input, it also accepts POST requests. Thus, a JQuery Ajax function was used to automate data storage in the form, using a POST method.

Figure 20 summarizes system implementation. The following examples are possible readings of this diagram: ThreeJS MeshBasic Material class allows 2D Flat graphic output using the Renderer class; ColladaLoader class creates a grain model with several colored grain objects while the Raycaster class is used in Neighbor filtering function to find neighbors and alter the visibility of the grain model; Customized FlyControls and DeviceOrientationControls classes directly control a camera object, moving the attached scene Camera using user input; A POST function of the Ajax class, provided by JQuery library, stores data collected from multiple scene objects in a Google Form spreadsheet.

### 3.3.3 Advanced capabilities

All methods described here are either needed for the sequence of the project or represents desirable evolutions of current ideas. Therefore, they were worth implementing in this first moment in a simplified fashion, not deviating from the primary objective.

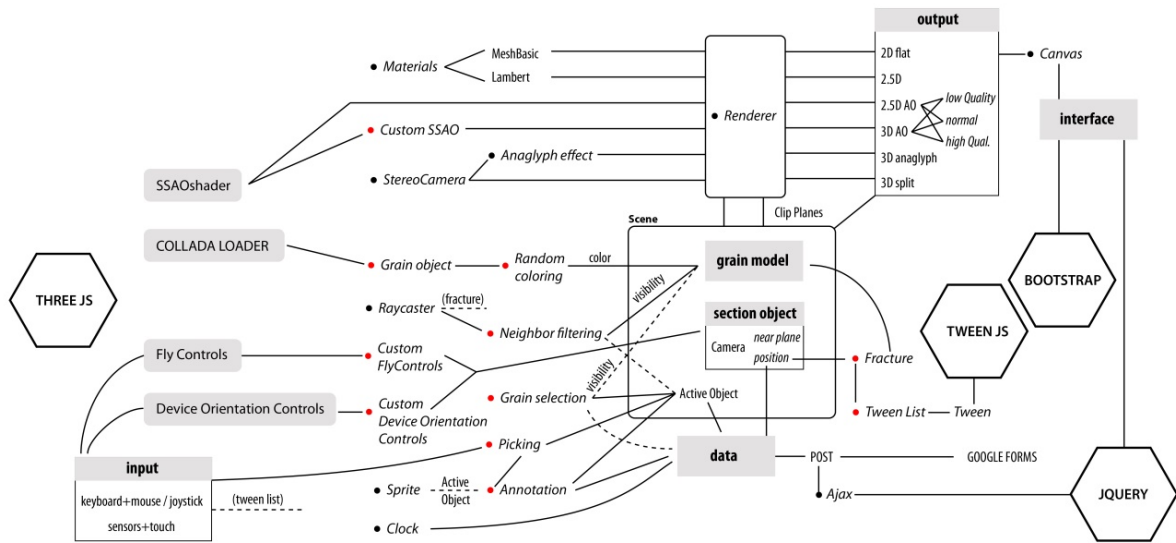


Figure 20 – System implementation diagram.

Grain neighbor detection is currently performed using a brute force raycasting approach since neighbor counts (and therefore grain face counts) are relatively constant (from 8 to 20), not representing a computational challenge. This method uses the chosen grain’s position and shoots rays to every vertex of the polyhedron, returning the first hit for each ray, resulting in an average of 100 rays. After removing duplicate hits, a list of neighbors is obtained. The selected grain and its neighbors are then isolated by hiding the rest of the sample (Figure 21, middle right). Raycasting is available in ThreeJS as a ‘Raycaster’ class and is performed with simple vector operations.

On section techniques, ‘Plane fracture’ filtering is a method to hide all intersected grains. The method shoots a finite number of parallel rays covering all the height of the section plane (camera near plane). The ray spacing is calculated as a third of the smaller side (bounding box) of the smaller grain of the sample, assuring the detection of all grains intersected by the sectioning plane. This results from 500 to 1000 rays per section, which is accomplished in less than a second on conventional computers. The performance is satisfactory since this method is not intended to be realtime. Each ray returns all hit grains, which are hidden to produce the fragile fracture effect (Figure 21, right).

Another useful filtering function is volume clipping, a typical computer graphics method that allows defining several infinite planes in the scene, hiding all geometry behind them. This allows defining arbitrary sample shapes, volumetric sectioning, and simultaneous multi-angle visualization of several section planes (Figure 22). This method is usually realtime, directly performed in the GPU. ThreeJS Renderer class has a handy ‘clippingPlanes’ array, where the user informs all desired clipping planes (as ‘THREE.Plane’ plane objects) to be used by the renderer. Interactive volume clipping is one prospective evolution of the interactive sectioning method described in the next sections.

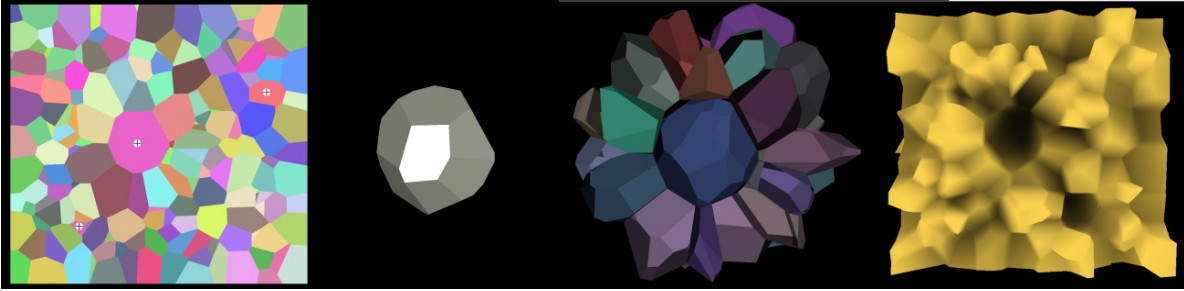


Figure 21 – Advanced operations: marking (left) and picking (center left) individual grains; Isolating neighbors (center right); Fracture (right).

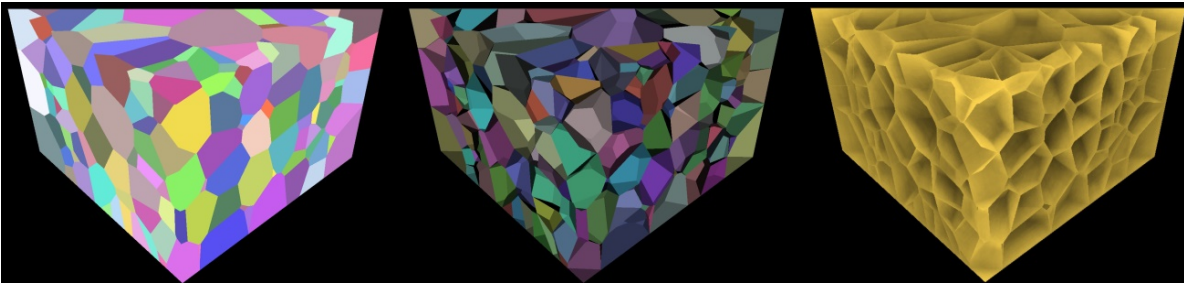


Figure 22 – Advanced operations: volume clipping. Three infinite planes filter geometry and define a new volume from the cubic sample.

Functions for procedural animations are offered by the TweenJS library, which interpolates values in time (or per frame in computer graphics). It is currently possible to interpolate camera position (vector) and rotation (quaternion) from one point to another, chaining how many points the user wants. The transition from one point to the other was called a ‘take’, inspired by cinematographic techniques. Each take has a user-defined time parameter and an interpolation function. Take chaining is automatically handled, and four interpolation functions are available (linear, circular, quadratic, cubic). The library also allows calling any Javascript function in the beginning/end of every take, which permits ‘animating’ or changing any system parameter, such as rendering mode and all the other operations already described (volume clipping, fracturing, selecting and isolating selected grains and neighbors, sending user data, loading models, etc). Take creation requires the user to navigate to the desired position, typing the take duration in a numeric menu box, choosing an interpolation function and adding the take to the animation sequence. Optionally, the tool offers an ‘active object’ parameter, allowing calling any function that requires a selected grain, as well as a ‘function’ menu field, which allows defining an existing function to be executed at the end of the take.

### 3.3.4 Interface design

As previously discussed in the introduction session, available digital tools for microstructure (grain structure) synthesis and visualization are not user friendly or adequate for educational uses. In his seminal work, Carroll ([CARROLL et al., 1998](#)) stands for a mini-



Figure 23 – Adaptive interface. Desktop version (left): note the minimal menu along the page bottom. Mobile version (right): note the collapsible menu.

malist design that exposes the user to system’s elements in a controlled manner, revealing functionalities according to the context, indicating that user interface can be a barrier to education software, which further justifies investments on logical and coherent visual design, aligned with contemporary aesthetic tendencies. The proposed interface approach was to combine contemporary aesthetics and minimalist design aiming to leverage user knowledge in such interfaces, minimizing estrangement.

This reflected on a very simple adaptive design privileging the visualization area and keeping menu items to a minimum. Leveraging Bootstrap’s capabilities, the system layout (web page) adapts to different devices and screen sizes, adjusting the visualization area aspect ratio and collapsing menus accordingly (Figure 23). The horizontal menu is positioned below the visualization area and is organized in three sectors. The first (left) sector contains elements to select rendering modes and quality. The central sector offers a few commands for narrative creation (section 3.5.3.1), and the last (right sector) has some handy user options such as fullscreen mode and a selector for mobile or desktop controls.

## 3.4 Realtime interactive sectioning technique

This project’s primary goal is to offer an improved learning experience over traditional educational media such as images, diagrams and animations or available visualization software, in order to tackle typical student misconceptions. The method of choice for inspecting microstructures is sectioning, performed with real or digital discrete sections.

Although the original video sequencing technique introduced by Lanzagorta *et al.* (1998) attacked the main limitation of static sections, it was not further investigated in the literature. Interactive sectioning in first person, when compared to videos or animations, allows more freedom and exploring the entire sample from any angle. However, in a first analysis, sections are computationally intensive procedures, limiting interactivity. Volumetric models sectioning requires surface rendering methods while B-Rep models

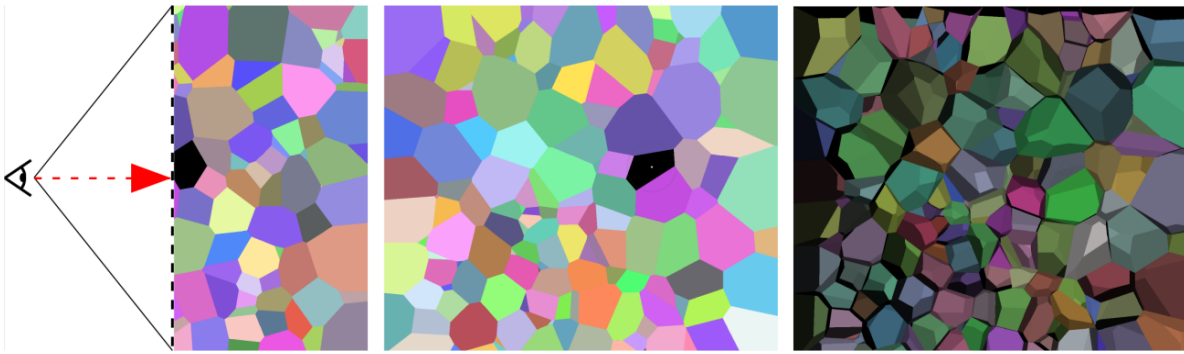


Figure 24 – Realtime sectioning implementation. Side view (left): camera near plane (dashed line) coincides with the sectioning plane. The camera is attached to the hinge point (red arrow) and is controlled by navigation commands. Flat 2D rendering (center) creates the illusion of a solid section by ignoring the shading component. Standard shading (right) : the user can see through the sectioning plane.

Boolean operations. This explains why most 3D software executes sections in discrete steps. Graphite, for instance, allows the user to interactively select the section plane angle, although the feature is limited to rotation around a central pivot (see figure 13).

The systematic review also indicated sectioning as a viable filtering solution in high occluded and cluttered environments. Therefore, implementing an interactive sectioning method in the chosen WebGL platform was considered the most natural starting point. However, as the standard rendering behavior allows the user to see the remaining portion of the grains, behind the section plane (Figure 24, right), a rendering method was developed to achieve a solid representation of the section. Instead of executing computationally intensive Boolean operations between the section plane and all sectioned objects, a simple technique was conceived to obtain sections in realtime, allowing first-person interactive flying navigation (RODRIGUES et al., 2017). The camera near plane was set as the section plane. Backface culling was disabled to make the interior side visible, and then rendering were configured to discard the shade component so all faces reflected the pure RGB color of the grain, creating the illusion of a solid section (Figure 24, middle). In order to visually distinguish the section plane from the rest of the sample, a transparent colored casing was created (Figure 25).

Navigation was implemented in a mixed fashion, fusing first-person (flying) with orbital navigation. A circle in the middle of the visualization area indicates the rotation pivot so the user can rotate around a known point, allowing interactive inspection of grain shape. The camera's near plane always coincides with this pivot, and since the field of view is fixed, the camera-pivot distance defines the zoom level (Figure 24, right). When the user moves, input from keyboard commands or sensor information actually translates and rotates this pivot and the camera only follows the movement, maintaining the relative position to this point.



Suggested environmental cues (directional lighting and shadows) for depth perception improvement were implemented using ambient occlusion techniques, simulating light behavior as if the model was really sectioned - the ‘holes’ are darker according to the depth of the grain in that direction. The effect responds properly to zoom as well. Ambient occlusion was the only available alternative to obtain this effect in realtime. Three quality levels are available and can be chosen according to the capabilities of the device (Figure 19). The lower level was designed to work even in middle-tier smartphones and tablets.

Filtering can also be performed by choosing individual grains and isolating its neighbors, using planar projections of mouse clicks, plain distance calculations to find the selected grain, and brute-force ray casting to find neighbors. Another filtering technique is the ‘planar fracture’, inspired in the fragile fracture phenomena known in materials science. The technique removes all grains intersected by the view plane, allowing the user to visualize the outer surfaces of the remaining visible grains. Impressive visuals akin to actual micrographs of fragile fractures can be obtained by combining this technique with ambient occlusion rendering (Figure 21, right).

The only spatial reference features implemented so far, considering the extensive list in section 3.1, are the cubic sample perimeter with its colored cage, fixed grain coloring and a tool to place 2D markers in space. This omission was intentional, in order to assess user spatial awareness in preliminary evaluations.

All atomic scale features and scale transitions were not considered since the scope was restricted to the grain scale. However, several proposed features can be readily implemented, such as grain statistics representations (shape factor, volume or size, and neighbors), grain grouping by statistic similarity, and transparent centroid representations, to name a few. Some of those possibilities will be explored in the second version (described in the next section). Other features such as lattice or crystallographic orientation representations and crystal defects (along with density and distributions) depend on more sophisticated modeling.

## 3.5 GrainCrawler technique

The GrainCrawler tool was based, as a whole, on prospective features of multiscale models and associated visualization and interaction techniques; proven pedagogical methods such as constructivism; motivation techniques such as gamification; as well as on feedback from preliminary user studies conducted on the interactive sectioning technique. One should note that the use of sections is also compatible with multiscale methods, so the qualities of different visualization mechanisms can be combined. However, the system presented here is still restricted to the grain scale.

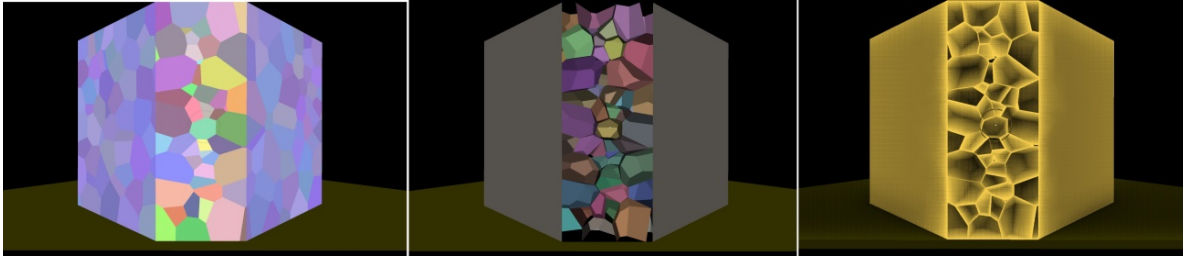


Figure 25 – Ground reference. Flat shading (left). Standard 2.5D (center). The ground looks more realistic with the shading effect created by ambient occlusion rendering in 2.5D depth enhanced mode (right).

### 3.5.1 Spatial reference features

Feedback from the preliminary user study indicates that the purposely missing spatial references (see section 4.4.2.5) would greatly improve task performance, especially for users with lower spatial ability. Thus, four features were introduced:

- ‘ground’ reference (Figure 25);
- a multipurpose cube, serving as a situation map and global map. The cube follows the visualization direction and may also serve as an orientation tool, allowing the user to quickly orient the sectioning plane by clicking in one of the six standard faces or, in the case of a beveled cube, any angle formed by two or three faces (Figure 26);
- section plane and relative user position in the cubemap, indicated by a crosshair (Figure 26);
- division of the cube map in visible octant sectors (1/8 of the cube – Figure 26, right);
- identification of sample faces (front, back, left, right, bottom, top - Figure 27).

Figure 28 illustrate the appearance of a tentative screen space with all the proposed spatial reference features combined.

### 3.5.2 Multiscale features

GrainCrawler technique includes prospective features of multiscale models. Despite the absence of multiples scales, the current model has a relatively large number of grains (more than 1000), and three orders of spatial magnitude were considered – sample, neighbors, and single grains. This means that features of multiscale models are highly relevant for grain structures alone, such as filtering, grouping, alternative graphical representations, level of detail techniques as well as visual representations of statistics (shape factor, volume or size, neighbors), that allow assessing sample or long-range parameters spatially, i.e., in the sample scale. On the other hand, sectioning techniques (S and DS) allow inspecting only a

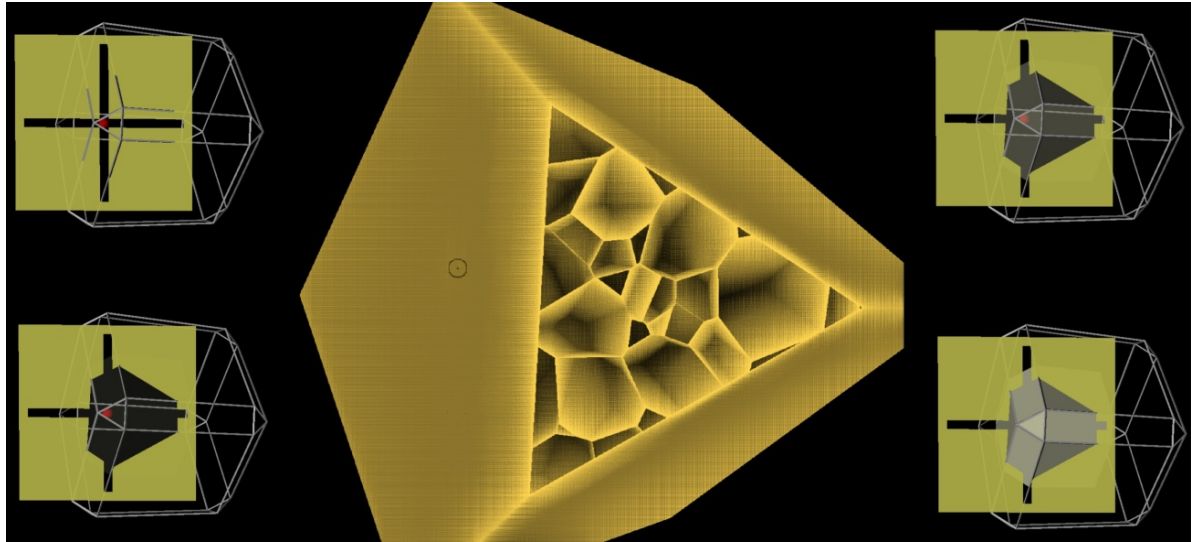


Figure 26 – ‘Cubemap’ design possibilities. Crosshair section plane, and beveled cube (top left). Section perimeter highlight (bottom left). Transparent octant (top right). Opaque octant (bottom right).

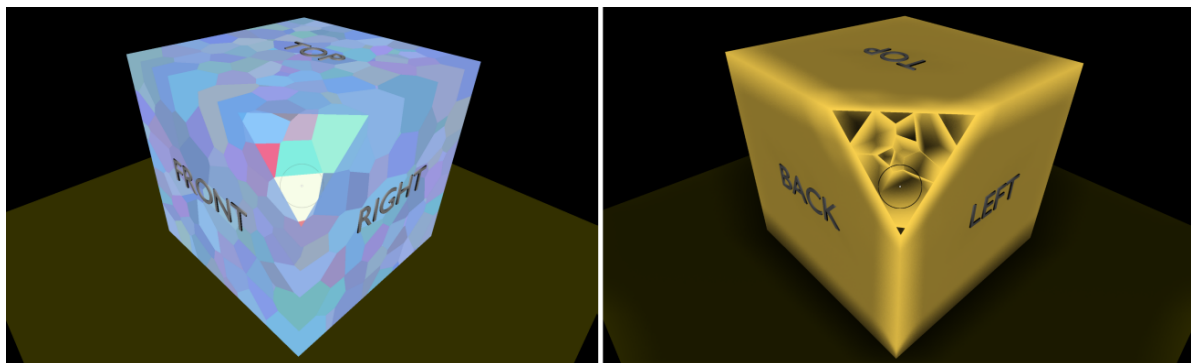


Figure 27 – Face identifications in flat and 2.5D depth enhanced modes.

single plane at any moment, and sample-wide characteristics depend solely on the user’s mental map. Therefore the new features of GrainCrawler will probably improve spatial comprehension of microstructural properties related to the sample scale and possibly the neighbor scale compared to the interactive sectioning technique.

The following configuration is proposed, considering the multiscale potentials identified for grain structures (section 3.1) :

- Distant sample-wide view: iconic representations for farther grains allow detecting heterogeneous regions for detailed inspection (background of Figure 29, left). The assumption is that spherical icons allow detecting scattering patterns;
- Focus view – neighborhood inspection: the view’s rotation pivot move to the center of the clicked grain, and the camera slowly moves to this new position. Representations change to wireframe and smaller icons as a function of proximity. This configuration may help to find the central grain of the region (Figure 29, middle);

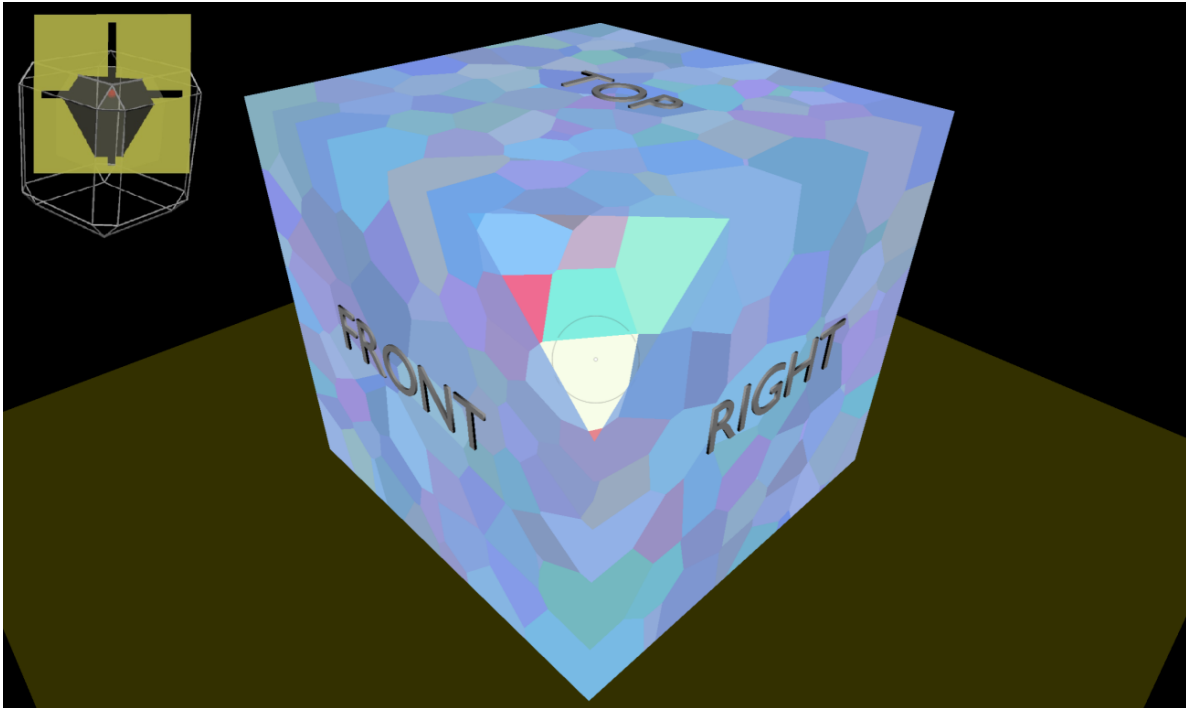


Figure 28 – Tentative configuration of the visualization area.

- Standard navigation view: farther grains are represented as solid objects and closer ones as wireframe cages, in a scene with progressive fog. This combination of features act as a visual filter, to facilitate detecting smaller and bigger grains (Figure 29);
- Proximal view – grain of interest and surroundings: the selected grain assumes a solid representation and neighbors get transparent to facilitate individual shape evaluation (Figure 29, right).

### 3.5.3 Prospective features

#### 3.5.3.1 Guided animations

Animation or features related to speed navigation appeared with a surprisingly high frequency in the systematic review (see section 2.3.3). In fact, animations are heavily studied and taken in high esteem in the educational literature, especially when it offers user control and involves active student participation.

An in-depth analysis of selected articles of the systematic review revealed that guided animations are not pointless movies or random sequences but intentionally designed steps that demand an established knowledge of the model, resembling cinematographic techniques. This opens endless possibilities for a teacher to suggest deep relationships among spatial elements or for a student to explore 3D models and create persistent records. In other words, the intelligence on microstructural models can be provided by teachers of material science classes, informing prospective guided or assisted animations. The

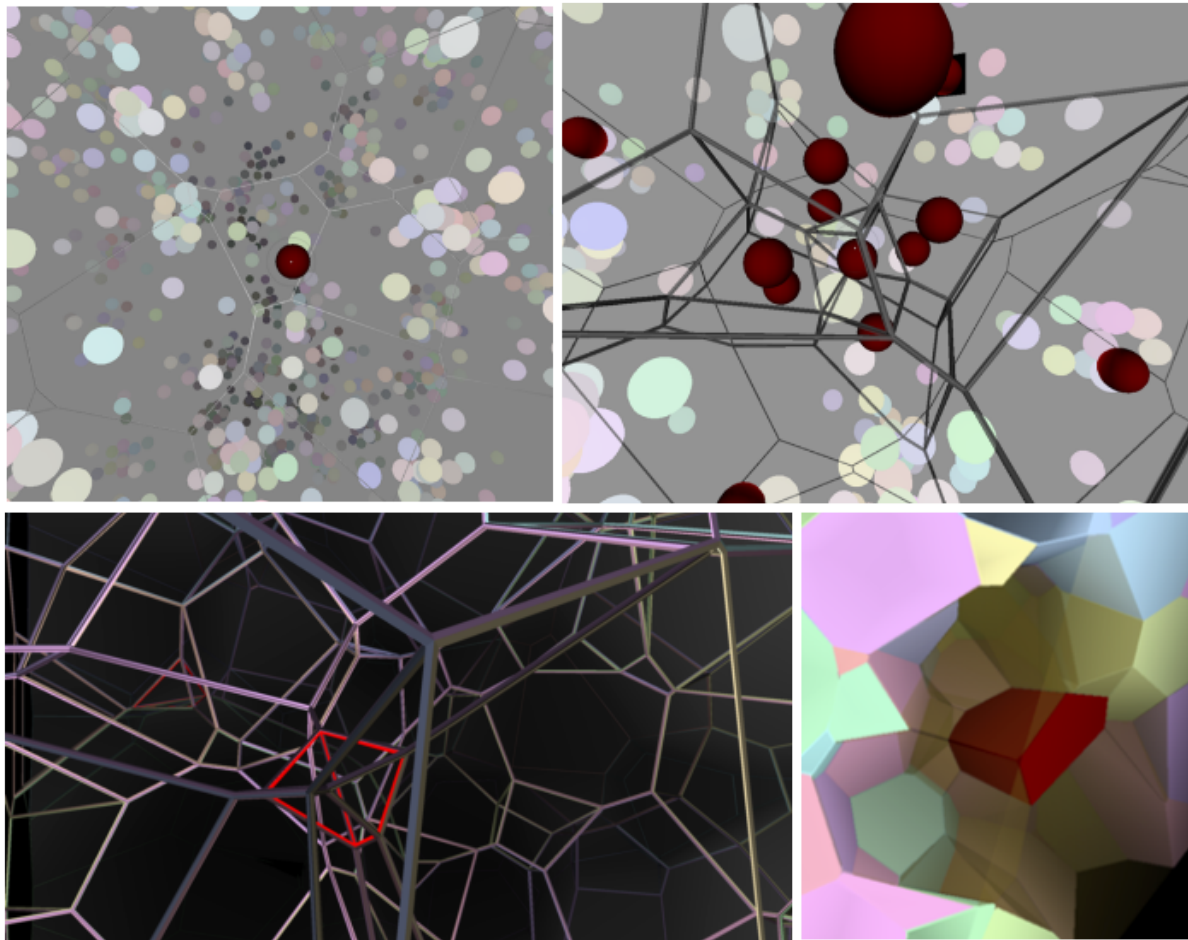


Figure 29 – GrainCrawler’s multiscale features. HR analysis (top row): distant view and focus view; Standard view for size analysis (bottom left) and proximal view for shape analysis (bottom right).

interactive sectioning tool (section 3.4) allows only free sample exploration, and learning gains will depend only on student persistence, exploration ability, and pattern recognition skills. Therefore, teacher-designed guided animations (or movements) are a promising and logical first addition to the system since the rationale for those movements is not established. In future developments of this research project, this tool may allow capturing common animation patterns that may turn into the desired guided or assisted animations for grain models and, ultimately, for microstructural models.

Additionally, methods for movement recording are valuable as a memory enhancement tool for users within a constructivist learning approach, where a student will be able to record relevant paths or movement sequences while crossing the sample, which would be otherwise too cumbersome to execute in free navigation. One will be able to record, for instance, a 3D path along the perimeter of a given grain, in order to investigate and define the number and character of its neighbors. This operation will release precious cognitive resources used on navigation tasks to service visualization, which will benefit especially users with lower spatial abilities, which is a commonly reported gain of animations.

However, executing animations in 3D is not straightforward, and there is no accessible open-source or free software that allows even the simplest animations to be executed by an unskilled user. Therefore, it appears relevant to build a simple tool that allows automating and editing navigation movements in a step-by-step fashion. The literature also mentions gains from the animation of other model properties such as color, rendering mode, and others. In fact, from a technical perspective, any function or property available in the system can be animated (see section 3.3.3). This opens endless and powerful possibilities to leverage all system features to their maximum potential. As the relationship to cinematography was already mentioned, this was called ‘Didactic narrative’ tool. Although testing the feature on user studies is outside the scope of this work, a version is already implemented and will hopefully be assessed in future studies.

### 3.5.3.2 Gamification possibilities to boost motivation in serious interactive software

Given that motivation plays a relevant role in learning and that most target audience members are expected to be involved in gaming, gamification was deemed a useful concept in this project. Gamification can boost motivation to engage, appealing to the intrinsic or high-level extrinsic motivators, tackling motivation, and encouraging behavior using a variety of principles and techniques (HAMARI; KOIVISTO; SARSA, 2014). Different system configurations can be devised to appeal to different user profiles, according to Amy Jo Kim’s social Action Matrix (KIM, 2018), based on the famous Bartle’s taxonomy of player types (BARTLE, 2004): collaborative, competitive, explorative and creative. Customization expands the system’s reach by allowing different profiles to participate, avoiding clashes due to divergent objectives. Relevant gamification procedures and guidelines suggested by both mentioned references are outlined in the remainder of this section.

After defining the student-gamer profile, a tailored gamified strategy can be applied to boost motivation, starting with meaningful learning objectives that can be accomplished by engaging the student into specific challenges or missions. It is argued that offering different activities that effectively leverages the model’s nature as well as its inherent features is the best strategy to achieve student engagement. Thus, an adaptive approach seems ideal, where multiple types of ‘missions’ are possible, involving different degrees of social interaction and varied activities that can be chosen and changed in real-time according to the detected profile and student behavior. The assumption here is that a given student can be stimulated by activities or challenges that effectively trigger his motivations.

In order to achieve milestones, users can progress in several categories, just like RPG-type games: personal objective achievement, reputation, and status, networking, number and variety of accomplishments, negotiation skills in discussions, and leadership. The user can evolve in each category, and the system can show how one can improve in that particular

sector. According to the core values of different profiles, users should be intrinsically encouraged to achieve specific and meaningful milestones, evolving to higher personal levels. Collaborative users may find it more valuable to improve reputation and networking. Competitive users will value aspects like status, number and variety of accomplishments and negotiation skills in discussions, while game elements can show progression and mastery in key profile skills and features. Experience (XP), ranking positions, badges, trophies, and system privileges are traditional rewards in gamified approaches. The choice of feedback must be carefully evaluated to reflect key and meaningful improvements, showing evolution towards mastery and specific user objectives.

#### 3.5.3.2.1 Relevant game elements

As gamification can be accomplished by several means, care should be taken to avoid excess, which could result in the opposite effect, i.e., discouraging students from using the tool. If the system is strongly advertised as a game, serious gamers may feel disappointed, expecting complexity and more challenging or long gameplay and may also find exaggerated gamification attempts to be foolish or childish (HAMARI; KOIVISTO; SARSA, 2014). Therefore, game elements and mechanics should be chosen wisely, kept simple and used in moderation, avoiding creating too complex or clumsy narratives. The goal is increasing motivation by setting some meaningful goals and introducing elements to make those goals clearer.

Interesting game elements in this work are contextual virtual HUDs (Heads Up Display), a mechanism that overlays data onto the view such as an aircraft or spaceship console, making the student feel as a miniaturized explorer inside the material. A HUD may also improve menu adaptability and performance, screen space utilization and interface unification, by bringing all active system elements together in an integrated manner, geometrically and semantically, clearly setting apart controls/menus from the content. If a vehicle or airplane-like console is adopted, for instance, the visual organization of the ‘buttons’ (or commands) can help grouping functions while the visual frame may provide spatial references (Figure 30), which might help to analyze grain shape. Currently, the viewing area’s extra space can house a simple virtual HUD, allowing integrating the three parts of the menu as hierarchical panels. The panels and buttons can be actual 3D elements, as mouse clicks (or finger touches) on 3D elements are already supported, facilitating development. Panels and menu items can be made retractable to avoid visual distractions, an essential feature for user studies.

Other elements that may help build a game-like feeling and boost motivation are contemporary and clean minimalist aesthetics and expected movement effects such as braking, dumping, and accelerations. Those elements can be freely introduced as long as they do not interfere with the system’s primary purpose. Finally, those features should fit

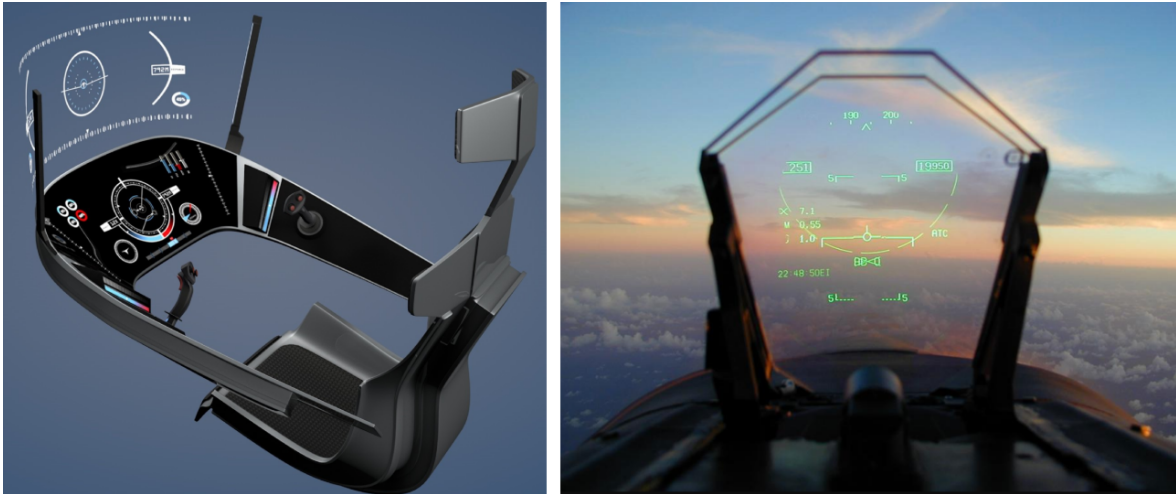


Figure 30 – Examples of virtual heads up displays. Panels can ease menu organization by grouping related functions in a consistent visual fashion (left). Frames may provide spatial references (right). Source: [www.free3d.com](http://www.free3d.com)

coherently into a context or a scenario design. Given the application and target audience, a suggestive and credible context could be a real or fictitious scientific tool simulator.

### 3.5.3.2.2 Game mechanics ideas

Suggested permanent game mechanics elements are experience indicators, accomplishments, social reputation or benchmarking, exploration history, interest analytics and activity traces on social networks. Ideally, all such elements must be adapted to distinct user profiles. Free exploration can promote improvements on all factors, but some specific activities can be proposed to each profile, appealing to specific motivators. Gamer profiles were proposed in the literature, basically summarized into four (HAMARI; KOIVISTO; SARSA, 2014): cooperators, competitors, explorers, and socializers.

The tutorial mode might push students to explore the model features, which might lead them to engage in other specific gamified activities. In this mode, a standard narrative informs the purpose, features, and available tools. Access to a library containing previously built narratives, other users (teachers and students) comments, and annotations may help build knowledge and conceptual connections between model features. Besides the tutorial mode, playful activities are suggested to students based on their profile, allowing them to improve in key personal categories:

- The first type of activity may focus on meaning building in the form of discovery games, which may involve tagging, posting, and digging information from discussions in chats or other sources. This suits beginners or cooperators, usually more interested in social aspects. Meaning can be related to the teacher's intents, in this case, a detective game that can be played alone or with other students to discover relevant



features, establish spatial or conceptual connections, and share insights. Teachers can create special narratives for some themes or ‘chapters’ considering possibilities allowed by the model. Interesting games can be created leveraging the learning purposes defined by the teacher.

- The second type of activity may focus on aesthetics, appealing more to beginners socializers interested in such aspects, luring them into interacting more with the model and discovering deeper relationships. The user seeks engaging visuals or composes its own narratives and post screenshots or links to that specific view/narrative, fostering others’ participation in social networks. A public ranking is displayed, and the student can see top-ranked images or narratives and engage in discussion. The teacher can comment and show the students relevant microstructural characteristics or properties revealed by each posting.
- Another possible activity appeals to explorers, stimulating them to find all model features, narratives, or user-created content in the form of detective games. The student can post findings and compare them to others’. In fact, this activity can also be appealing to cooperative and competitive profiles. The student tags features and sees others found by other students. More information can be harvested in lists about that model, where evidence collected by others can be viewed.
- An interesting activity for explorers is to escape the ‘heat-map’. The student is urged to visit neglected areas, comment about them, find features or patterns, and suggest why they seem less appealing than others.

Most of these ideas rely on online forums or asset libraries, and this kind of functionality is not implemented. Currently, only online data storage is available for testing purposes. In fact, the prototype’s scope in this work is restricted to the execution of formal user studies, and prospective gamified features may interfere directly with experiment results. However, given gamification’s potential to boost interest, motivation, increase task time awareness, and diminish some negative extrinsic factors, the matter will be further discussed.

#### 3.5.3.2.3 Prospective mechanics for experimental setups

In the experimental stage, gamification may improve interaction learning by motivating the participant to engage and concentrate and to spend more time on the tutorials. All tasks can be presented as stages or missions. Visible feature counts and positive feedback upon task completion may boost motivation. Subtle warnings may appear whenever a participant gets stuck in a position or stay for too long in a sector.

Time is an import issue, both for the experiment logistics and for the results. Logistics imposes a time limit per task, especially on collective sessions. However, time is also an

independent measure, and mechanisms that hasten participants may produce the opposite effect and reduce performance. Therefore, a subtle way must be found, and gamification may help with this issue.

A gamer profile questionnaire can be included in the background survey to tune the gamified features accordingly, applying an aggressive time and performance setting for competitive users - they certainly will desire to be the fastest and to consume all available missions. An intermediate setting should suit explorers and a subtle one should be enforced for collaborators and creatives, which probably may feel stressed by a timer. If previous profiling is not possible, a simple game mechanism that may fit all profiles can be a limited resource, such as 'fuel' or energy or to avoid overheating the sensors. In this case, sensitive participants may feel somewhat pressed, but not necessarily in a negative way, as resource limitations are pervasive in anyone's life. This is a matter of perspective, i.e., hurrying a person is a different situation than recognizing that a task must be accomplished before a valuable resource ends.

On the other hand, performance boosting may interfere with experiment results, as most mechanisms require giving hints or indications of what is good or bad, such as detection accomplishments. However, such mechanisms may be freely used in tutorials in order to show the participant what a valid feature looks like (how a big or small grain looks - see section 4.6.8). On the other hand, mechanisms can help to avoid bad performances caused by extrinsic factors such as distractions, laziness, and spending too much time stopped or stuck in a sector. The preliminary experiment indicated that it is usually better to give the rest of the model a chance and explore it at a constant pace than getting stuck in a given sector trying to find a feature or to confirm whether it is a valid feature. Warnings can be shown, or game elements might urge the participant not to stay stopped for too long. An example is setting a maximum time for using the 'ship stabilizers' or forcing the player to search and consume 'energy units', which would be evenly distributed through the sample. The later mechanism will also urge the participant to explore the entire sample, where the total available energy in the scene represents the time available for that task. Rewards or hints can urge participants to move and explore when they are too slow, get stuck or have not selected a feature after some time. In this case, the view can move slowly to the closest valid feature, and a message indicates that it looks like a valid feature, urging the participant to find similar ones.

This section's purpose was discussing gamification possibilities to tackle enduring experimental challenges of participant behavior, which could help increase the scientific reliability of studies' outcomes. Thus, implementations of such features should be thoroughly tested by further studies, and it would be risky to apply them in this work. However, some steps of the process, such as pre-tests and tutorials, may adopt pertinent features, not interfering with the main experiments.

## 3.6 Development summary

An interactive visualization system for microstructural models was developed, aiming improvements in materials science education, given the limitations of didactic materials in supporting the comprehension of fundamental microstructure topics. This work particularly pursues improvements on the spatial perception of such honeycomb structures. Despite the focus on a target audience of engineering students, other related areas were considered, especially physics and chemistry. Since engineering students are expected to have good spatial abilities, it was hypothesized that the proposed methods should perform well for the target audience.

The multiscale paradigm was deemed a promising approach to enhance the learning of microstructures, given such models' nature. A systematic review indicated the challenges, limitations, best practices, and prospective features of this visualization paradigm. Since the grain structure (individual cell) is considered the most fundamental feature of microstructures and no specific interaction method exists yet, this work's scope was restricted to methods to interact specifically with grains.

The ThreeJS library was adopted in this project for WebGL development and the grain models used in this work were synthesized according to the procedure described by [Rodrigues et al. \(2018\)](#). The system loads models and offers navigation controls using either conventional input methods (keyboard, mouse) or mobile controls. Available rendering modes allow balancing image quality, rendering performance and introducing relevant depth enhancing effects. The user interface is minimalist and adapts to different devices. Advanced operations are supported, such as object picking, grain neighbors filtering, planar grain fracture, volume clipping, and user-defined animations or 'didactic narratives'. Such operations sought supporting didactic activities.

Three techniques were developed, namely, discrete sections (S), realtime sectioning (DS) and GrainCrawler (GC). The systematic review indicated sectioning as a viable filtering solution in highly occluded and cluttered environments. Implementing an interactive sectioning method in the chosen WebGL platform was considered the most natural starting point. S and DS are section-based visualizations with five rendering modes: flat 2D, 2.5D, and ambient occlusion, while 3D is available as Anaglyph or split-screen stereo. Flat 2D discards the shade component, so all faces reflect the pure RGB color of the grain, creating the illusion of a solid section. 2.5D adopts the standard shader, allowing the user to see the remaining portion of the grains behind the section plane. Depth perception improvement was implemented in 2.5D rendering using ambient occlusion techniques, simulating shading as if the model was really sectioned.

GrainCrawler is not section-based and pursued a combination of prospective features of multiscale visualization, mainly level of detail techniques, object filtering, grouping,

spatial references (ground plane, orientation cube, face identification), and alternative graphical representations. The sought improvements over section techniques were increased field of regard and better preattentive features, purposefully selected after each perceptual objective (detect extreme sizes, heterogeneous regions and shapes), resulting in four visualization modes: the sample-wide view allows detecting heterogeneous regions for detailed inspection, by visualizing scattering patterns; the focus view assists finding the central grain of the heterogeneous region; the standard navigation view aims to detect smaller and bigger grains, representing farther grains as solid objects and closer ones as wireframe cages; in the proximal view, the selected grain assumes a solid representation, and neighbors get transparent to facilitate individual shape evaluation.

Prospective educational features were proposed and discussed, particularly guided animations. The system already allows teacher-designed camera paths, opening endless possibilities either for a teacher to suggest deep relationships among spatial elements or for a student to explore 3D models and create persistent records. Given the relationship to cinematography, this tool was called ‘Didactic narrative’.

Gamification was also discussed as a powerful feature to increase interest and motivation by setting meaningful goals and introducing elements to make those goals clearer, according to distinct user profiles. In experiments, gamification may help improving learning the interaction and perceptual goals by motivating the participant to engage and concentrate and also to spend more time on the tutorials.

## 4 Evaluation

Given this project's educational goals, the focus of this manuscript up until this point was on ideal features, holistically integrated into the proposed tools. However, the limited time and resources available to this project call for restrictions in scope that impede the full scientific assessment of such a complex software system and its features. Therefore, experimental versions were developed to study prioritized factors briefly discussed in the previous chapter. Therefore, the remainder of this manuscript focuses on the experimental system and user study issues. The conceptual framework upon which the experimental system was based is discussed in sequence.

### 4.1 Conceptual framework: the interaction-visualization immersive continuum

Unconstrained exploration of virtual 3D environments requires two basic features: a navigation method that allows 6 degrees of freedom and rendering techniques that emphasize the model's most relevant characteristics. As discussed in section 2.2, navigation and visualization require cognitive resources at a level compatible with their complexity. In this case, they are performed simultaneously, and as the interplay between them is not likely to be linear, the cognitive load tends to be higher than the sum of individual requirements. Thus, given the proposed techniques (interactive sectioning and GC) and following cognitive research indicators from the systematic review on multiscale systems, it is hypothesized that spatial comprehension of grain structures can be improved incrementally as interaction and visualization reach 'natural' levels, i.e., as both interaction and visualization abstractions are reduced.

Further research on this topic led to the classic immersion concept, which is pervasive and heavily discussed in the virtual reality community due to alleged potentials for cognition enhancement. Recently, interactive and visualization technology gained momentum and achieved enough maturity, allowing assessing immersion concepts in educational applications. Immersion can be conceptualized in two complementary ways that can be merged into a single theoretical framework, helping to conceive an immersion progression. McMahan (2011) concept is based on the traditional notion of immersion as perceptive simulation, being total immersion the full experience of reality by human senses, suggesting a continuum where immersion level rises with diminishing interaction and visualization abstraction. On the other hand, Thon (2008) built his immersion theory on the process of attention channeling to other activities, in this case, a computer program, evaluating

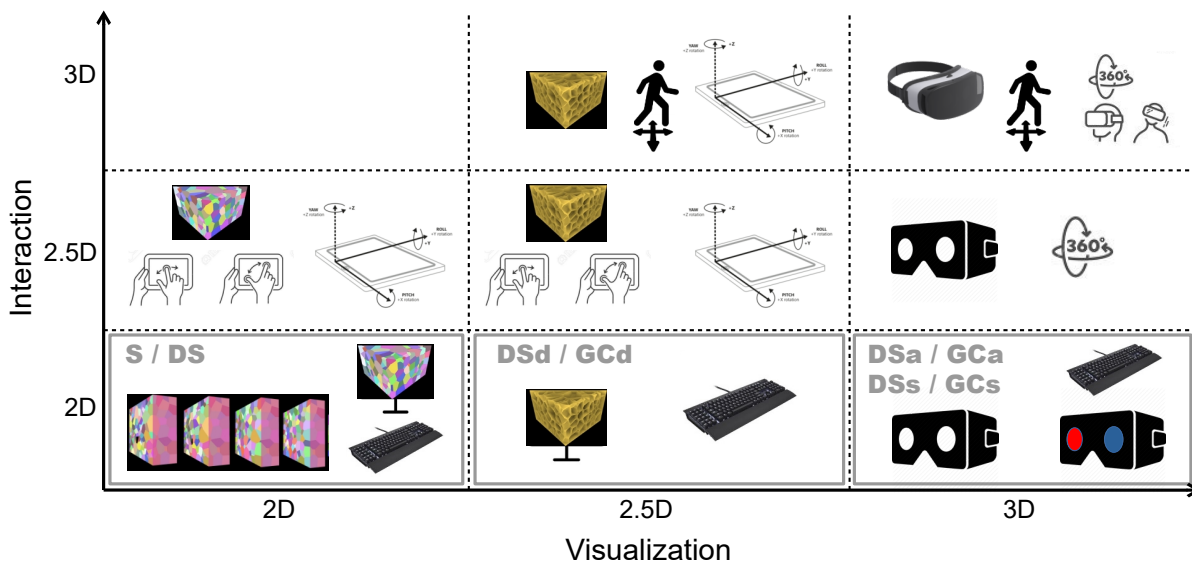


Figure 31 – System variations resulting from relevant combinations of interaction (from abstract to natural) and visualization modes (from flat 2D to split 3D). 2D interaction relies on abstract commands in devices such as keyboards and mice. Rotations and touch commands (2.5D) with hands are considered more intuitive, while walking and head rotations are familiar natural movements (3D). Visualization, in turn, progresses from 2D, depth-enhanced 2D, to high FOV stereoscopic vision. Combinations highlighted in grey boxes are within the scope of this work.

four dimensions of immersion, each one based on a different attention attractor (fun, challenge and sociability, aesthetics, spatial factors). In fact, stimulus and curiosity are proven essential factors to educational applications (BYRNE, 1996), as also indicated in the preliminary user study (see sections 4.4.2.4 and 4.4.2.5).

Therefore, the union and application of both concepts might unify interactive seamlessness with ludic stimulus, inspiring several system configurations, where immersion level is supposed to rise (Figure 31) as a function of diminishing interaction abstraction (2D with keyboard, 2.5D with touchscreen, 3D with gestures and walking) and more natural visualization (2D with monitors, 2.5D with improved depth perception, 3D with anaglyph and split modes). Proposed configurations are bounded as well by technology costs and accessibility criteria that are relevant for educational applications. For more details on those criteria, see section 4.4.

Although the comprehensive evaluation of this immersive continuum is outside the scope of this work, it is important to lay down the conceptual framework upon which the system is founded. As perception and cognition are considered a priority in this first exploration, the experimental system will focus on the visualization axis of this continuum, and the mentioned visualization modes will be assessed on experiments (2D, 2.5D, 3D anaglyph and 3D split-screen).

## 4.2 Selection of dependent variables and evaluation strategy

Given DeHoff (2012) and Hornbogen (1986) microstructural formulations, learning recommendations, and the main goal of this research - improving spatial comprehension of honeycomb structures - the dependent variables must be able to express the desired improvement precisely, i.e., to be directly related to spatial comprehension. The chosen microstructural features of interest reflect three ‘scales’. The sample scale, or aggregate properties, is related to localized variations of grain size. The neighborhood scale is related to a given central grain and the relationships with its neighbors. The grain scale concerns individual grain size and shape.

Recalling the discussion in section 2.2, the approach of most HCI studies is to measure time and performance on one or more tasks, in order to compare variations of visualization or interaction methods. Given the three relevant scales described above, one can say that a successful technique for spatial comprehension of relevant grain structure features will first allow detecting heterogeneities and extreme values in the sample scale, then grain size and shape factor in the grain scale. We can also affirm that the superior technique will allow faster detection of those features. Typical dependent variables for such visual search tasks (see section 2.2.2) are detection performance and reaction time.

Martin (2007) presents a concise review of experimental techniques and procedures applicable to HCI studies, including the relevant statistical analysis methods. Aside from the proper choice of independent variables, another relevant decision is the experimental design. Typically, HCI studies are conducted with quasi-experimental designs, due to the practical difficulties in enforcing the strict rules of experimentation, i.e., achieving proper randomization in participant acquisition and assignment. In the absence of an ideal scenario, where groups are entirely randomized, several measures are usually adopted to minimize internal validity threats. Typical threats are maturation, history, mortality, selection, testing, and statistical regression, often introducing biases and confounding factors. Students are usually the ‘most available’ type of participant, which may introduce biases, though this is not an issue here as college students are the target audience. Still, most threats to internal validity can be accessed by modern quasi-experimental designs, and robust statistical methods and tools are available to perform and analyze the results of this type of study (MARTIN, 2007).

Another possibility suggested by Martin (2007) is single-subject and small-N baseline designs, requiring only one or a few participants. The assumption here is that in some cases, the change in behavior caused by the introduction of a ‘treatment’ or manipulation of a variable can be reversed, and substantial differences can be detected. This is not the case when the most relevant behavior involves memory or learning, where reversion is not possible, or widely varies between individuals. Advantages are higher control of experimental conditions, deeper insights into individual behavior, easier interpretation and

analysis of results, and focus on larger effects. However, experimenters must be able to reverse the effects consistently; otherwise, it is not clear if the transition-state behavior is due to manipulation or to confounding variables. Additionally, small effects that might be important are difficult to detect, and effects cannot be generalized, as individual responses may differ radically. This issue can be dealt with by including a few other participants.

Given the possibilities described above and the nature of this work, a small-N baseline design is a possible alternative to detect strong effects and help discover and detail all behavior relevant to the tasks. On the other hand, quasi-experimental group testing is suited to find both large and small effects, and allows confident generalization of results.

### 4.3 Preliminary task design

Three tasks were designed to assess whether participants are able to identify relevant features of grain structures with the three techniques, given the dependent variables defined above. It is important to remind that only sectioning techniques were tested on both preliminary experiments.

#### 4.3.1 Size tasks

A task was designed for grain size identification, where the participant is instructed to pick 6 to 10 grains among the larger (Figure 33 and 38) and 6 to 10 among the smaller grains (Figure 34 and 39) in the sample using only (see details on section 4.1) forward and backward movements. As discussed so far, choice time and size rank are the main measurements. Movement restrictions were enforced to simulate the kind of visualization afforded by real samples when serial sectioning techniques are available, i.e., rotation was not allowed since real materials' sections are parallel and the sample is destroyed after the sectioning procedure, not allowing imaging sections in other angles. However, the main purpose of restricting movement to a single axis is limiting the influence of interaction on feature detection and also to allow comparison between techniques.

For the remainder of this manuscript, the experimental task of detecting smaller grains is denoted by Ss, while Sb is used for bigger grain detection.

#### 4.3.2 Shape task

The second task was designed to check whether the participant was able to evaluate the shape of a specific grain with continuous 6DOF navigation. Therefore, this task initially aimed to assess if the proposed technique allowed detecting grain shape at all, as well as the main difficulties and limitations, in order to improve techniques before detailed formal studies. Three different grain shapes were proposed initially based on an arbitrary



‘stretching’ parameter (see figures 35 and 36). Afterward, four shape factors were defined (equiaxial, brick, discoid, stretched disc) based on an algorithmic analysis performed on the digital grain model (figure 42). The bounding box of a grain can represent its physical shape, though it may be much different in appearance than the grain itself. Shape can then be defined by the size ratios between the box sides. In order to simplify this description, each shape can be specified by a combination of a larger and a smaller face, and the third face just depends on both. Given three generic edge sizes (short, medium, long), the shape of a single face can be square (S), rectangular (R) or long (L - a stretched rectangle) and, therefore, the nine possible shapes are Ss, Sr, Sl, Rs, Rr, Rl, Ls, Lr and Ll. However, the algorithm found only Ss (equiaxial), Sr (discoid), Rl (stretched disc), and Ls (brick) shapes in the sample (highlighted in Figure 32).

The participant inspected three grains per shape class by rotating around each one, using keyboard arrows and Q and E buttons. The system interface provided a simple menu so the participant could specify one of four available shape factors.

### 4.3.3 Heterogeneous regions tasks

Finally, the third task aims to assess if students are able to capture the unusual concept of ‘heterogeneity’, defined simply as clustering of small or big grains. Participants were asked to detect cluster ‘epicenters’, for big grains clustering (Figure 33 and 40) and small grain clustering (Figure 37 and 41), using interactive forward-backward navigation, observing the same restrictions enforced on the size task.

Eventual rotation was allowed on the GrainCrawler technique to help users to find epicenters with precision (Figure 41), as the perspective view usually suffers from occlusion. When the user starts rotating, movement speed reduces drastically to increase precision and impede free sample exploration in different angles. Then the user can hit the Q button to reset the position and restore normal speed.

For the remainder of the manuscript, the term heterogeneous regions is designated by ‘HR’ while HRb designates the experimental task of detecting bigger grain clusters, and HRs stands for smaller grain clusters.

### 4.3.4 Possibilities for grain neighborhood analysis

The neighbor scale was left out of the scope of this work; however, the results of the third task (HR) will indicate whether it is possible to detect spatial neighboring relationships and efficiently inspect the surroundings of a given grain with interactive rotation. However, the simpler first approach seems to check if is possible to count the neighbors of a single central grain by rotation. This task heavily relies on memory and, therefore, the best technique should enhance performance (total time and count accuracy)

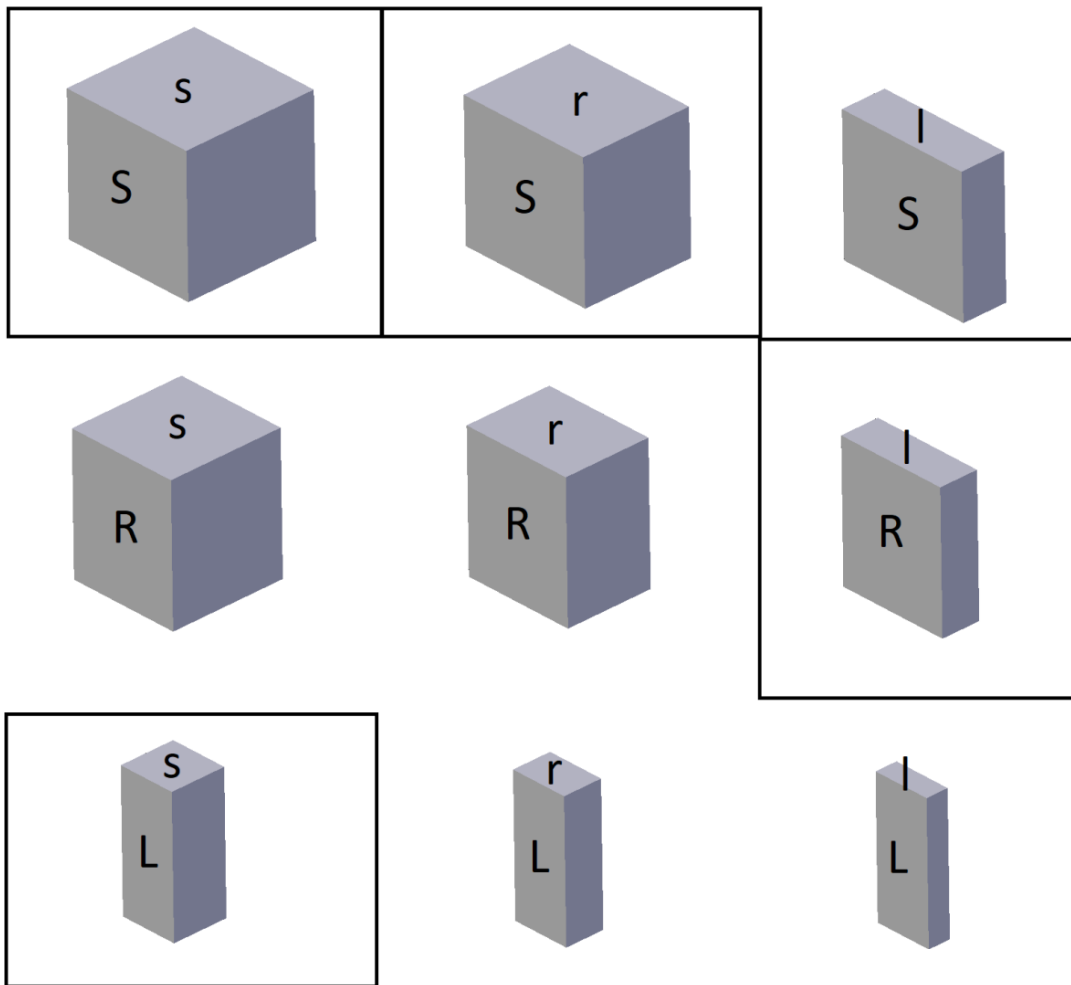


Figure 32 – Nine possible shapes of a grain, given the combinations of three predefined 2D shapes (Square, Rectangle, Long rectangle) for each of the two main sides. Only the four highlighted shapes were found in the models.

by decreasing memory needs. This might be accomplished by mnemonic references, the ability to show more neighbors instantly, and to build stronger spatial maps. Flat 2D, for instance, shows only sectioned neighbors, requiring rotation to count them all (Figure 43), while 2.5D techniques show the remaining internal faces, which allows seeing more neighbors at a given moment. Therefore it is argued that the user will need less time using 2.5D and probably achieve higher accuracy as the subject can count all grains using only two positions ( $0$  and  $180^\circ$ ). Improved depth perception might help to build stronger spatial maps and a better perception of neighbor-neighbor relationships, which will probably improve count accuracy. Figure 43 shows that, for a small grain, at least three views are necessary in 2D flat mode to detect all nine neighbors and a fourth to reach the mirrored image and stop counting.

Figure 44 shows that two mirrored views ( $180^\circ$  rotation) suffice in theory for 2.5D; however, a good resolution is required to perceive small faces (see the zoom box in 2<sup>nd</sup>

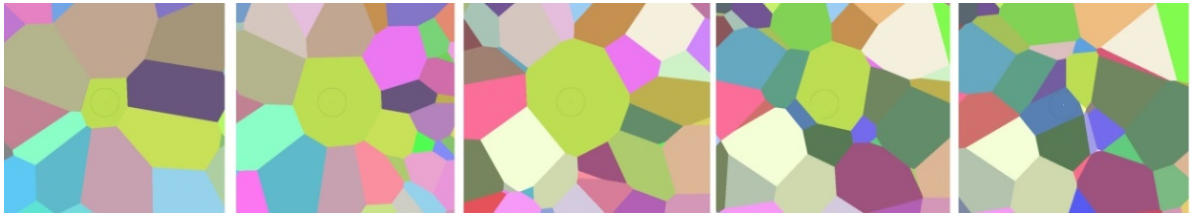


Figure 33 – Example of a large grain within a cluster: note the ‘flower’ pattern around the central grain, formed by other large neighbors.

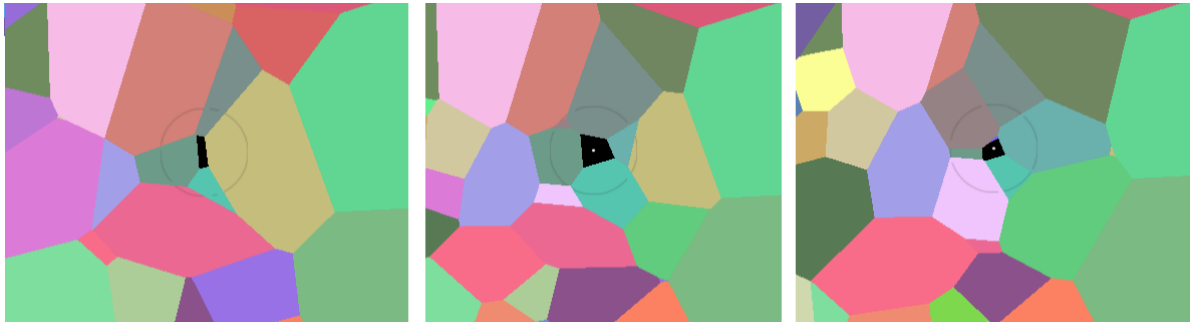


Figure 34 – A small grain appears briefly during translation.

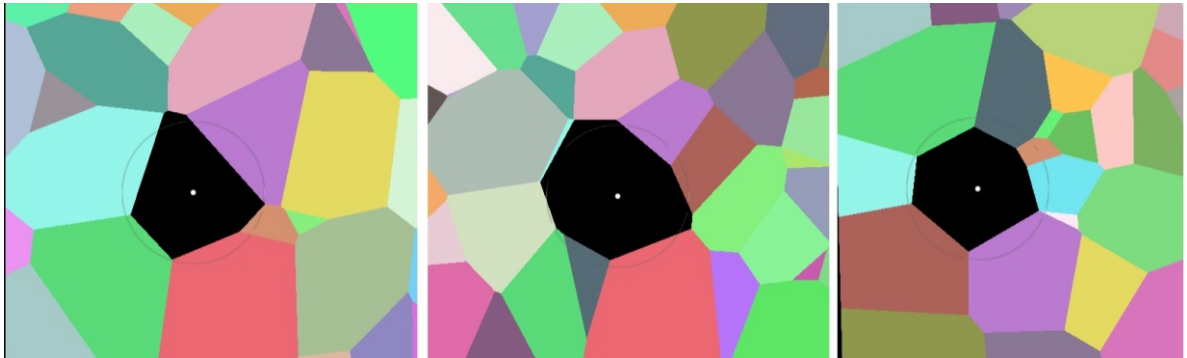


Figure 35 – Rotation around equiaxial grain.

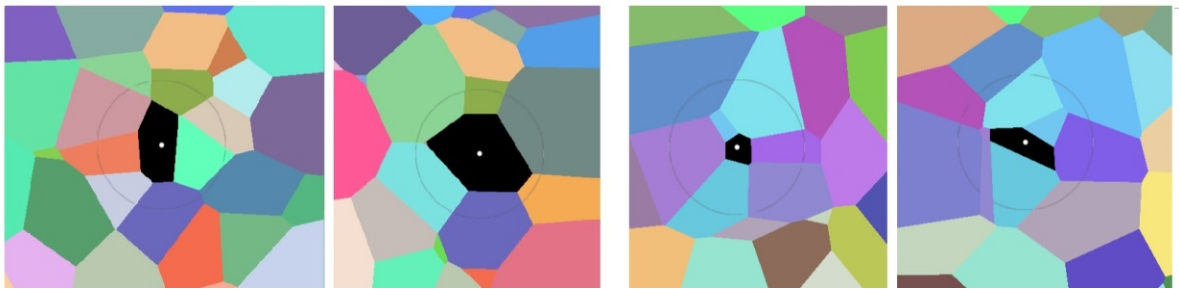


Figure 36 – Rotation around stretched (two images at the left) and highly stretched grains (two images at the right).

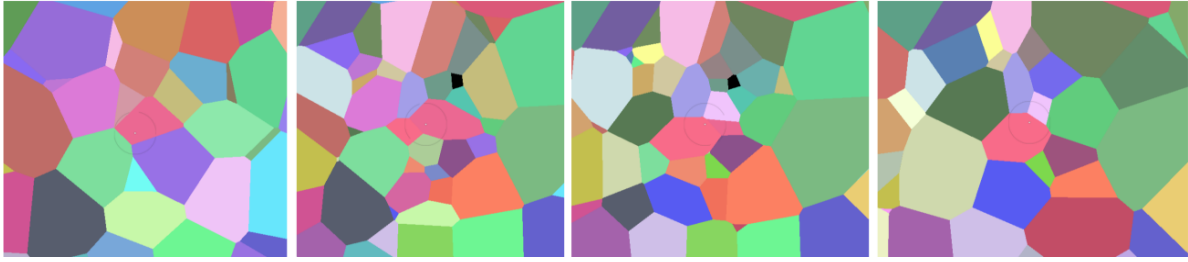


Figure 37 – Small grain heterogeneous region: note the clustering of small and medium-sized grains around the black one.

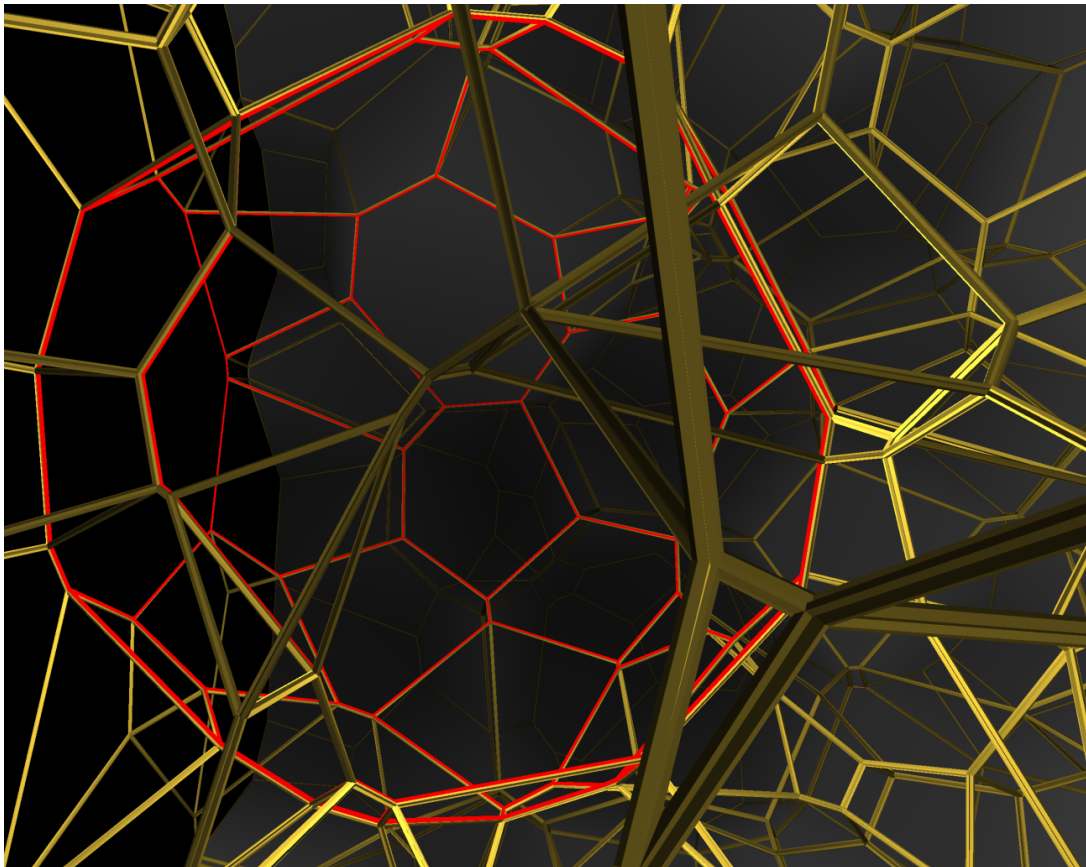


Figure 38 – A large grain visualized in GrainCrawler (highlighted in red).

image) and some spatial ability to recognize which face belongs to sectioned grains and to imagine a mirrored image. The resolution issue appears in enhanced 2.5D (Figure 44, zoom box of 4<sup>th</sup> image) as the user cannot tell if there is a new small face in the far back or if it is just a shadow. As color seems to be an adequate mnemonic aid here, the absence of color in enhanced 2.5D may in fact cause a decrement in time and accuracy. Therefore, testing a colored enhanced 2.5D version seems highly relevant, at least for this task.

This first analysis indicates that flat 2D might be a better technique to count grains, while standard 2.5D may afford superior spatial reasoning. Thus, a different test procedure is necessary, assuming that building a good spatial mapping is more relevant than counting neighbors. Additionally, counting grains seems not to rely as strongly on movement - a

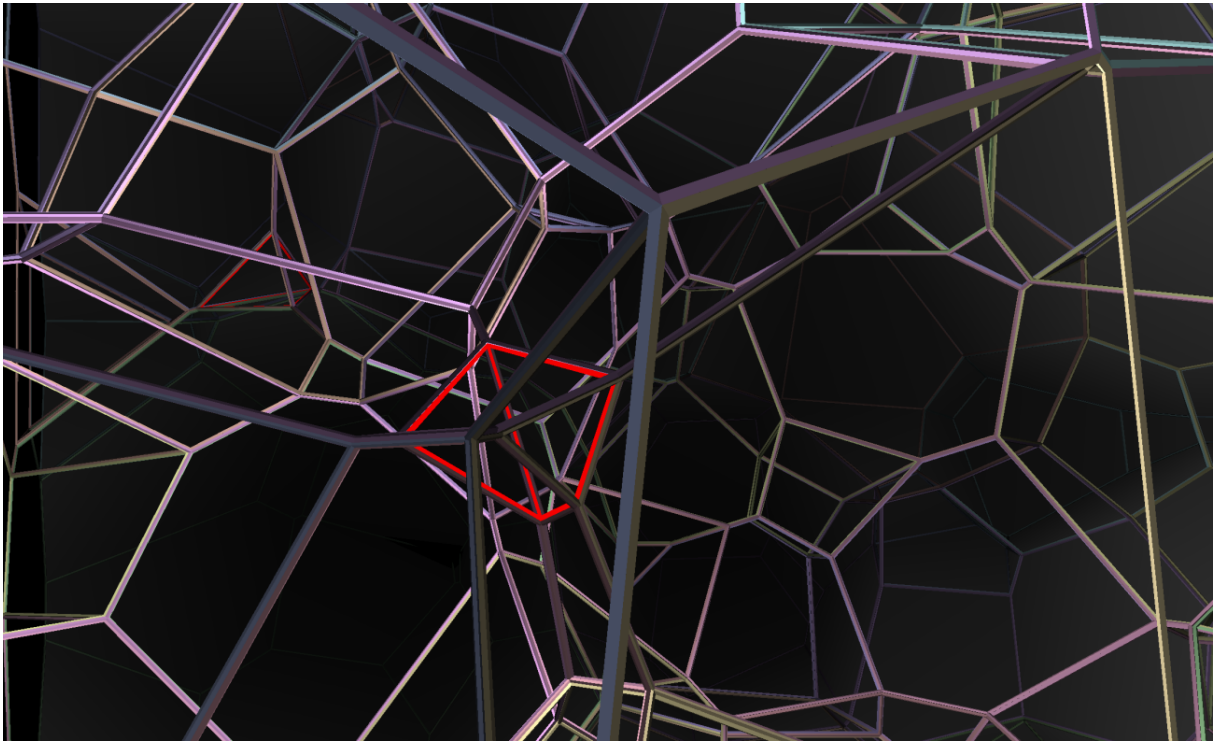


Figure 39 – A small grain visualized in GrainCrawler (highlighted in red). Note the higher concentration of edges around the small grain. In this task the participant is instructed to pick 6 to 10 grains among the smaller using only translation movements, i.e, rotation is not allowed.

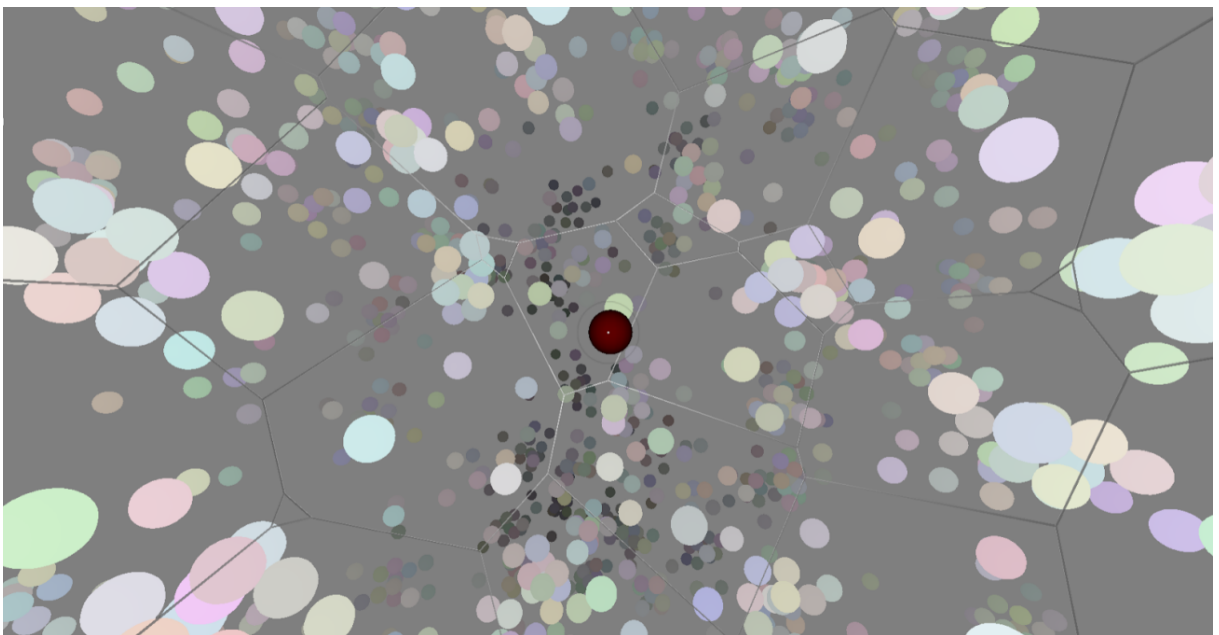


Figure 40 – GrainCrawler: visualizing a heterogeneous region comprising big grains. Note the wide space between adjacent grain centers.

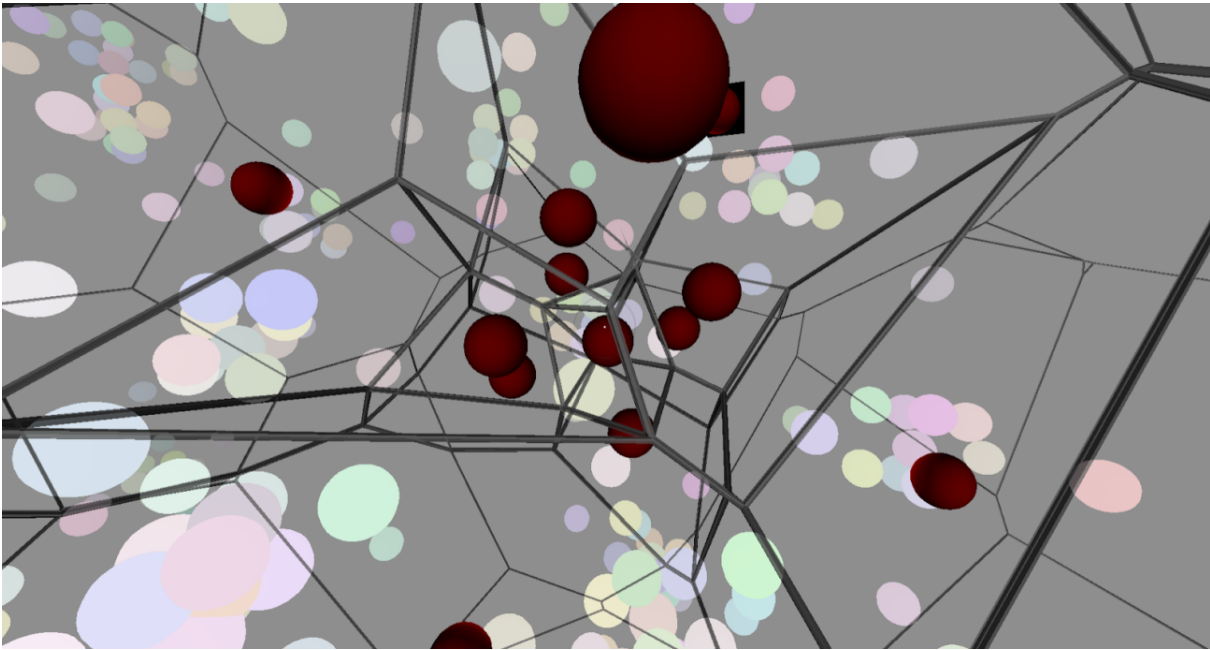


Figure 41 – GrainCrawler: visualizing a heterogeneous region comprising small grains. Note the short distance between adjacent grain centers.

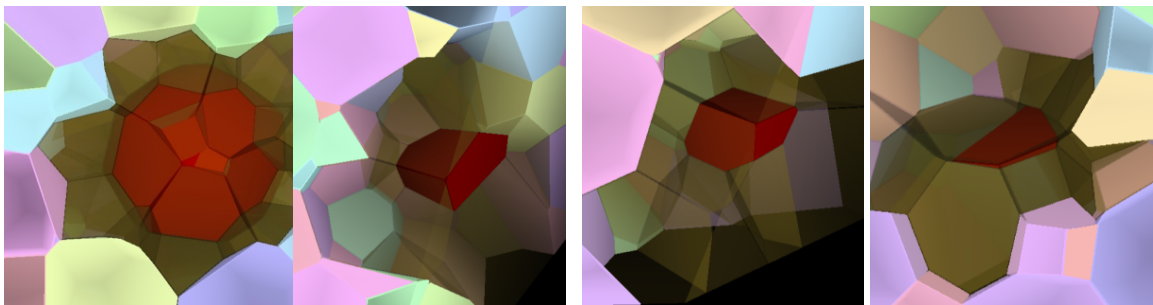


Figure 42 – GrainCrawler: visualizing each of the four shapes (equiaxial, discoid, brick and stretched disc).

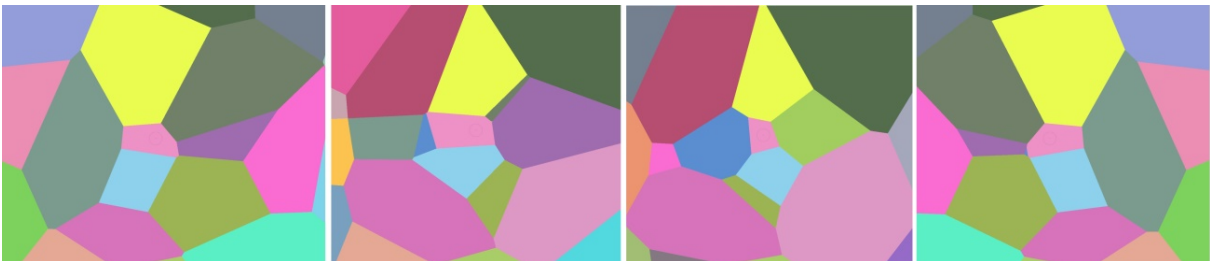


Figure 43 – Detecting neighbors with a  $180^\circ$  clockwise rotation in flat 2D visualization. At least four angles are necessary to detect the 9 neighbors of the central pink grain and color is clearly a good mnemonic aid. Note the mirroring between first and last images.

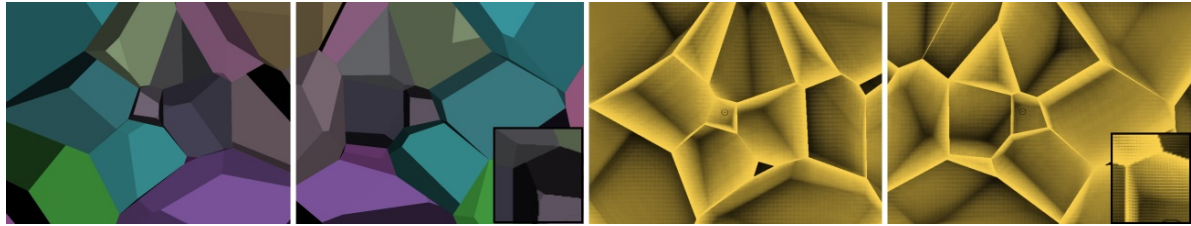


Figure 44 – Although two angles might suffice to detect all neighbors in 2.5D modes, resolution limitations (left, zoom box) and shading properties of ambient occlusion rendering (right, zoom box) might mask small faces.

fundamental advantage of interactive techniques - as does detecting topological relationships. Finally, the above example is a simpler small grain with only nine neighbors in a sample that averages 14 neighbors per grain and can reach more than 20.

On the other hand, GrainCrawler may be a better technique on both ends, offering two ways to count grains or analyze a particular grain's surroundings. The perspective view of wireframe cages and sphere representation of all grain centers in a single view (Figure 41) reduces memory needs and shows spatial relationships clearer than sectioning techniques. The shape analysis mode (Figure 42), on the other hand, seems better than the others as it still shows half of the central grain with a superior sense of shape. Neighbors are clearly highlighted with transparency, depicting individual shapes better and revealing the remaining neighbors in the back. Even better would be representing neighbors with colored wireframes, improving mnemonic aids.

## 4.4 Preliminary user study

The goals of the preliminary user study are:

- indicate if tasks allow efficient assessment of chosen dependent variables;
- observe tendencies and variations of participant behavior and strategy;
- evaluate tutorials efficiency;
- detecting if the target audience can complete the tasks within a reasonable time;
- check the character of response distributions;
- check for relationships between tutorial and task performances;
- correct bugs and limitations;
- suggest technique improvements;
- collect initial user feedback;

- detect relevant logistical issues: time requirements, comprehension of task instructions, other unforeseen issues or restrictions.

### 4.4.1 Execution

The study was realized with a single group, testing the DS/S platform. This group was composed of 10 undergraduate students from engineering, physics, chemistry and biology courses, and the study was performed over 1 hour as an exercise, applied in a computer laboratory after a materials science class. The platform is the 2D desktop (keyboard controls, mouse for selections, monitor for visualization).

Before starting the tasks, students were given basic information about grain structures, a short briefing about the system and the tasks, and a five-minute tutorial to get familiar with the interaction technique. This tutorial was freely executed in the first attempt and then repeated for time as soon as the students were considered ready. Although the tutorial's main goal was assessing navigation proficiency, an indirect indication of spatial ability levels could also be obtained.

The four tasks were executed in sequence. Each task's procedure begun with a description, some time for questions, and then students typed their ID code in a dedicated field, identifying the user and the task. When ready, students clicked on a start button, initializing the task clock. A counter was available to show selected grains. Given that all students explored an identical grain structure, each student was assigned a unique angle (six cube sides plus four diagonals) to avoid mutual help.

Students were asked to give written feedback and answer ten questions about their background in computer games, 3D graphics applications, drawing classes' performance, relevant personal traits (persistence, patience), stimulus or joy using the system, and mental stress in task execution. The questions aimed to indicate previous experience or training in the navigation technique, spatial ability level, mental stress during and after the exercise, and any factor (stimulus, impatience) that may impact the results. Such factors are usually assessed in user studies as they might explain unexpected values or behavior and usually correlate with performance results. Finally, students will be asked for feedback about the procedure.

### 4.4.2 Desktop platform results

#### 4.4.2.1 Response distributions

Choice time and score (size rank) response distributions, as well as their residuals, were found normal according to Shapiro-Wilk tests, conforming to ANOVA normality assumption. Only T1 ( $p=0.0076$ ) and T4 ( $p=0.035$ ) scores deviated from normal, although



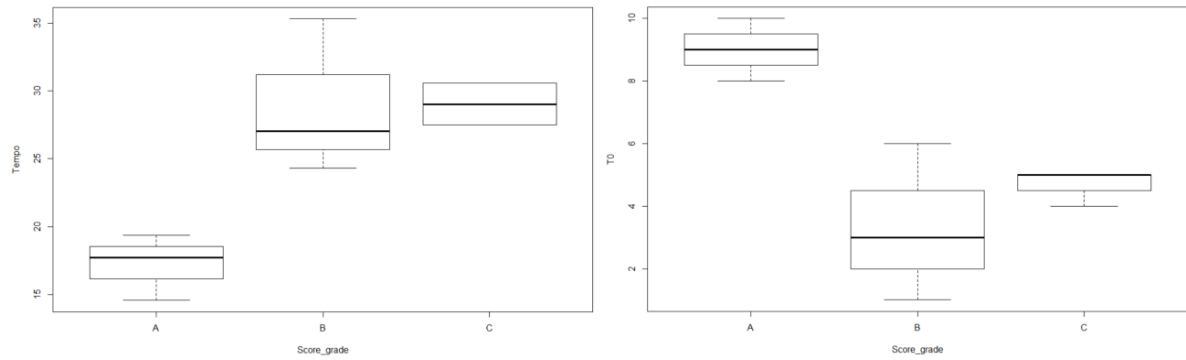


Figure 45 – Left: overall performance X total time. Right: tutorial performance (time) X overall performance.

T1 is probably normal for their residuals did not deviate significantly from normal. T4 was further tested and passed the Kolmogorov-Smirnov log-normality test.

#### 4.4.2.2 Results

A detailed data analysis procedure was performed to assess tendencies identified in the preliminary analysis. One should note that this process can only reinforce tendencies, not prove or disprove them, as this test had few data points, and standard deviations were not always compatible to allow proper pairwise comparisons.

- Overall performance X total time

Three groups (A, B, C) were defined according to overall performance. A shorter total time was expected from the best overall performers (ascending line). This hypothesis was not rejected (A clearly better than B+C), although a distinction between B and C groups could not be adequately detected (Figure 45, left). Standard deviations are pretty close on three groups.

- Tutorial performance (time) X overall performance

A superior overall performance (descending line) was expected from top performers in T0 (tutorial). ‘A’ group was apparently superior to B and C. Although a significant difference could not be detected between C and B, there is a suggestive superiority tendency towards C (Figure 45, right).

- Tutorial performance (time) X performance on each task

Three groups (A,B,C) were defined according to performance in T0 (tutorial). The best performers in T0 were expected superior performance (descending line) in all tasks (H1) or only in T3 and less in T4 (H2). An apparent difference was found in T1 (Figure 46, left) between A and B groups, inverted between A and C, inverted between B and C. A and C were notable superior to B. An apparent difference was

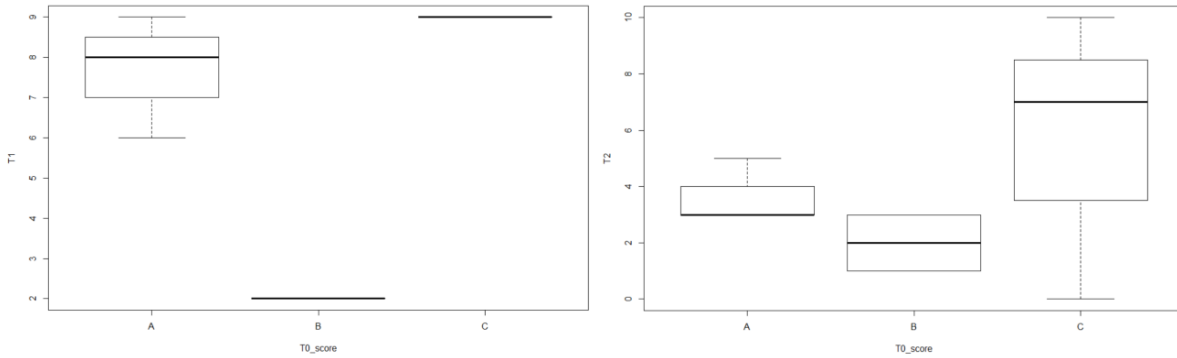


Figure 46 – Tutorial time X performance on each task on T1 (left) and T2 (right).

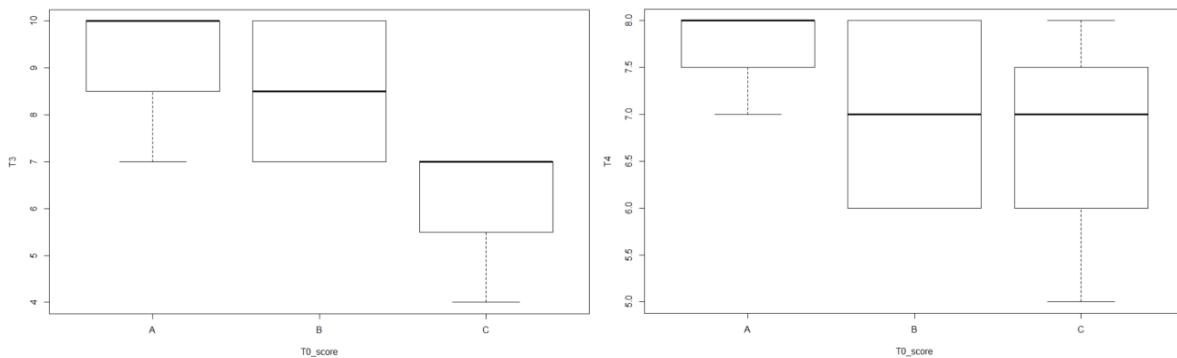


Figure 47 – Tutorial time X performance on each task on T3 (left) and T4 (right).

found in T2 (Figure 46, right) between A and B, inverted between A and C, inverted between B and C, with a large standard deviation in C.

There is an apparent difference in T3 (Figure 47, left), particularly between A+B and C groups. A large standard deviation in B allows only observing a tendency between A and B. Likewise, only a tendency was found in T4, between A and B+C, as the large standard deviation in C and particularly in B do not allow proper distinctions (Figure 47, right). The results do not support the first hypothesis, suggesting a tendency towards the second.

- Overall performance X task time or performance

Three groups (A, B, C) were defined according to overall performance. Superior performance (descending line) and smaller time (ascending line) were expected in each task from the best overall performers. An apparent positive correlation was found between overall performance and performance in T1, although a large standard deviation was found in B and C groups (Figure 48, left). An apparent positive correlation was also found in T2 (A superior to C, B superior to C) with a large standard deviation on B (Figure 48, right). Results suggest that B might be superior to A and C in T1 and T2.

An apparent positive correlation was found in T3 (A superior to C, B then C) and a suggestive tendency of A over B (Figure 49, next). Differences could not be detected

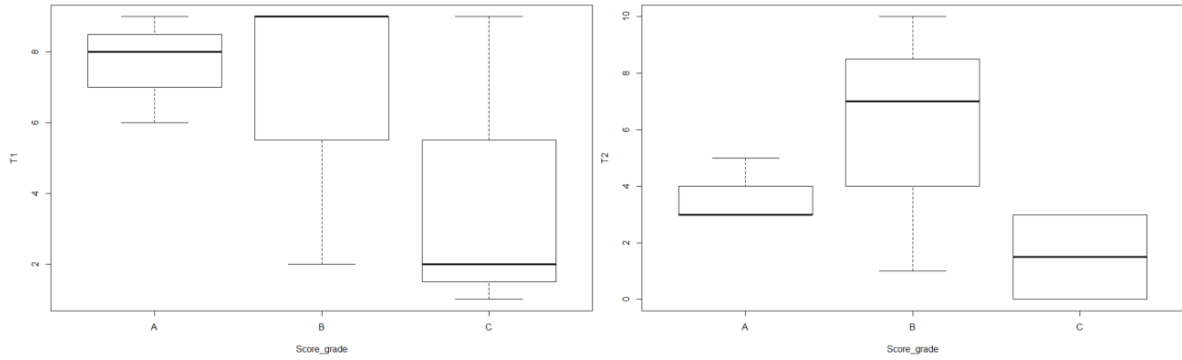


Figure 48 – Overall performance X task time on T1 (left) and T2 (right).

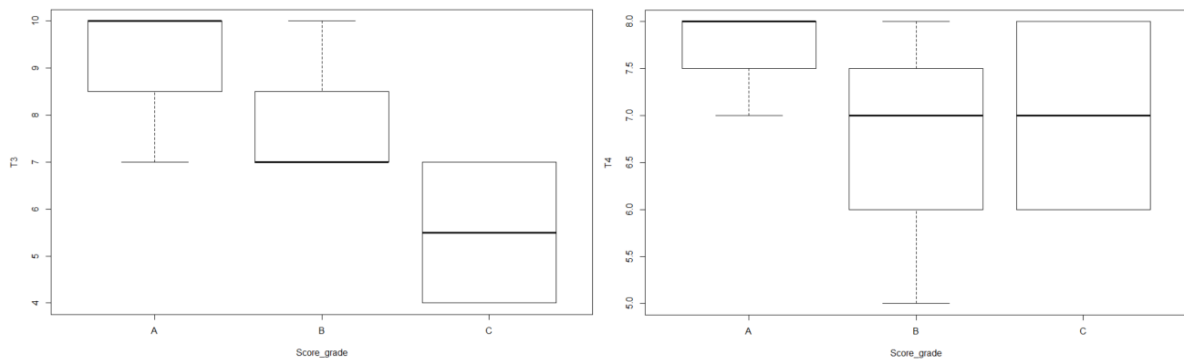


Figure 49 – Overall performance X task time on T3 (left) and T4 (right).

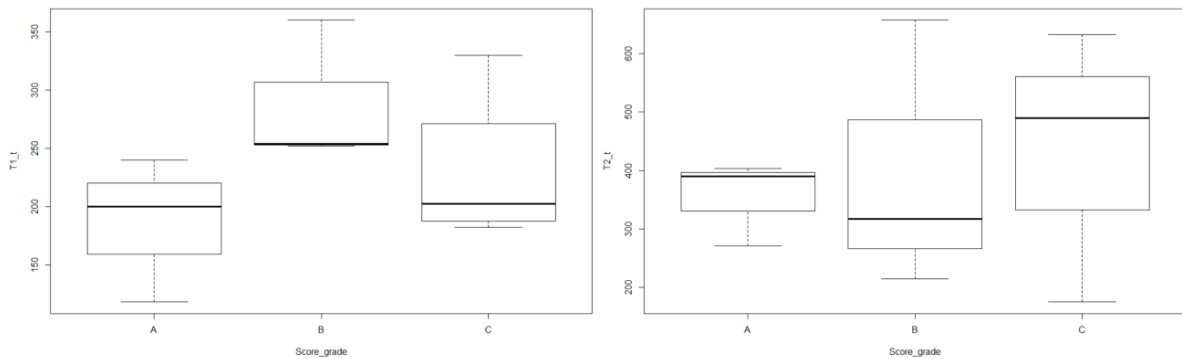


Figure 50 – Correlation between performance and time in T1 (left) and T2 (right).

in T4 (large standard deviations on B and C) despite a suggestive tendency of A over C and A over B (Figure 49, right).

An apparent positive correlation was found between overall performance and T1 time only between A and B (Figure 50, left). T2 analysis is inconclusive despite the smaller standard deviation found on A group (Figure 50, right). An apparent positive correlation was detected in T3 only between A and B+C, (Figure 51, left). An apparent positive correlation was also found in T4, between A and B, A and C and inverse between B and C (Figure 51, right).

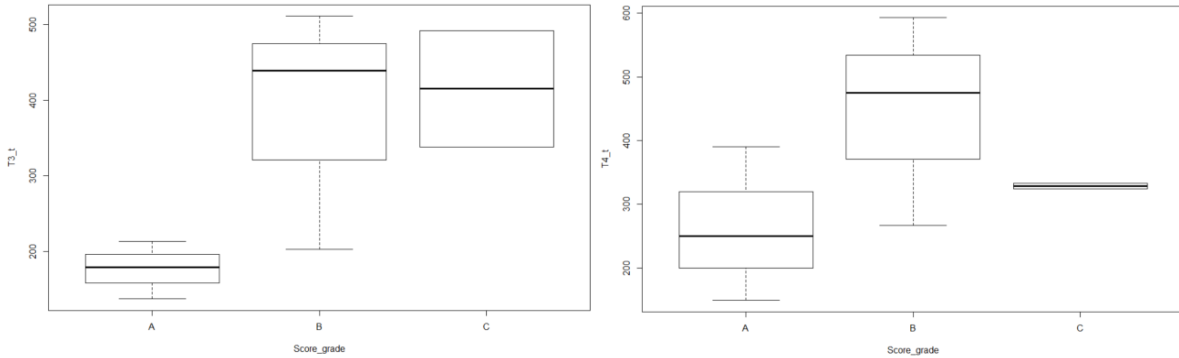


Figure 51 – Correlation between performance and time in T3 (left) and T4 (right).

#### 4.4.2.3 Results analysis and impressions

After an initial analysis, it appears that overall score ranking correlates with overall time ranking and tutorial time, i.e., top performers needed less time to complete all tasks, including the tutorial. The overall score also seems to correlate with tutorial time.

The tutorial task was designed to allow learning and some habituation with the navigation technique. Still, top performers in this task could have previous gaming experience, as the navigation is similar to FPS games, and they may also have good spatial abilities. As only the tutorial (T0) and T3 required all commands (6 DOF navigation), T0 performance probably had little influence in T1, T2, and T4 results, where navigation was straightforward. However, top T0 performers figure in top overall positions, which might indicate that T0 also depends on relevant cognitive capabilities such as spatial ability.

Students performed better in T3, followed by T4, T1, and T2. The better performance in T3 and T4 might reflect the continuous interaction, simpler tasks, and the few available response options or their subjective nature. T3 had only three response options for shape - equiaxial, stretched, or highly stretched – and the ‘stretching’ level can be deemed a subjective factor. Although three students obtained the highest possible score on T3, the few available options or task score granularity have not allowed a proper distinction between medium and high or medium and low scores, which might have biased the results towards higher scores on this task.

Finally, T4 had no right answers, i.e., clearly defined ‘heterogeneity epicenters’, and the evaluation relied on a loose definition for heterogeneity, resulting probably either in user uncertainty or subjective grading, particularly for ‘big grain’ grouping. However, considering that heterogeneity is not an established terminology or a typical property of grain structures, all students could still identify at least two correct epicenters, as stated by experienced evaluators.

Despite the similarity between T1 and T2 (visual search for grains by size), students performed better in T1 probably because bigger grains are easier to detect, and not all

smaller grains will appear on the available discrete sections. Additionally, smaller grains should be harder to detect in larger sections - four students analyzed diagonal sections in the cubic sample. The limited available data does not allow detecting significant differences in task performance from students that explored the sample in diagonal orientations. In fact, one of the top performers belongs to the diagonal group. T1 and T2 tasks had higher score granularity and evaluation accuracy when compared to T3 and T4 due to the objective scoring - scores are proportional to the position in the size ranking. For bigger grains, picking the first five grants the higher score, while the 6<sup>th</sup> up to the 15<sup>th</sup> yield diminishing values. Since smaller grains are harder to detect and will probably not appear equally to all students, a higher tolerance was allowed (higher score for ten smaller and diminishing scores from 11<sup>th</sup> to 30<sup>th</sup>).

Therefore, considering size detection tasks only, students performed better on T1 and more time was spent on T2, as expected, probably due to the necessity for longer and more detailed searching. It also appears that the fastest participants had a poor performance in T1, T2, or both, probably due to impatience, anxiety, or cognitive overload, as such tasks demand detailed or longer searching time. The preliminary analysis supports this hypothesis: the best in T2 was the slowest in this task; top performers in T2 tried to search more grains than requested, performed well in T1 and are among the top overall performers, which might indicate high persistence and motivation; good performers both in T1 and T2 took longer (average total time); poor performers in T1 (simpler search task) performed poorly also on T2 (difficult search task) and either got similar T1 and T2 scores or a higher T1. Finally, no evidence of learning effect was found on task order (T1->T2), as students who took longer on T1 did not necessarily perform better on T2.

Interestingly, the top performer was faster overall and fastest in most tasks, had top scores on T3 and T4 but only intermediary scores both in T1 and T2. Along with another top performer, this student persisted in finding more grains, showing higher engagement or curiosity, suggesting that motivation might impact performance positively. The only student who performed very well in T1 and T2 took longer overall and in most tasks, suggesting that T1 and T2 strongly rely either on visual search proficiency or a combination of patience, mental stamina, and exploration time. Top performers on T4 also spent more time in this task, which demands skills in visual searching and pattern recognition.

#### 4.4.2.4 Student feedback

The last in the score ranking had limited spatial abilities as expressed by self-declared low performance on drawing and mental rotation tasks, as well as limited navigation ability on virtual 3D environments, as expressed by little 3D gaming and no 3D modeling experience. Curiously the same student affirmed that the experience was enlightening, for typical 'superficial' images observed on microscopes does not allow having the volumetric

sense of the grains allowed by the interactive sectioning prototype. Still, the background and ability limitations as well as the confessed cognitive overload and impatience, probably lead to low performances in T1, T2, and T4.

Another student from the bottom of the score rank reported no experience in 3D gaming or modeling (navigation skills included) and poor drawing skills, which are likely related to poor performance in T1 and T2. Accordingly, only those two students had low performance both in T1 and T2.

On the other hand, one of the top overall performers reported 3D gaming experiences on console and PC platforms (including 6 DOF keyboard navigation), good performance on drawing disciplines, advanced 3D CAD training, no cognitive overload, and curiosity to perform the tasks. Curiously, the top performer in T1 and T2 was either the slowest or more patient (overall and in T0). This student reported some gaming experience, no CAD training, and poor drawing knowledge. However, the self-declared persistence and curiosity probably contributed to the good performance on T1 and particularly to the top score on T2. The intense effort probably contributed to the high reported cognitive load and impatience upon the execution of the last two tasks. The alleged motivation was the curiosity to ‘discover how grains really look like’.

#### 4.4.2.5 Preliminary user study final remarks

It is argued that dynamic real-time sectioning can deliver more in cognition of grain structures than static sections, videos and animations, or conventional 3D visualizations, given the possibility to assess fundamental characteristics such as grain size, shape, neighbors, and sample-wide heterogeneities.

The prototype allowed users to visualize ‘infinite’ sections within the virtual material sample, dynamically setting up any arbitrary combination of position and angle. Students found the tool compelling, accessible, and simple to use and learn. Individual grain size and shape could be detected by translation and rotation movements, while smooth interactive navigation along the sample allowed students to get a good sense of the structure and to discover spatial heterogeneities such as big or small grain clusters. User feedback and T4 results showed that students could properly understand the concept of heterogeneity, detect spatial neighborhood relationships, and inspect the surroundings of a given grain using interactive sectioning.

Although they had limited knowledge of grain structures, most agreed that interactive sectioning provided a unique spatial experience of grain structures, given available educational media. It is clear that the interaction leverages the qualities of traditional sections and surpasses limitations associated with their static nature that can easily mislead students, i.e., a single section tells little about the properties of any single grain but real-time sectioning can tell much about the geometric properties of the whole sample.

In fact, students' comments and T2 results reinforced the inherent limitations of discrete sections, particularly in identifying and characterizing smaller grains because they tend not to appear in any section. Most students complained about the discretization of sections in T1 and T2, affirming that finer continuous movements would allow for better recognition of smaller grains and their shapes.

As previously stated, spatial references were purposely left out of this preliminary study. Almost all students suggested including spatial referencing systems such as a miniature or global map, Cartesian axis, and orientation cube, to name a few. Thus, as predicted, the absence of spatial references contributed to difficulties in locating inside the sample or estimating rotations. The introduction of such structural elements could probably reduce the reported cognitive overload and support students with limited spatial abilities.

Study results, along with student feedback, indicate the following improvement points for subsequent user studies:

- navigation should be assessed before and after the tutorial in order to properly account for individual variations of navigation skills in test results;
- scores should be measured with more details: size ranking was not enough and the scoring system was arbitrarily chosen;
- previous similar experiences, as well as spatial ability, should be properly assessed with other methods;
- shape factor classes should present unambiguous and clear options to allow automating scoring procedures and reducing or eliminating evaluation subjectivity;
- all tasks may feature motivational elements, as engaged participants are likely to perform at their fullest potential. The use of VR technology (3D vision and natural interactions), high-quality graphics features (such as depth enhancing 2.5D), and gamified task elements will probably contribute to this matter. Such features might also impact learning in educational settings, as previously discussed;
- spatial references (position and rotation) will probably impact performance and cognitive load, especially for students with lower spatial abilities, making tasks easier and allowing longer experiments;
- cognitive overload might be further reduced with clean, minimalist and easy menus, along with simpler and intuitive controls, both in desktop and VR versions.

Concerning preliminary hypothesis assessment, the study indicates that the merger of proposed features allows extracting more information and improves spatial comprehension of grain structures, suggesting an increased learning value over single sections, discrete

sequences, and previous interactive attempts such as [Lanzagorta et al. \(1998\)](#). The preliminary user study also suggests that the designed tasks are effective enough to assess the proposed factors, as students were able to perform all of them in a feasible time, and resulting response distributions follow known functions. However, reported cognitive overload might indicate that each task or the entire procedure is too long, which could be reduced by introducing spatial references and more intuitive commands and menus. Therefore, the identified trends indicate that prototype options combining visual depth enhancements (2.5D), upgraded spatial reference systems, intuitive or natural commands and stereoscopic vision (see next section) are likely to perform better (task time and performance) than the first 2D prototype.

## 4.5 Second preliminary experiment

The first preliminary experiment provided valuable feedback, but it has not yet compared visualization modes nor techniques, so the next experiment design should approach the complexity level expected for the main experiments. Therefore a second and larger experiment was planned, comprising about 30 participants, with the following goals:

- apply a mixed design with between and within-subjects factors;
- assess distinct scoring possibilities;
- improve upon pre and post-experiment tests;
- inspect response patterns and apply statistical treatments;
- assess methods for online data collection;
- assess consequences of all changes on the experimental logistics;
- further investigate the study's hypothesis.

Although still limited to sectioning techniques, this experiment introduced and compared two 2.5D visualization modes as a between-subjects factor and color modes as a within-subjects factor, keeping the experiment in a complexity level appropriate to the 30 available participants.

### 4.5.1 Discussing and detailing 2.5D modes

It is argued that 2.5D sectioning allows faster and better recognition of grain size and shape; however, standard shading schemes do not express the expected natural lighting behavior. Therefore it is reasonable to assess performance differences between shading modes.



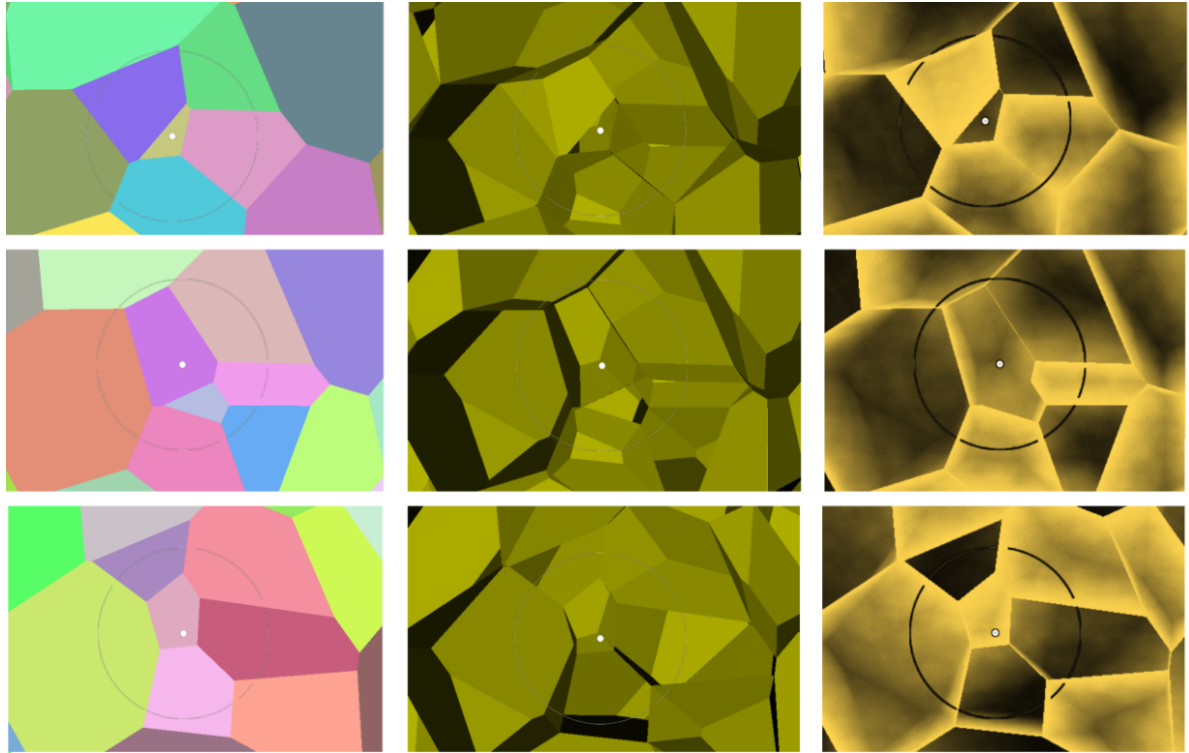


Figure 52 – Visualizing a grain (white dot) with 2D and 2.5D modes. 1<sup>st</sup> row: entering the grain, clearly detectable with enhanced 2.5D (right). 2<sup>nd</sup> row: larger section of this grain, depth differences are noted. 3<sup>rd</sup> row: the end of the grain is detectable with 2.5D (center) and unambiguously detected with enhanced 2.5D (right).

As previously discussed, 2.5D modes are supposed to excel flat 2D sectioning on size perception, as the user can always see the remaining portion of the sectioned grains (Figure 52) and can determine whether the section is positioned at the beginning or the end of a grain and to guess faster and accurately where the larger section might be. It is argued that, although individual sizes are easily perceivable with the ‘shadeless’ standard method, size differences or clustering (groups of grains with similar size) can only be perceived with improved depth perception. In this case, shadeless refers to the typical shading algorithm that only takes into account the angle of the light source, ignoring any surface or object in between, consequently not reproducing the dimming in light that should normally occur.

This example (Figure 52) illustrate this point showing that, given a grain visualized with a flat 2D section, one cannot tell whether it is a small grain or an extremity of a larger grain. In this case, it is a medium-sized grain, and conventional 2.5D may allow one to distinguish extremities, although the shading method may induce visual confusion. It is evident that the enhanced 2.5D shading method allows a clearer distinction, at least on static images. However, this advantage must be assessed in interactive scenarios.

Detecting clusters requires spotting size differences not only between single grains but between regions. Once again, flat 2D sectioning might mislead the observer when most

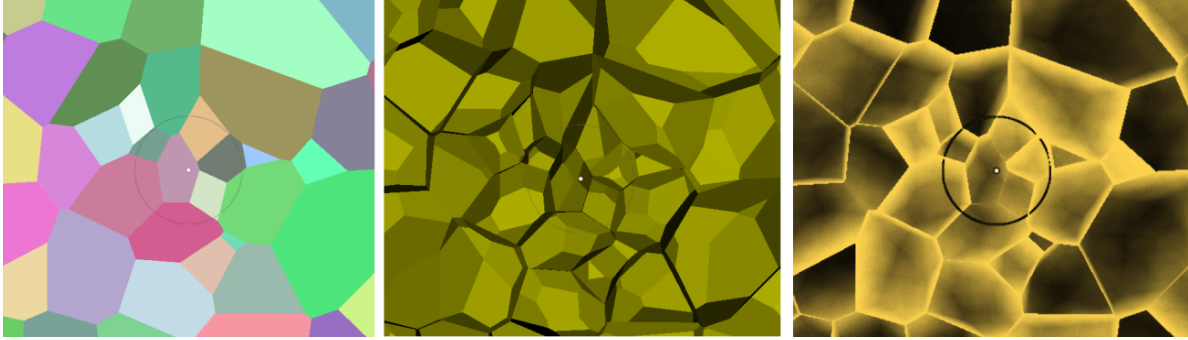


Figure 53 – Cluster detection. Flat 2D: the user cannot tell the difference between small grains or the beginning of bigger ones (left). 2.5D allows seeing the end and guessing that a cluster might be present (center). Depth-enhanced 2.5D shows clearly that adjacent grains have similar depths (right).

contiguous grains are at the beginning of the current view. Although 2.5D allows resolving this uncertainty, depth-enhanced 2.5D will also show if neighboring grains have similar depths (Figure 53). Based on fundamental concepts from visual search theories, discussed in section 2.2, the expected detection advantage of depth-enhancements may lie on its capacity to simply and clearly highlight features of interest with a single characteristic: shading tone differences. Bigger grains should appear as darker holes and their clusters as darker regions; smaller grains should appear as brighter small spots while their clusters as brighter regions. In this sense, the inconsistent shading of standard 2.5D rendering and the random colors of flat 2D act as distractors, requiring scanning all the visual field actively, which happens to be slow in a large 3D search space, as proper scanning demands movement. This also indicates why standard 2.5D should perform better than flat 2D, as the former reduces the amount of movement necessary to scan a sector. One should remind that although the examples given here are illustrated with static images, the user can move and analyze the sample dynamically. Still, what is being discussed at this point is how efficiently each visualization mode encodes such high information density. The difference is clear considering static images and the available theory suggests that the same is probably true for interactive sectioning, but only the user study will show whether this holds.

#### 4.5.2 Color

Another relevant issue is color in enhanced 2.5D modes. In order to emphasize shape on study models, designers typically adopt a single brighter tone, increasing the contrast between lighted and shadowed areas. Adopting multiple colors interferes with shadow tone, affecting particularly darker colors, hindering proper perception of shape and depth. In theory, any bright tone can accomplish this effect, but it is a good practice to avoid brighter colors in electronic displays, as they may not look well on every device or can even disappear depending on display settings. The chosen color for enhanced 2.5D is a sepia tone, a classic color used in drawing and particularly in photography due to the



Figure 54 – Color palettes obtained with the HSL color model. The brighter palette (top) was achieved with 95% saturation and 85% on lightness while the bottom darker palette adopted 100% and 20% respectively.

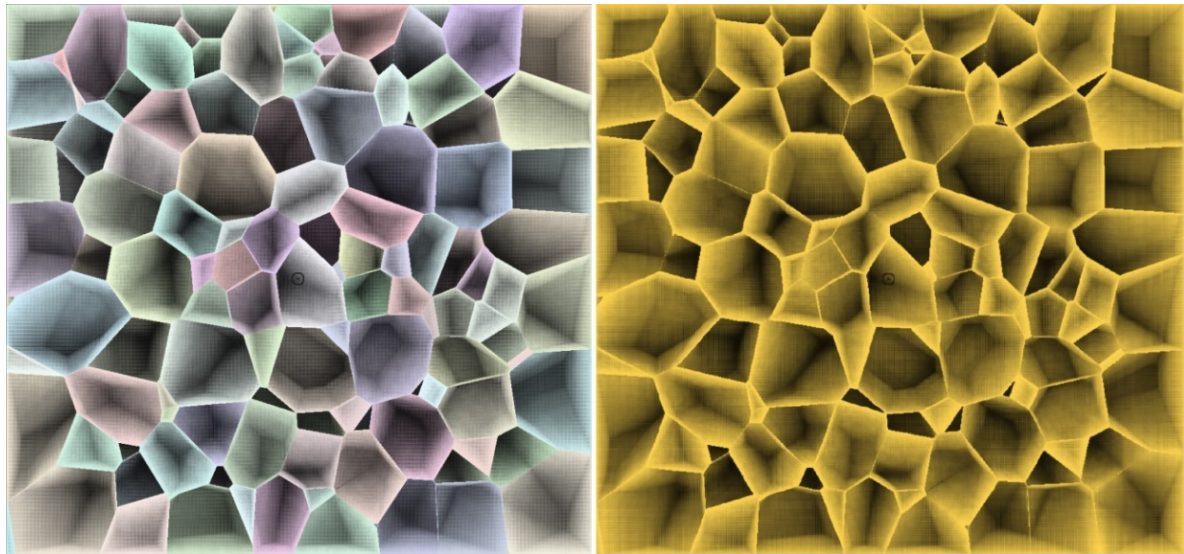


Figure 55 – Color in 2.5D depth enhanced mode. Although the depth perception seems inferior using multiple colors, note that the perception of darker spots (‘holes’) is well preserved compared to a single sepia tone.

alleged superior visual comfort. The ‘warmth’ feeling typically attributed to this color is due to the resemblance to sunlight, as warmer colors are related to relaxation. As noted in section 3.3.2, the sepia tone was tuned to reduce ghosting issues in anaglyph 3D. However, as already noted, although color may disturb optimal shadow and depth perception, it plays an essential role as a mnemonic aid, which may help to perform recognition and counting tasks and to build stronger spatial maps.

Therefore, colored 2.5D should be assessed, and brighter colors favored. The HSL (hue, saturation, lightness) model maps RGB in a cylindrical coordinate system, based on relevant perception criteria. Conveniently, by fixing saturation and lightness values, a perceptually uniform color palette is obtained (Figure 54). A bright color palette was easily created (Figure 54, top) by setting saturation at 95% and lightness at 85%. Three.js allows setting colors using HSL and each parameter ranges from 0 to 1, resulting in 0.95 saturation and 0.85 lightness for the desired palette. The Hue parameter range was then equally divided by the total grain count, obtaining a unique color for each grain, theoretically. Finally, each grain was randomly assigned one of these colors to avoid equally colored neighbors (Figure 55, left).

### 4.5.3 Experiment design - 2.5D modes and color

This procedure intends to assess feature detection differences between two shading modes on ‘2.5D’ visualization as well as the influence of color. According to the preceding analysis, it is argued that enhanced 2.5D will outperform 2.5D, which will, in turn, surpass flat 2D for grain size and cluster detection. The test will focus on tasks where these detections are more challenging, i.e., finding smaller grains and small grain clusters. Both tasks are described in section 4.3. Participants will choose at least six grains in each task. Half the group will experience conventional 2.5D and the rest depth-enhanced 2.5D (between-subjects factor), where each participant will test single and multiple colors (within-subjects factor) in a mixed-effects experimental design. As color is a secondary factor here, it will be tested as a within-subject variable, allowing studying more phenomena with the same number of participants, increasing the reach of the experiment (MARTIN, 2007). Another difference from the preliminary test (section 4.3) is that continuous movement will be allowed instead of discrete sections, and the user will use two arrows (front and back) to navigate.

A participant profile questionnaire and a navigation and task tutorial will be administered remotely (see description in section 4.6.8) some days before the main study, and visualization mode order will be fully counterbalanced, demanding multiples of 2 participants. Although not that relevant, task order can be easily counterbalanced with multiples of two participants. At the end of the experiment, a questionnaire will be administered to check for effort and preference on colors and visualization mode.

Three scores were collected and calculated in this experiment: choice time, count of choices within the 30 smaller (Top 30), the percentage of choices within five strata (smaller of the sample, top 5 smaller, top 20, top 70, top 250), and a factor calculated after a linear descending score scheme, which gives top score to the smallest and the minimal score after the 30<sup>th</sup>.

### 4.5.4 Results

The collective experiment took less than an hour in total, with 30 minutes of active work on the tasks, as part of a graduate Materials Science course class. Despite the initial intent to execute the experiment in a computer lab, it was agreed that the best option was using students’ laptops. All accepted participants filled preparatory questionnaires and performed tutorials at home using online forms, as a prerequisite to taking part in the experiment - a data collection system allowed to collect the data generated in the tutorial successfully. Upon the end of the tasks, they filled an online preference questionnaire. It is important to note that the experiments performed in this work were approved by an ethics committee (project code CAAE - 04436918.8.0000.5390).

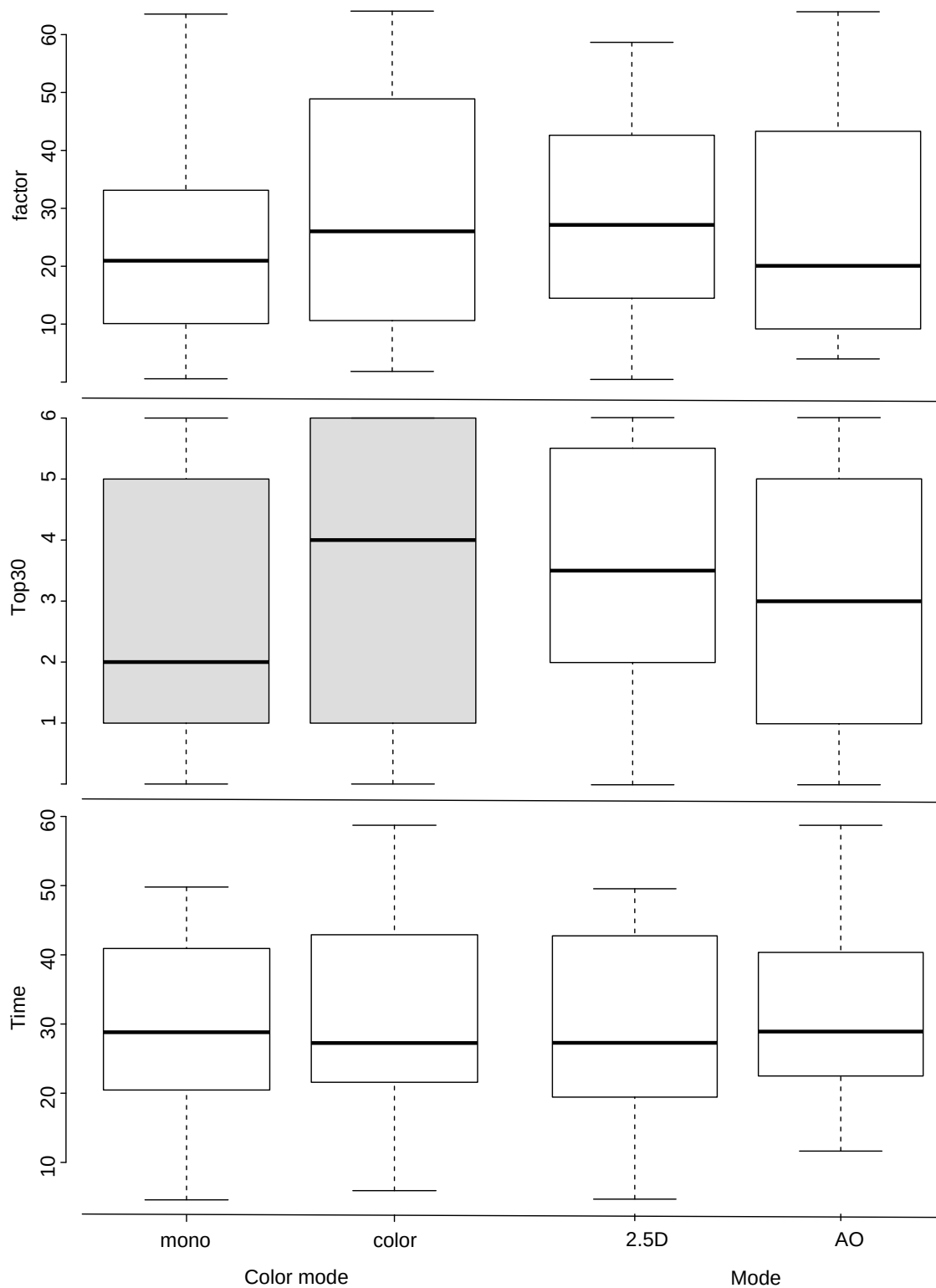


Figure 56 – Size task: boxplots of Time, Top30 and Factor responses per color mode and visualization mode. A significant difference was found in color mode on Top 30 (highlighted in grey).

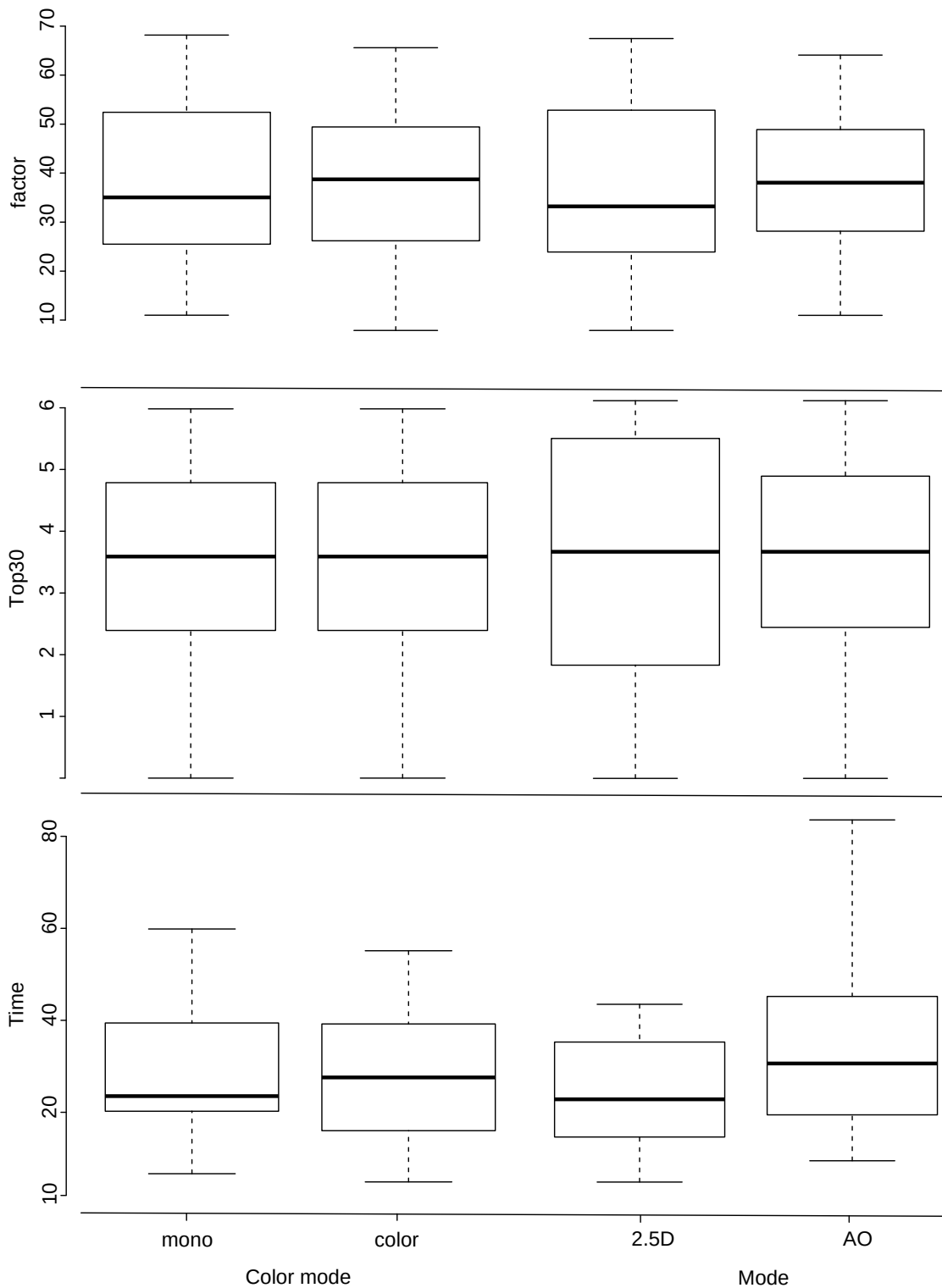


Figure 57 – HR: boxplots of Time, Top30 and Factor responses per color mode and visualization mode. No significant difference was found.

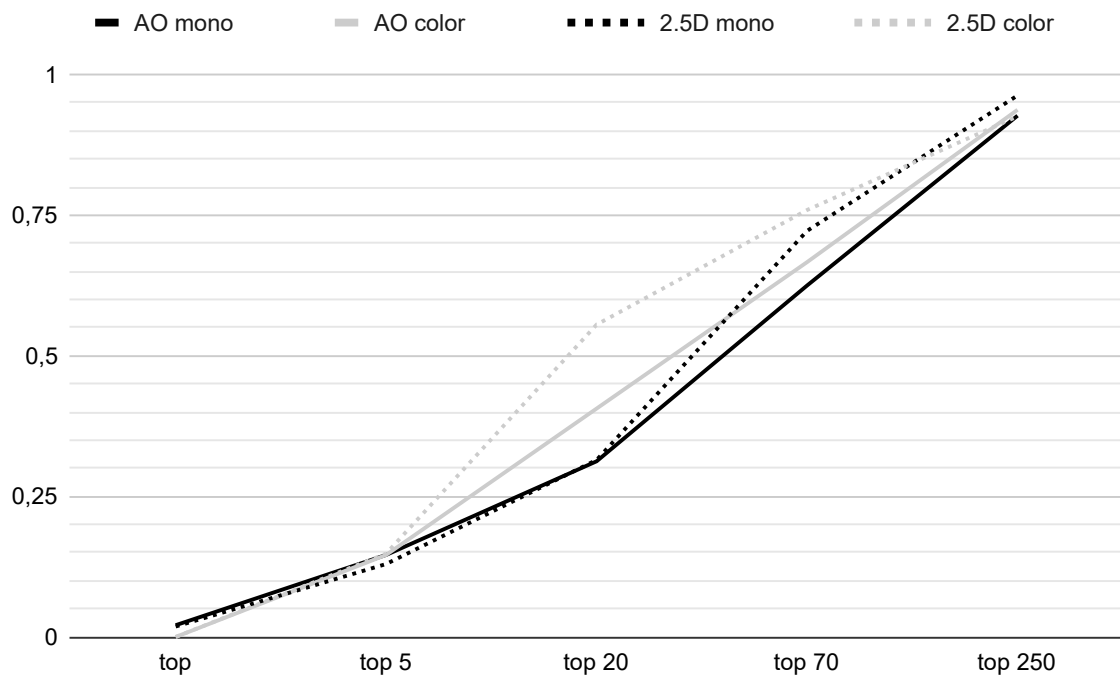


Figure 58 – Size task: percentage of choices within five size ranking strata (smaller of the sample, top 5 smaller, top 20, top 70, top 250).

No distribution conformed to parametric ANOVA assumptions, nor conformed to any known distributions; therefore, all scores were analyzed using the non-parametric Aligned Rank Transform procedure (ART). Some options to improve distributions' profiles are increasing choices per participant, assigning more participants to each visualization mode or changing this factor type to within-subjects. An increase in the quantity of collected data may help find the real character of response's distributions, ease finding the appropriate ANOVA method for statistical analysis, and find more significant differences between visualization modes if they exist.

The broader selection of scores increased analytic possibilities, and a significant effect was found. The Top 30 score was significant for color on the Size task (Figure 56) with a difference of 26 % in favor of color. Further, analysis of choices per strata (Figure 58) allowed to attribute the difference found to the colored 2.5D mode, due to the substantially superior performance at top 20 (77%) and top 70 (22%) strata. AO color was only marginally better in the same strata, suggesting an interaction between visualization and color. Thus, other scores may help analyze the phenomenon through distinct points of view and make deeper distinctions, escaping the reliance on statistical significance alone.

Finally, participants expressed strong preferences for color (83% versus Mono at 17%) and 2.5D modes (76% versus Flat 2D at 14%) considering aesthetics and self-assessed detection performance. No significant differences were found in HR tasks (Figure 57).

Experiment logistics ran smoothly and without major problems. A new online data collection mechanism worked properly and robustly, even with more than 30 participants feeding data concurrently. However, it is necessary to collect more data to assess responses properly. Still, other scoring mechanisms allowed finding differences in the Color mode response. Overall, this experiment allowed the assessment of a more complex mixed design and detailing logistics requirements.

Regarding pre and post-tests, more detailed and established spatial ability tests are still necessary. After a brief survey of available questionnaires, the Vandenberg Mental Rotation Test, or MRT (VANDENBERG; KUSE, 1978), and the Santa Barbara Sense of Direction, or SBSOD (HEGARTY et al., 2002), were considered appropriate. The MRT intends to assess the participant capacity to process geometric operations, such as rotating objects. The SBSOD, on the other hand, is used as a measure of spatial ability on the broader environmental scale. Such tests will improve the assessment of proficiency in core skills required to complete the tasks successfully. The results are typically used in studies for an in-depth analysis of the main experiment's outcomes, either strengthening observed tendencies and objective results or supporting the resolution of exceptions, conflicts, and unexpected findings. After the experiment, the application of the well established NASA Task Load Index (TLX) (HART; STAVELAND, 1988) will improve the evaluation of the cognitive load associated with each task.

## 4.6 Main study design

After two preliminary experiments, it is possible to design a robust and safe experiment, considering several logistics and experimental requirements. It is estimated that 60 participants are adequate to investigate techniques, visualization, and color modes in a factorial mixed-design experiment. Ideally, more data yields more reliable results, and so far, there is no reason to rule out within-subjects designs except for extra time consumption on the part of participants.

The logistics of pre and post-experiment questionnaires, remote tutorials, and data collection are well resolved. However, MRT and SBSOD questionnaires will substitute previous pre-tests while TLX was introduced as post-test. Among visualization modes, 2D and 2.5D have been examined on preliminary user studies. Thus, 3D modes remain to be assessed for the first time and may present logistic challenges. Anaglyph 3D is executed on conventional 3D desktops, so it does not bring novel logistic issues. 3D split-screen will demand scheduled single-user sessions for each participant, which only increases experimental time requirements.

A few other issues were not yet detailed or assessed in experiments and deserve more discussion and planning before the main experiment's design.



### 4.6.1 3D anaglyph issues

Anaglyph 3D is a relatively accessible visualization option, as simple anaglyph goggles are very cheap and do not require special displays. Thus, conventional computers can be used in larger experiments. However, good performance depends largely on the quality of the goggles and display calibration. However, it is a known fact that anaglyph 3D is not universally comfortable and that some users cannot perceive the 3D effect, mostly due to ghosting issues and heterogeneous filtering quality across the color spectra (LIVATINO et al., 2015). The ghosting issue, already introduced in section 3.3.2, which influenced the choice for the specific sepia tone shown in some illustrations along this document. However, if colored depth enhanced 2.5D is adopted (Figure 55, left), the strategy to use an optimized single color will no longer be possible. The technique also incurs into light losses, as the goggles filter a significant portion of the light. An advantage is that the field of regard is preserved when compared to other 2D modes.

As hardware variability is expected and end-users will not usually bother optimizing monitor configurations, the experiment should not attempt to control ghosting or color matching, assessing the overall performance while preserving such inherent technology issues (i.e., as random variables).

### 4.6.2 Influences of HMD parameters on 3D split-screen visualization

3D split screen stereoscopy technology presents several interactive and visual variables to be considered, such as latency, interaction range (body and hands), precision and sensor interference, field of view, resolution, availability, and cost. Low latency is a requirement for proper interaction as inadequate latency is associated with VR sickness (MEEHAN et al., 2003). Interferences and tight sensor ranges will probably interfere in task time and performance, as tracking disturbances or glitches consumes time and interrupts the task flow. Although current HMD sets (Figure 59) have similar resolutions, some institutions might possess only older devices. Modern high-end smartphones have a superior resolution but may lack enough graphics and processing power to assure low latency. Some devices such as Hololens, albeit advanced, present low resolution and a very restricted field of view. However, tasks can be adjusted to rule out field of view restrictions.

Therefore available HMDs (Oculus Rift, HTC Vive, Microsoft Hololens, high-end smartphones or Gear VR) should be assessed more thoroughly, at least qualitatively, looking for possible impacts in task time and performance of factors such as precision, interference, user discomfort (including latency) and resolution. The following table (Table 1) shows relevant properties of split-screen 3D devices. As already mentioned, technology choices must be based not only on performance factors but also on accessibility - devices should not restrict proper task execution and should be available to students.



Figure 59 – Advanced HMD systems: Oculus Rift CV1, HTC Vive and Microsoft HoloLens.

Table 1 – HMD comparison table (09/2017).He:Head, Ha:Hand, W:walk, G:gestures, B:buttons

	Rift CV1	Vive	HoloLens	Rift DK2	Gear VR	Cardboard
F.O.V.	110	110	30	100	100	110
Resolution	2160x1200	2160x1200	HD/eye	Full HD	1960x1440	up to 4k
Size	3.61"/eye	3.61"/eye	Small	5.7"	5.1-6"	5-6"
Refresh Rate	90	90	60	75	60	60
Range	<3m	4.6x4.6m	5m	<3m	NA	NA
Precision	High	High	Medium	High	NA	NA
Interaction	he,ha,w	he,ha,w	he,w,g	he	he,b	he,b
Weight	470g	550g	580g	440g	370g	400g
Cost (US)	600	800	3000	NA(used)	750	350

DK2 is inferior in FOV, resolution, sensor ranges, and possible interactions. HoloLens has several limitations, especially low FOV (only 30° horizontal), and its main strengths (see-through visor, range, gestures) are not relevant to the tasks. Therefore, remains the comparison between Cardboard and Vive, where the differences lie in FOV, resolution, refresh rate, and sensor adequacy (range and precision). Concerning accessibility, the most affordable options are smartphones equipped with Cardboard frames and Gear VR devices. However, Gear VR works only with a few high-end devices of a specific brand. Therefore, the most accessible option is the Cardboard. As shown in the table, both options lack position tracking. However, interaction is limited in the experimental system and is not investigated in this work.

Another relevant factor is the system's current WebGL platform, designed to be accessible and device-independent, running on most modern browsers. However, as Web VR technology is still in its infancy, the system does not run properly on VR headsets yet, and the development of a new version for HMDs was deemed unpractical. Web VR will allow practical platform unification in the near future, allowing the use of advanced VR devices such as Oculus Rift, HTC Vive, and HoloLens (Figure 59). Therefore, a VR adapter equipped with a high-resolution smartphone was considered adequate to perform user studies in 3D versions of the system. In this particular setup, a Bluetooth keyboard was used as input method.

### 4.6.3 List of independent variables

The technique factor has three levels:

- Benchmark technique: simulated discrete 2D sections sequences (S);
- Proposed tool 1: dynamic sectioning (DS);
- Proposed tool 2: GrainCrawler (GC).

Visualization factor comprises four levels:

- 2D (monitor, flat shading) – not applicable to GC;
- 2.5D (monitor, conventional or depth enhancing rendering);
- 3D anaglyph (monitor, 2.5D rendering);
- Immersive split-screen 3D (only 2 tests in 2D interaction).

The color factor has two levels and applies only to size and HR tasks on DS and size tasks on GC:

- Mono (sepia or white) – DS and GC only;
- Bright HSL color palette (see section 4.5.2) - sole level on S.

Thus, the number of factorial combinations per technique are:

- S – 1;
- DS – 8;
- GC - 6.

### 4.6.4 Discussion of experimental design possibilities

Given the factor combinations outlined above and considering at least 20 people in between-subjects groups, 160 participants would be necessary for an ideal between-subjects (BS) factorial study. Considering budget limits and the difficulty of mobilizing so many people, an optimized study design was pursued to reduce participant requirements and tasks per participant while maximizing data quality. A typical technique is changing the type of a secondary factor to within-subjects (WS)([MARTIN, 2007](#)). The preliminary experiments indicated about 30 participants per condition.

Turning visualization into a WS factor reduced the required number of participants and procedure length since there are only three techniques, requiring about 90 participants. In this design, each participant would perform two tasks on S, eight tasks on DS, and six on GC. Preliminary experiments have shown that about seven minutes are required to complete each task, resulting in 14 (S), 56 (DS), and 42 minutes (GC). Good practice (MARTIN, 2007) indicates that a section should last no more than one hour, and longer sections should be split. It is known that researchers usually rely on university students as a source of participants for user studies in exchange for course credits. This approach allows for long term commitment, as courses typically run from 4 to 6 months. However, Brazilian law do not allow paying experiment participants in any way, requiring them to be volunteers. Therefore, even if the experiment fits as a natural part of the course, the activity cannot be mandatory, and an alternative activity should be offered. In this case, grades for experiment involvement should only be assigned based on correct participation. This approach will be pursued in this work, trying to achieve about 30 participants per experimental group in two offerings of a Virtual Reality course.

For practical reasons and given the similarity between S and DS, they will be tested by the same 30 participants in two 45 minutes collective sessions at a computer laboratory, decreasing the number of participants to 60. According to the post questionnaire, the second preliminary experiment ran for a total of one hour and 30 minutes of active work, with acceptable working loads. On the other hand, GC will be tested by another group of 30 in one collective session for 2D and 3Da modes. HMD 3Dst mode (DS and GC) will be tested individually at scheduled sessions. A complete counterbalancing scheme was employed on the GC group and a 4x4 latin square on DS/S group. Turning technique into a between-subjects factor will also help reduce cognitive load since each participant will only have to learn one type of interaction, avoiding inducing users to confound commands when transitioning from one platform to another. Finally, this design pattern allows a superior assessment of differences between the levels of the technique factor.

Tutorials and navigation testing will be performed remotely following the logistics of preliminary experiments, reducing total session time. Keyboard navigation should be somewhat familiar to most participants, although this interaction can be fully trained and evaluated in remote home sessions since all users will probably have access to a computer. In fact, 2D, 2.5D, and 3D anaglyph versions could be fully evaluated in remote sessions.

## 4.6.5 Details of experimental techniques

### 4.6.5.1 Benchmark technique: discrete sections

As previously discussed, the discrete section technique (S) will be assessed only on size and HR tasks, in a single visualization option (2D), in order to simulate real sections. Interaction is straightforward, adopting only two buttons to go back or forward through

the discrete sections. Thus, one tutorial is necessary to train size perception of large/small grains and another to capture the concept of heterogeneous regions. The first session comprises a size tutorial, followed by size task and a brief 5-minute rest. The HR task session follows the same scheme. Each session is expected to last from 10 to 15 minutes and participants can execute this version (S) remotely. Since this technique comes in a single version (2D), it is more practical to be executed by a participant from the DS group, instead of creating a third experimental group with a small workload.

Color was considered a secondary factor and was not explicitly defined in the experimental scheme. Therefore, color levels will be equally assigned on a random basis to participants on size and HR tasks. Only colored modes are considered relevant for testing this technique as a single color does not allow visual distinction between grains in a 2D section.

In order to inhibit participants from learning the model as tasks are performed, a distinct viewing angle will be assigned to each task. Since there is rigorously only three unique views (one for X, Y, and Z), where the opposite side is just spatially mirrored, straightforward 3D mirror and rotation operations can be applied to each standard view to achieve substantially distinct variations, in terms of visual perception. Then, the variations produced can be assigned to different sessions and distinct tasks, reducing the likelihood of recalling the model and the possible advantage, if any recall eventually occurs. Based on the rationale behind spatial ability tests, good variations should succeed in breaking standard spatial relationships such as left-right, up-down, and back-front. Examples of good variations are 180° rotations and 90° rotations with vertical mirroring (Figure 60, upper row). Changing grain colors makes even more difficult to detect such transformations (Figure 60, bottom row).

Although such operations change the model aspect substantially, the topological aspects are kept intact. This procedure also reduces possible biases introduced by different models and makes for easier comparison of the results. Whatever the case, it must be guaranteed that each viewing axis yields an approximately equal section area, model depth (number of sections in the case of discrete sections), and feature availability (quantity of smaller or larger grains, and heterogeneities). Finally, in order to avoid mutual help in open class experiments, participants should also be optimally distributed considering the individual task order.

#### 4.6.5.2 Technique 2: interactive sectioning

Technique S essentially differs from interactive sectioning (DS) in the distance between ‘sections,’ which results in larger gaps over 24 sections in S, and a smoother traveling through the sample in DS. Interactions are also slightly more complex, and the possibility to rotate allows the execution of the shape factor task, which will follow the same scheme

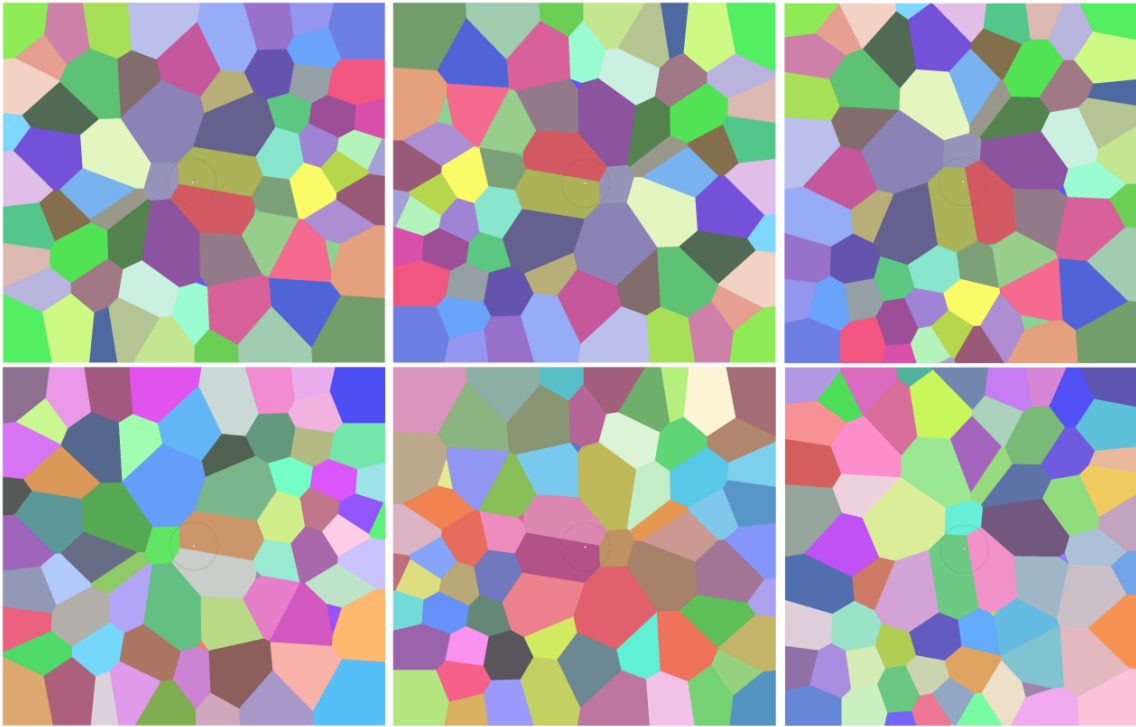


Figure 60 – Examples of 3D geometric transformations on the model. Standard model, front face (left images); 180° rotation (center); 90° rotation and mirroring on the vertical plane (right). On the first row only geometric changes are shown, while on the second row colors were altered as well.

of the others, i.e., a tutorial followed by a task. The tutorial will inform the user about the interaction technique and show examples of each of the four shape factors.

Following the evaluation guidelines, each participant should be exposed to a single interaction and close to three tasks per session, allowing up to two visualization options per day in two sessions. A very comfortable pace for a university student enrolled in a course is performing one session per week. This approach may also increase the chances of taking part in the study, as the workload is diluted across several weeks. Another incentive is allowing home sessions.

Except for split-screen 3D, which demands scheduled individual sessions on VR devices, the experiment was planned for home sessions. The optimal experiment design for this group is executing one version per week along a month. The S technique was included in this group, as both visualization and interaction are very similar, totaling five weeks. In order to avoid order effects, the sequence may follow a 5x5 latin square, requiring multiples of 10 participants, which is satisfied by a group of 30 participants. Task order can also be counterbalanced through the sessions, which can be accomplished in five sessions.

Color levels will be assigned on a random basis to participants on size and HR tasks. The two levels will be tested, except on 2D mode, where only colored modes are relevant.

Following the procedure described in the previous section, a unique viewing axis will

be assigned to each try of size and HR tasks to avoid model learning as the user progresses through tasks. As only eight distinct starting views are necessary for this technique (two tasks and four variations per participant), the transformation operations illustrated in Figure 60 meet the demand, as nine perceptually distinct conditions can be achieved. Finally, the grain shape task simply requires selecting a few grains of each shape factor to be analyzed.

#### 4.6.5.3 Technique 3: GrainCrawler

The experiment design for this technique follows closely the scheme defined for testing Interactive Sectioning. However, 2D flat rendering is not pertinent here. Therefore, the variations are:

1. PC 2.5D – remote session
2. PC 3D anaglyph – remote session
3. HMD 3D - scheduled in-person session

The experiment design for this group requires one test per week for only three weeks, presenting a reduced workload compared to the DS group. In order to avoid order effects, the sequence will follow a complete balancing scheme, requiring multiples of 6 participants, satisfied once again by a group with 30 volunteers. Task order can be counterbalanced through the sessions, which can be accomplished in three sessions with a 3x3 latin square.

Color levels will be equally assigned on a random basis to participants only on size tasks. This technique employs two distinct visual representations of grains, according to the distance from the observer. However, color is only applied to the wireframe representations, while background grains are always rendered in an opaque grey color.

Movement on GC was also restricted to a single axis to allow comparisons to the other techniques. However, the first-person view reduced the field of regard, and lateral movements are necessary to cover the entire sample. This was accomplished by expanding the navigation to the other two degrees of freedom, using six keyboard keys. Since rotational movement is very limited, allowed only in slow speeds for temporary inspection of a given region on HR tasks, the axis-aligned approaching direction (perpendicular to a given face) suffices to create perceptually distinct model variations, in order to avoid learning the model. The evaluation of this technique demands only six unique views (three variations and two tasks). One must note that GC experimental group is distinct from DS/S, i.e., not sharing the same participants and, therefore, the same grain model can be used on both experiments and the same nine unique approaching views.

## 4.6.6 Consolidated experimental procedure

An ethics committee approved user studies (check the original approval document on Annex A). Participants filled the consent form before experimental procedures started (see the original document on Annex B).

### 4.6.6.1 Pre and post tests

Three pre-tests will be applied, two of them assessing aspects of spatial ability and a graded tutorial. Spatial ability tests such as the ‘Purdue visualization of rotations’ test (BODNER; GUAY, 1997) intend to assess the participant capacity to process geometric operations, such as rotating objects mentally, which seems particularly relevant for the shape analysis task. The Vandenberg Test of Rotations (VANDENBERG; KUSE, 1978) was considered the most appropriate of such tests (check the original form on Annex C). Other tests such as the SBSOD (Santa Barbara Sense of Direction) (HEGARTY et al., 2002), in turn, check spatial awareness in a broader environmental scale, which seems more pertinent to size and HR tasks (see the original form on Annex D). MRT and SBSOD were implemented as online forms and will be applied once per experimental group.

The graded tutorial (see next section) intends to reveal individual difficulties on each task, train the participant in the relevant navigation method, and ensure compliance to a performance standard, which depends on the task. This step reduces the impact on the results caused by the learning curve of each navigation method. Objective measurements of attempts and total time obtained in this tutorial also allow detecting participants with more difficulty and comparing the learning curve of different techniques. Then, participants shall perform a perception task, where the standard for grain size, shape and heterogeneous regions is presented, showing examples of very small or very large grains, the four shape types and, finally, large and small grain heterogeneous regions. Tutorial and experimental tasks are very similar - differences are the feedback about their choices and the adoption of a distinct grain model. The participant will be able to experience a similar sample, and to distinguish size extremes, being trained to capture size differences properly. Both tutorials will be executed before each experimental condition and accessed through specific web addresses, provided every week.

This opportunity will again reduce the learning curve impacts on experiment results, as every participant will already possess a good sense of what to look for (size perception) and how to perform close to the individual optima (search strategy). The expected overall benefit of this tutorial to the main experiment is to increase the likelihood that most of the detected performance differences will be due to variations of factors levels.

The procedures described above assess proficiency in the core skills required for the successful completion of the tasks. The results allow in-depth analysis of the main ex-



periment's outcomes, either strengthening observed tendencies and objective results or supporting the resolution of exceptions, conflicts, and unexpected findings. The typical usage of pre-tests results is to group (in the analysis stage) participants based on their skill level in order to evaluate performance differences (MARTIN, 2007).

Upon completion of each task, a NASA Task Load Index (TLX) (HART; STAVELAND, 1988) will be filled (online form) to verify the associated cognitive load (check the original form on Annex E).

#### 4.6.6.2 Main experiment

Considering the remote execution of experiments in individual home sessions, participants performed one condition per week. The recommendation was to perform tutorials in one session, then, later in the same day, executing the three experimental tasks, each followed by a TLX task load test. This division allowed rest in between sessions, potentially improving experiment performance. However, this split was optional, considering the variations in individual capacities, preferences, and available time. Some participants preferred to immediately execute the tasks after the tutorial while others felt tired and either took a short break or executed the rest of the experiment later in the day.

All 2D, 2.5D, and 3D anaglyph modes require a recent computer connected to the internet with at least a 15" monitor. Anaglyph goggles were distributed to all participants and, when necessary, access was provided to computers at the university. All equipment was checked for minimum performance requirements. Framerate was kept at a minimum of 30 fps, and distance to the screen was set to 45 cm for a 24" monitor, being adjusted accordingly to different sizes for a standardized field of view. The WebGL software platform has two versions, DS/S and GC, presenting minor differences. The files were uploaded to a private server and the HTML pages accessed by students using the Chrome browser. Specific web addresses were provided weekly, as the experiment unfolded and students were identified in the system by imputing their university ID. Data was recorded on online spreadsheets (Google Forms).

The immersive 3D mode (3Dst) required VR adapters, provided on individual sessions conducted by the experimenter. Participants were offered a weekly schedule to choose the most convenient date and hour. Due to time and logistic constraints, each participant had one hour to execute all three tasks. Each task was preceded by the corresponding tutorial, and URLs were automatically sequenced by the system. Participants had a break at the end of every task while filling the TLX and could extend this rest for a few minutes if desired. In this case, the TLX was filled on paper, given the technical difficulties to answer online forms on an immersive system. This break was also convenient to provide participants some ocular rest.

#### 4.6.6.3 Dependent variables and auxiliary measurements

Size and HR tasks are evaluated by four scores. Selection ‘strength’ or performance was defined after main scores A and B. The rationale of both scores is detectability.

Score A is simply the position within a size ranking, being one the highest score. The lowest possible score varies per sub-task (Sb: 30; Ss: 60; HRb: 10; HRs: 45) and was determined after preliminary experiments. The assumption is that scores outside this range indicate issues with participant performance as a whole and should be discarded due to the negligent execution of tutorials or the experiment itself. Score A’s intent is assessing the absolute detectability of the experimental condition (technique or pair), assuming that better conditions yield superior overall detectability.

Score B, on the other hand, indicates selection’s quality given the visible grains at the moment, measuring relative detectability. The maximum score B is 10, which means that the higher-ranked was selected amongst the visible. Lower scores mean that other visible grains are better positioned on the ranking. While the ranking logic is clear on size tasks, on HR tasks grains are ranked by the degree of clustering of their neighbors, which takes into account how many neighbors the concerned grain has and how big they are. Finding the exact ‘epicenter’ of a given HR is not trivial, and tolerance was built into the choice algorithm that searched for the higher-ranked HR within immediate closer neighbors. Thus, the participant was credited for the most likely target of the clicked position.

Score C and D are secondary scores, calculated only for good selections, as defined by score A. Once more, the precise cutoff varies per sub-task (Sb:  $A < 10$ ; Ss:  $A < 30$ ; HRb:  $A < 10$ ; HRs:  $A < 15$ ). Both scores are based on exposition or reaction time. Such secondary scores intend to detail the analysis of principal scores and resolve ties between conditions.

Score C is simply the absolute reaction time, counted from appearance to choice, measuring how fast the object was spotted and chosen. Score C is always reset when the object escapes the view. Score D is a factor indicating the relative exposure time, taking into account the total time that a given grain was exposed before selection, obtained by dividing score C (instant reaction time) by total exposure time. Therefore, maximum score D is 1, meaning that the grain was chosen upon the very first appearance. Every time the grain is missed, the score gets progressively smaller as the denominator (total exposure time) only gets larger, and the numerator (current exposure time) is reset to 0. Therefore score D also indicates how many times the object appeared before detection.

On grain shape tasks, the main dependent variable shows if the selection corresponds to the correct aspect ratio. Choice time is also recorded for correct choices, as an auxiliary evaluation variable.

Other relevant data collected by the experimental system:

- total time per task;
- elapsed time since last selection;
- clicked grain and click position per selection;
- movement details per selection: time stopped, quantity and sequencing details of forward and backwards movements.

#### 4.6.7 Possible confounding factors

Learning effects may happen as a transference of task strategy and navigation proficiency from one task to another. Possible countermeasures, based on [Martin \(2007\)](#), are:

- Navigation tutorials for each interaction technique;
- Size and shape perception tutorials;
- Technique as a between-subjects factor;
- Latin square scheme on the S/DS and full counterbalancing for GC technique, related to visualization mode order;
- Latin square counterbalancing for task order: five sub-tasks demands at least ten participants or multiples. Groups of 30 participants meet this requirement.

#### 4.6.8 Tutorials details

##### 4.6.8.1 Navigation

Minimal navigation proficiency is necessary, particularly for highly abstract commands and movements with multiple degrees of freedom, in order to avoid consuming experiment time for learning purposes and impacting results with individual learning curves. Experiment tasks on S and DS demand only forward and backward movements (one degree of freedom), while GC requires another degree of freedom (DOF). However, the shape analysis task requires all 6 degrees of freedom. Therefore, a navigation tutorial should allow enough practice time to achieve proficiency on all required movements, attested by a standardized minimal completion time. This was defined as the average time it takes for three habituated users to complete the tutorial using only one degree of freedom at a time (translation or rotation in one axis). This procedure must be executed for each interaction mode. Desktop (2D and Anaglyph 3D) tutorials can be performed remotely using personal devices, where optimized system variations must be designed to guarantee a minimum acceptable performance even on middle-tier laptops. Tutorials for split-screen 3D devices will be performed on an individual basis before the main experiment.

The task is to reach the sample region and frame a given face with precision, which require all six degrees of freedom. The participant should complete the tutorial in the minimal time, from two different but equally distant starting positions at least, as repeating the task always from a single starting point would induce biased learning. This position can be randomized in case a participant needs several attempts. In the case of 3D VR modes, distance and angles will be adjusted considering the allowed range of rotation.

#### 4.6.8.2 Feature detection sensitivity

Proper grain size and shape detection depends on a sense of what is big or small within the sample and which proportions correspond to the defined shape levels. In order to avoid consuming experiment time or impacting performance measures with the learning curve and distinct subjective scales, a tutorial was deemed appropriate to show participants the expected detection sensitivity. This was accomplished by a remote desktop tutorial, using a topologically similar model. The expected overall benefit of this tutorial to the main experiment is increasing the likelihood that most of the detected performance differences are due to factors levels.

The task consisted of free exploration of the sample until the correct selection of 10 grains of each size class (large, small). Correct hits are confirmed by visual feedback and, in order to avoid guessing and stimulate engaged searching, one can only pick another grain after five seconds. The second task is correctly classifying the shape of eight grains, where each of the four predefined shape options appears twice. In this task, feedback is given by text whenever the participant picks the wrong choice, instructing to try again until the correct choice is made.

#### 4.6.9 Data analysis strategy and procedures

Three factors were defined for the main user study, namely, technique, visualization, and color modes. Discrete sections (S), dynamic sectioning (DS) and GrainCrawler (GC) are the proposed techniques, defined as a between-subjects factor. Flat 2D sections, 2.5D see-through sections, anaglyph 3D, and split-screen 3D (using VR adapters) are the four visualization modes, defined as within-subjects factors, which demands a counterbalancing scheme for order effects.

This mixed factorial study design, which includes a within-subjects factor, demands Linear Mixed Models (LMM) or Generalized Linear Mixed Models (GLMM) depending on the distribution of responses, for proper evaluation of both fixed effects (caused by factors) and random effects (participants) (WOBBROCK; KAY, 2016). Although such methods are more computationally intensive, they are much more robust to missing data and do not require spherical or balanced data.

The first analysis is performing normality (Shapiro-Wilk) followed by homoscedasticity (Levene) tests on time and performance responses for each relevant combination of technique, visualization, and color modes along with their residuals (fitting an analysis of variance model). Although normal data is usually expected from experimental responses, counts such as errors or hits (success) are usually random or Poisson distributed instead, while task time is usually log-normal (WOBBROCK; KAY, 2016). If data is not normal, it will be tested for log-normality and finally, for fitness to the other distributions. Log-normal distributions can be adjusted for use in LMM. Whenever a distribution deviates from normality or log-normality, a Generalized Linear Mixed Model will be used. Finally, if data do not fit any distribution, a non-parametric Aligned Rank Transform (ART) will be adopted instead (WOBBROCK; KAY, 2016).

If the data is within acceptable standards, an LMM ANOVA follows to search for order effects on within-subjects factors, the visualization and color modes in this case. Despite all the measures taken to avoid learning effects, task order effects will also be analyzed for completeness. Technique and exposition order of visualization modes are set as fixed effects and subject as a random effect. If order effects are not found, then the same test is repeated without the order factor to finally check for significant differences between the three techniques and between eight pairs (combinations of techniques and visualization modes). If this omnibus test reveals any significant difference, then techniques can be compared, and post hoc pairwise comparisons between pairs are possible, applying a multiple corrections adjustment (Holm-Bonferroni). The most relevant comparisons are:

- overall comparisons: best technique (S x DS x GC);
- the best visualization mode for each technique: differences of visualization modes within DS and GC techniques;
- best technique for each visualization mode: S x DS x GC;
- ranking of pairs: top-performing combinations.

\* Notice that GC does not have a 2D Flat visualization mode. Whenever 2D visualization is compared between techniques, GC will participate with 2.5D modes.

This statistical analysis process concerns main scores A and B and auxiliary scores C and D. In addition to variance analysis, all scores were further analyzed by other means to assess identified differences or found unidentified ones. Score A was further analyzed by counting choices by ranking stratum ( $A < 60$ ,  $50 < A < 60$ ,  $40 < A < 50$ ,  $30 < A < 40$ ,  $20 < A < 30$ ,  $10 < A < 20$ ,  $4 < A < 10$ ,  $2 < A < 4$  and  $0 < A < 2$ ). This analysis may indicate detection sensitivity to specific strata, with particular interest to the first three ( $A < 10$ ). Score B was further analyzed by calculating the percentage of choices with score B = 10, which means that the

best visible grain was chosen. This analysis was performed per technique, pairs and color mode. Score C was also analyzed by score A rank strata, showing reaction time mean per stratum. Score D was also further analyzed by counts, as D=1 means detection on the first appearance. This analysis was performed per technique, pairs, color mode, and score A rank strata.

#### 4.6.10 Hypothesis statement

The main hypothesis is that GC technique shall perform better than DS, followed by the benchmark technique S, and stereoscopic visualization modes shall perform better than 2.5D modes, followed by 2D Flat mode. Thus, considering pairs, the expected sequence is GCs > GCa > GCd > DSs > DSa > DSd > DS > S, and the superior combination shall be GC with split-screen stereoscopy.

### 4.7 Alternative: Single-subject and small-N baseline study designs

This section will briefly discuss possible advantages, scientific concerns, and known limitations of this evaluation method, widely used in experimental psychology. Compared to experimental or quasi-experimental studies, the main difference is that only one or a few participants perform a long series of tries per pair, reducing the need for participants drastically. The main issues with this design are learning effects related to memorization, navigation, task proficiency, and differences in individual responses. In this work, this type of study would adopt the same variables, tasks, techniques, and measurements described on the last sections.

Model memorization can be counteracted using different views and transformation operations on each available model, as described in section 4.6.5.2, and by creating other similar models. When a new model is created, care must be taken to enforce the closest possible characteristics related to performance (hits and time), mainly sample size, shape, and density of features (Ss, Sb, and HR tasks). Finally, neighborhood and grain inspection require only a substantial number of different grains of varied shapes, a condition satisfied by one or a couple of models.

New models can be created using the techniques described by [Rodrigues et al. \(2017\)](#) and several grain measurement functions are already implemented in the current version of the software, in order to assure that all characteristics conform to the standards. Participants may share the same models and a common sequencing scheme to ease comparisons. Since a given model can be reused by approaching it from different angles or applying transformation operations, the sequencing scheme can further reduce the likelihood of model recalling by alternating the model on every try. For example, if four models are available (A, B, C, D), the sequence A1-B1-C1-D1-A2-B2-C2-D2 (and so on) can be

adopted, where 1 and 2 are model variations. As stated before, nine variations are possible per model when evaluating S and DS and at least three for the GC technique.

Concerning learning issues, the baseline notion assumes that task performance cannot be infinitely improved, i.e., a stable performance condition is achieved after a finite number of tries, regarding all skills relevant to the task. This means that performance will stabilize after a few tries, reaching a unique baseline for each technique variation. Order effects do not apply as long as a baseline is achieved. Long term performance improvements are likely to be smaller and systematic. Thus, the baseline concept relies heavily on the power function of [Newell e Rosenbloom \(1993\)](#), discussed in section 2.2, which states that most learning occurs upon the beginning of training and further minimal gains require a disproportionately high amount of time. However, one cannot use this type of study to assess factors that involve forgetting acquired knowledge, such as the learning curve of a technique, as a person cannot undo learning ([MARTIN, 2007](#)).

Additionally, learning effects can be reduced by grouping variations per technique and allowing a training period. If a stable proficiency in the technique is achieved, then performance differences can be attributed to the visualization method. Therefore, the experimental design should guarantee enough tries for every variation in order to achieve a stable performance baseline and assure that each baseline can be reproduced closely, which requires at least a second (and shorter) battery of tries.

Results of single-subject or Small-N Baseline experiments do not require special statistical tools, only direct comparisons between baseline values and transition-state values ([MARTIN, 2007](#)). Results are firstly analyzed by comparing baselines of first and second batteries, for each technique variation, per individual. Large differences are not expected to be found, which will confirm that learning effects did not occur and that other confounding factors were absent. Success in this analysis then allows a direct comparison of techniques by simply comparing baseline values. Other interesting insights can be obtained, such as deriving the learning curve for each technique, which is indicated by the average number of tries (and total time) taken to achieve the baseline on the first battery. However, given that learning cannot be reversed, as discussed above, this can only be realized once per participant. Finally, pairwise comparisons of participant results can be performed to confirm the technique ranking consistency or find relevant individual variations.

This type of experimental design has several advantages, mainly the reduction of participant needs and faster detection of large effects. However, smaller but relevant effects cannot be properly detected, and the method is still not widely adopted nor well investigated in HCI literature. Therefore, its application in this work was considered risky, and a study with multiple participants was preferred over Single subject or Small N-baseline studies. The primary purpose of this discussion was to broaden the assessment of available experimental methods.

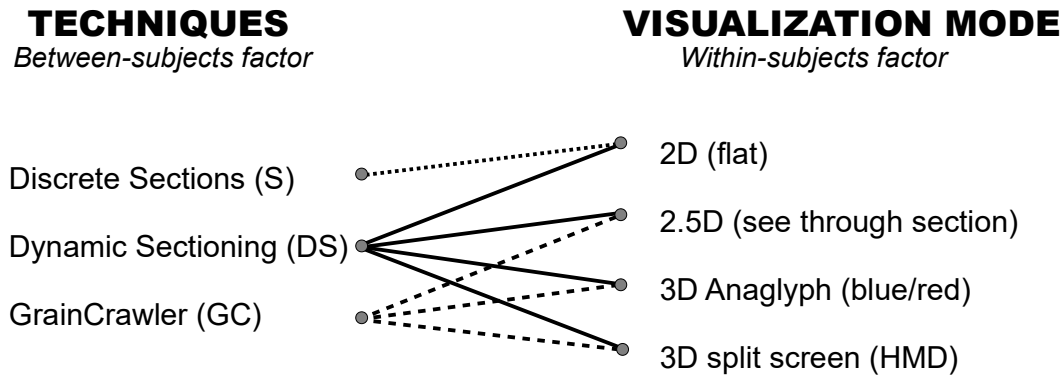


Figure 61 – User study summary. Note that S is tested only in Flat 2D and GC do not use 2d flat rendering.

## 4.8 Evaluation summary

The evaluation followed an immersive continuum conceptual framework, upon which the experimental system was founded. As perception and cognition were considered a priority, the experimental system focused on the visualization factor of this continuum, and visualization modes were assessed on experiments.

Dependent variables were chosen after spatial comprehension objectives, concerning sample and grain scales. Grain scale concerns individual grain size and shape while the sample scale is related to localized grain size variations. Thus, techniques should allow detecting heterogeneous regions and extreme values in the sample scale, then grain size and shape factor in the grain scale. Three experimental tasks were designed: grain size identification, shape evaluation (four shape factors were defined - equiaxial, brick, discoid, stretched disc) and detection of heterogeneous regions (HR) or clustering of small/big grains. Movement was restricted to a single axis on size and HR tasks to limit the influence of interaction on feature detection and allow comparison between techniques.

Two preliminary experiments were performed to assess task efficiency, experimental design, and to further detail logistic requirements (instructions, time and space, scheduling); assess distinct scoring possibilities, check the character of response distributions and decide upon data collection methods; collect user feedback and improve techniques; assess pre and post questionnaires.

Size and HR tasks are evaluated by four scores. Selection ‘strength’ or performance was defined after main scores A and B. The rationale of both scores is detectability. Score A is simply the position within a size ranking, assessing the absolute detectability. Score B indicates how good the selection is, given the visible grains, measuring relative detectability. Score C and D are secondary scores, calculated only for good selections, as defined by



score A. Score C is simply the absolute reaction time, measuring how fast the object was spotted and chosen. Score D is a factor related to the relative exposure time, indicating either if the grain was chosen upon the very first appearance or how many times the object appeared before detection.

The proposed user study design has three factors or independent variables: technique, visualization mode, and color mode. Technique has three levels (S, DS and GC), visualization has four (2D, 2.5D, 3D anaglyph, and immersive 3D), and color has two (mono, colored). Although a few experimental designs are available, such as single-subject and small-N baseline, quasi-experimental designs were considered more appropriate. In order to reduce the number of participants required for an optimal between-subjects factorial user study, visualization and color were set as within-subjects factors while technique remained a between-subjects factor. Each experimental group has 30 participants, one for DS and S and another for GC, totaling 60 participants. On DS/S group, each person performs one visualization mode per week (three tasks) for five weeks, including S in one of those weeks, while GC requires only three weeks. Figure 61 summarizes the formal user study design.

Except for immersive 3D visualization mode, which is performed in scheduled individual sessions along the experimental period, experiments are executed on remote home sessions on a browser-based system, including online pre/post tests. Participants perform SBSOD and MRT spatial ability pre-tests once. Interaction and feature detection tutorials are performed before each task and a TLX task load questionnaire is filled after the task. The presence of within-subjects factors in this mixed experimental design (combining WS and BS factors), along with random effects (participant variability) demands the application of Linear Mixed Models for normally distributed responses, Generalized Linear Mixed Models (GLMM) for responses distributed according to other known functions, or the Aligned Rank Transform (ART) for other cases. After checking the responses distributions, data is checked for order effects on the WS factors, proceeding to an omnibus test looking for statistically relevant differences and, finally, performing pairwise comparisons between the relevant combinations of techniques and visualization modes. Color effects are analyzed separately.



## 5 Results

This chapter starts with a summary of MRT and SBSOD spatial ability pre-tests and TLX task load post-tests, followed by task order influence assessment and, finally, a summary of experimental results of the three experimental tasks (Size, HR, Shape), including a ranking of techniques, pairs and color modes, shown for each task. Refer to Appendix A for the complete and detailed quantitative analysis of each task and score, including responses histograms, significant results, and significance levels. Comparisons are shown both in tables and boxplots.

### 5.1 Pre-tests on experimental groups

GC and DS experimental groups were characterized by spatial ability tests - the test of Mental Rotations (MRT), and the Santa Barbara Sense of Direction (SBSOD). Data distributions deviated from normality and were analyzed using Generalized Linear Mixed Models (GLMM) adopting the Gamma distribution family.

#### 5.1.1 MRT

The GLMM ANOVA was not significant for MRT, and the mean differences are negligible (Table 2, figure 62).

Table 2 – Mean differences of MRT scores per experimental group

	MRT Mean	MRT s.d.	MRT dif.
DS	20,42	4	-
GC	20,67	3,1	-

#### 5.1.2 SBSOD

The GLMM ANOVA was significant for SBSOD ( $P < 0.01$ ); however, the means are extremely close (Table 3, figure 62).

Table 3 – Mean differences of SBSOD scores per experimental group.

	SBSOD Mean	SBSOD s.d.	SBSOD dif.
DS	6,83	1,4353	4,6
GC	6,53	1,392	-4,4

Considering the MRT and SBSOD tests, experimental groups are considered homogeneous when it comes to spatial abilities. This was expected as both groups consisted of students from two subsequent offerings of the same undergraduate course.

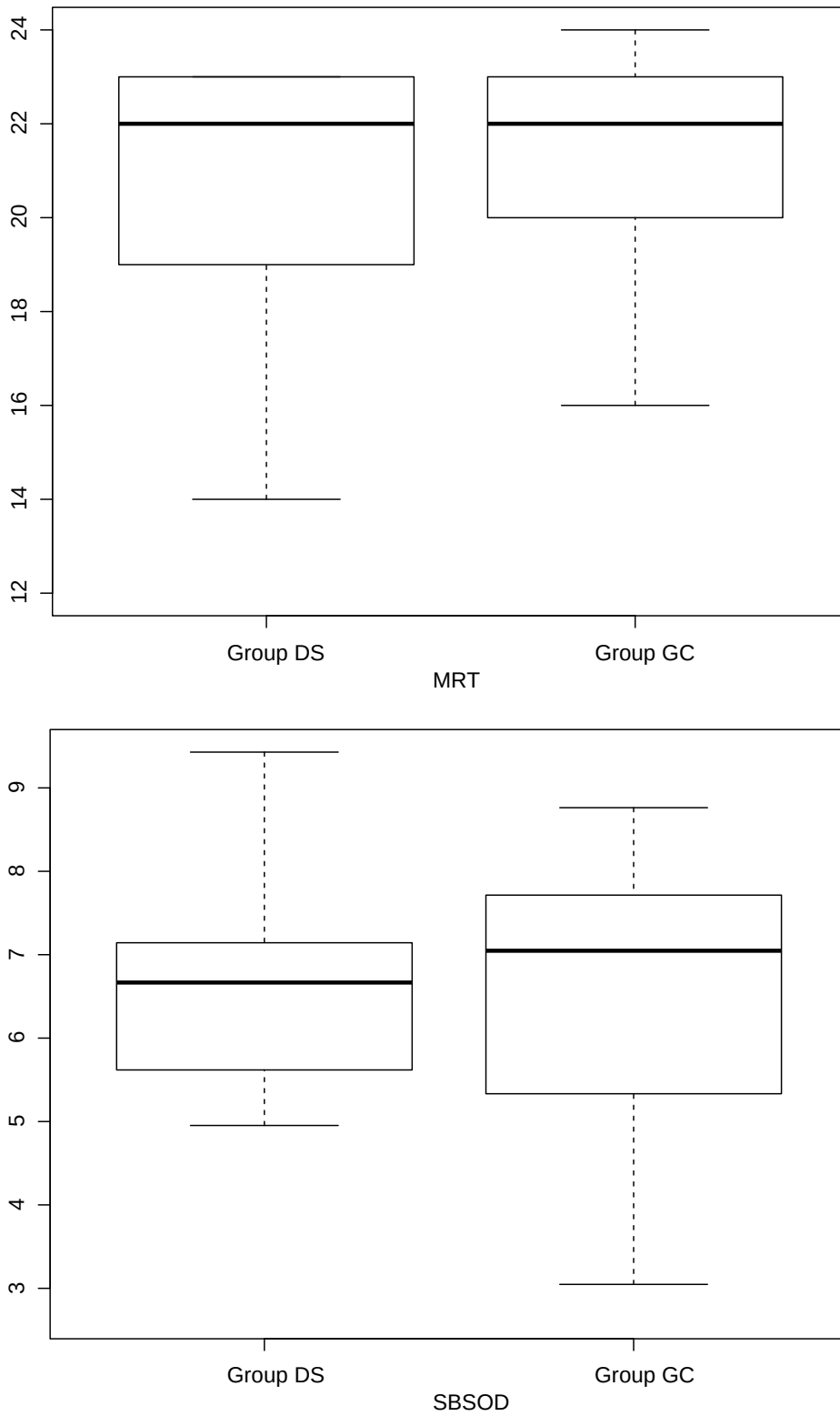


Figure 62 – Boxplots of MRT (top) and SBSOD (bottom) scores.

## 5.2 TLX analysis

The Nasa Task Load Index test was administered to each participant after the conclusion of each experimental condition (technique versus mode combinations).

### 5.2.1 Size tasks

In this section, the results of TLX are presented, considering only size tasks (one TLX for Sb and Ss tasks). Data distributions deviated from normality and were analyzed using GLMMs, adopting the Gamma distribution family. It is known that large data sets are more likely to deviate from normality; however, QQ-Plots confirmed the deviation. GLMM ANOVA results was not significant on techniques (Table 4) but were significant for pairs ( $p < 0.01$ ), visualization modes (DS n.s., GC  $p < 0.01$ ) and color modes (omnibus  $p < 0.01$ ; DS n.s.; GC  $p < 0.01$ ). Pair boxplots (Figure 63, top) clearly shows how GCa and GCs stand out from all others with higher TLX scores (26 to 39% higher versus DS and S, see table 6). Therefore we can conclude that technique DS differs only from 3D variations of GC, where Anaglyph GC is the most stressful pair, 10% higher than GCd.

Considering color modes, the difference within GC technique seems to explain the difference in the omnibus test (see figure 63, bottom and table 5). Mono GC score was 26% higher than Color GC.

Table 4 – TLX analysis: significant differences (%) per technique on Size tasks

Technique	Mean	s.d.	Difference
DS	3,95	1,01	-
GC	4,51	1,53	-

Table 5 – TLX analysis: significant differences (%) per Color mode on Size tasks

Color mode	Mean	s.d.	Difference
DS Color	4,02	0,97	-
DS Mono	3,8	1,09	-
GC Color	3,75	1,16	-20,98
GC Mono	4,75	1,56	26,55
Color	3,88	1,08	-
Mono	4,64	1,54	-

### 5.2.2 Shape task

One should first notice that the S technique and Color modes on GC do not apply to the Shape factor task.

GLMM ANOVA results was significant only for techniques ( $p < 0.01$ , see table 7) and pairs ( $p < 0.01$ ). Boxplots (Figure 64, top) clearly shows how GCa and GCd stands out

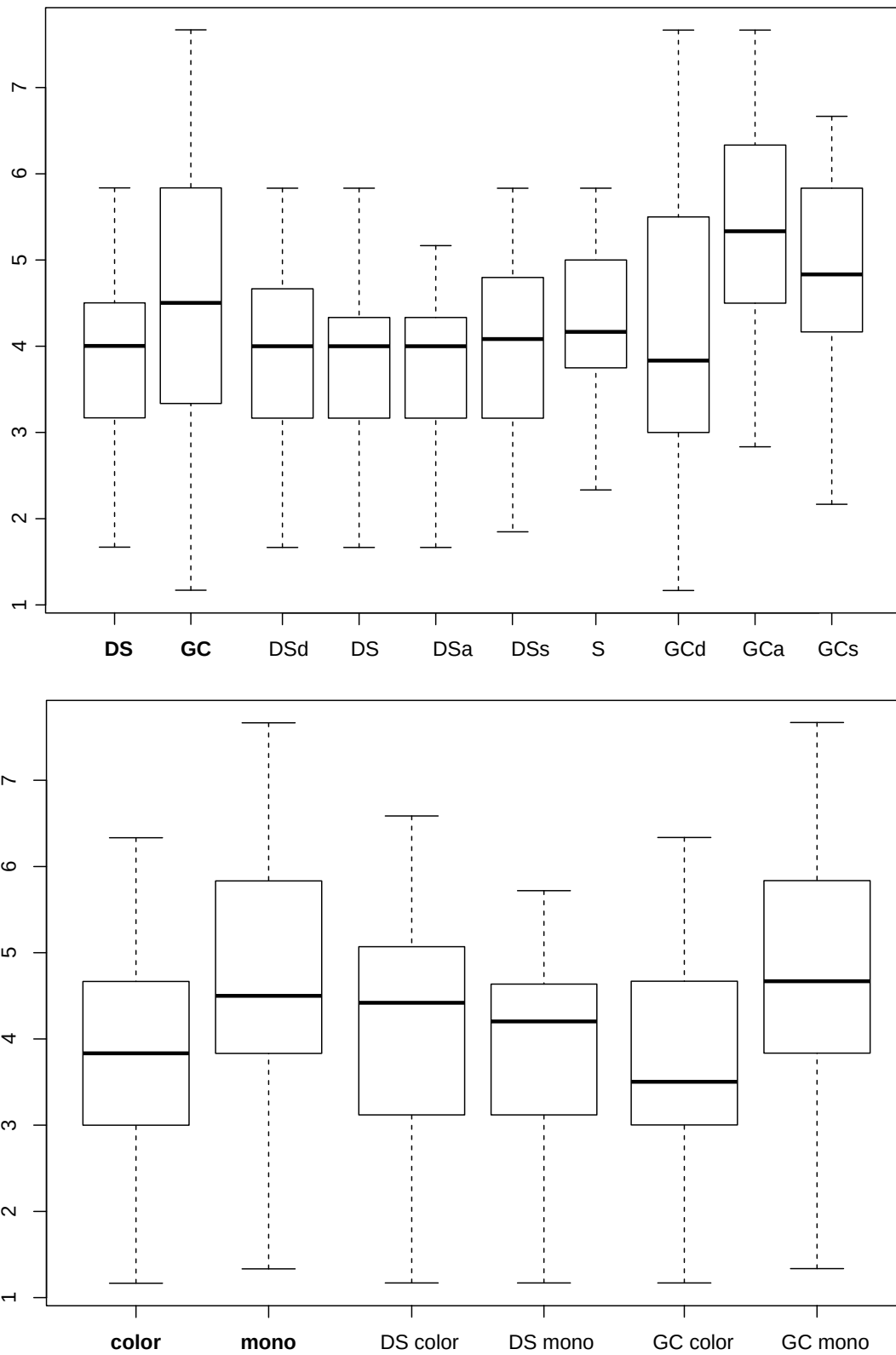


Figure 63 – Size tasks: Boxplots of TLX per pairs (top) and color modes (bottom).

Table 6 – TLX analysis: significant differences (%) between pairs on Size tasks

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	4,16	0,96	0	-	-	-	-	-	26,3	15,3
DS	3,81	0,94	-	0	-	-	-	-	37,8	25,8
DSd	3,96	0,97	-	-	0	-	-	-	32,5	21
DSa	3,88	1,18	-	-	-	0	-	-	35,2	23,5
DSs	3,78	1,04	-	-	-	-	0	-	39	26,9
GCd	4,1	1,56	-	-	-	-	-	0	28	16,8
GCa	5,25	1,44	-20,8	-27,4	-24,5	-26	-28	-21,8	0	-8,6
GCs	4,8	1,2	-13,3	-20,5	-17,3	-19	-21,2	-14,4	9,5	0

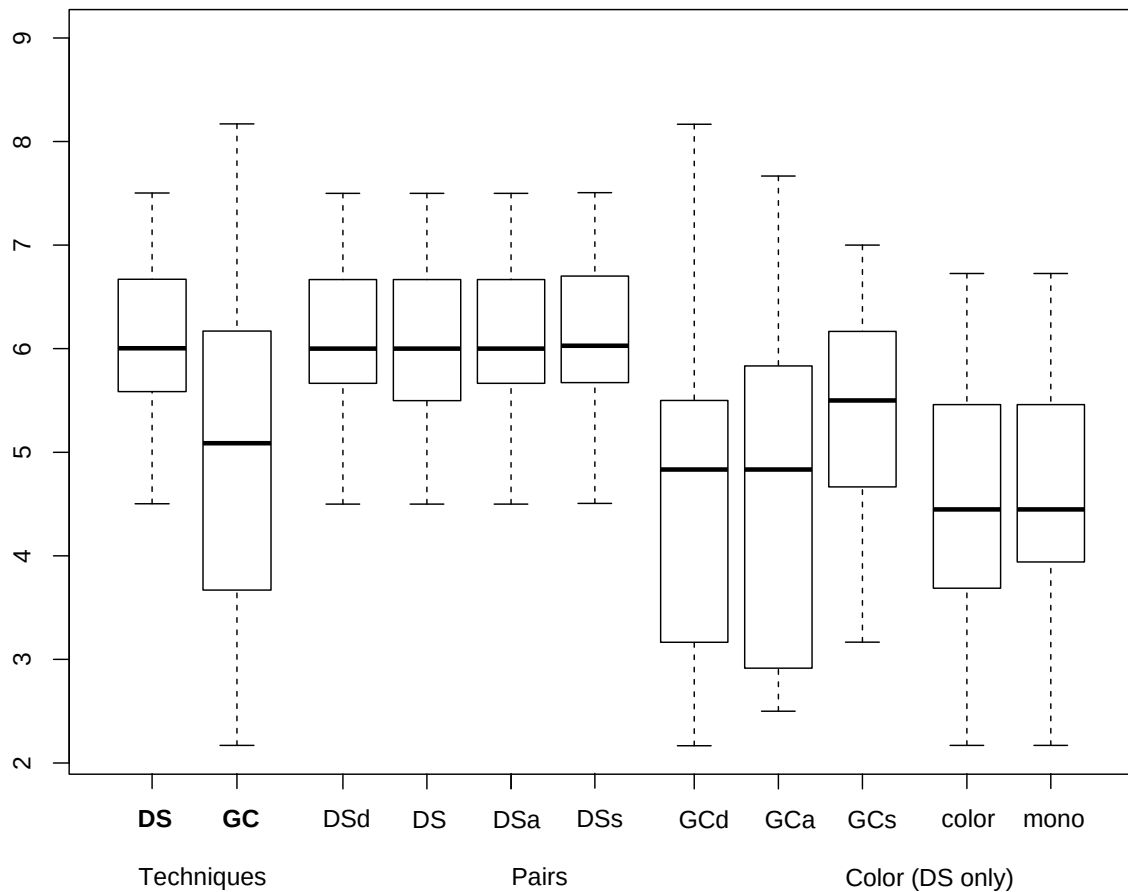


Figure 64 – Shape task: Boxplot of TLX showing techniques (two left), pairs (central seven) and color modes (two right).

from all others with lower TLX scores (GCa 25 %, GCd 27 % versus DS, see table 9). Color modes on DS was not significant and the difference is negligible (Table 8)

Therefore, we can conclude that DS and GC techniques differ, except for GCs, and color played no role. GCd and GCa were deemed less stressful.

Table 7 – TLX analysis: significant differences (%) per technique on Shape tasks

Technique	Mean	s.d.	Difference
DS	6,21	1,03	24,84
GC	4,97	1,39	-19,9

Table 8 – TLX analysis: significant differences (%) per Color mode on Shape tasks

Color mode	Mean	s.d.	Difference
DS Color	6,17	1,03	-
DS Mono	6,26	1,03	-

Table 9 – TLX analysis: significant differences (%) between pairs on Shape tasks

Pairwise comparisons (pairs Technique x Mode)									
	Mean	s.d.	S	DS	DSd	DSa	GCd	GCa	GCs
S	6,2	1,09	0	-	-	-	-27,6	-25,3	-
DS	6,2	1,07	-	0	-	-	-27,6	-25,2	-
DSd	6,22	0,92	-	-	0	-	-27,8	-25,4	-
DSa	6,19	0,97	-	-	-	0	-27,4	-25,1	-
DSs	6,2	0,95	-	-	-	0	-27,6	-25,2	-
GCd	4,49	1,46	38,2	38,2	38,6	37,8	0	-	-
GCa	4,63	1,56	33,8	33,8	34,2	33,5	-	0	-
GCs	5,35	1,17	-	-	-	-	-	-	0

### 5.2.3 Heterogeneous Regions tasks

It is important to remind that Color modes on GC do not apply to this task. GLMM ANOVA results were not significant. Boxplots (Figure 65) show no sharp distinctions between any technique, pair, or color mode.

Therefore we can conclude that no condition was particularly more stressful than any other, and the mean was around 5 (see tables 10, 11, and 12).

Table 10 – TLX analysis: significant differences (%) per technique on HR tasks

Technique	Mean	s.d.	Difference
DS	4,9	0,98	-
GC	5	1,52	-

Table 11 – TLX analysis: significant differences (%) per Color mode on HR tasks

Color mode	Mean	s.d.	Difference
DS Color	4,93	0,93	-
DS Mono	4,88	1,06	-

### 5.2.4 Discomfort on 3D conditions and VR sickness

Several participants reported ocular distress, headaches, and even nausea while using anaglyph goggles. This was not an issue on the HMD, where only a few VR sickness cases were reported. In fact, many more cases were expected due to the keyboard-based movements of the experimental system. Refer to section 6 for more details on those issues.



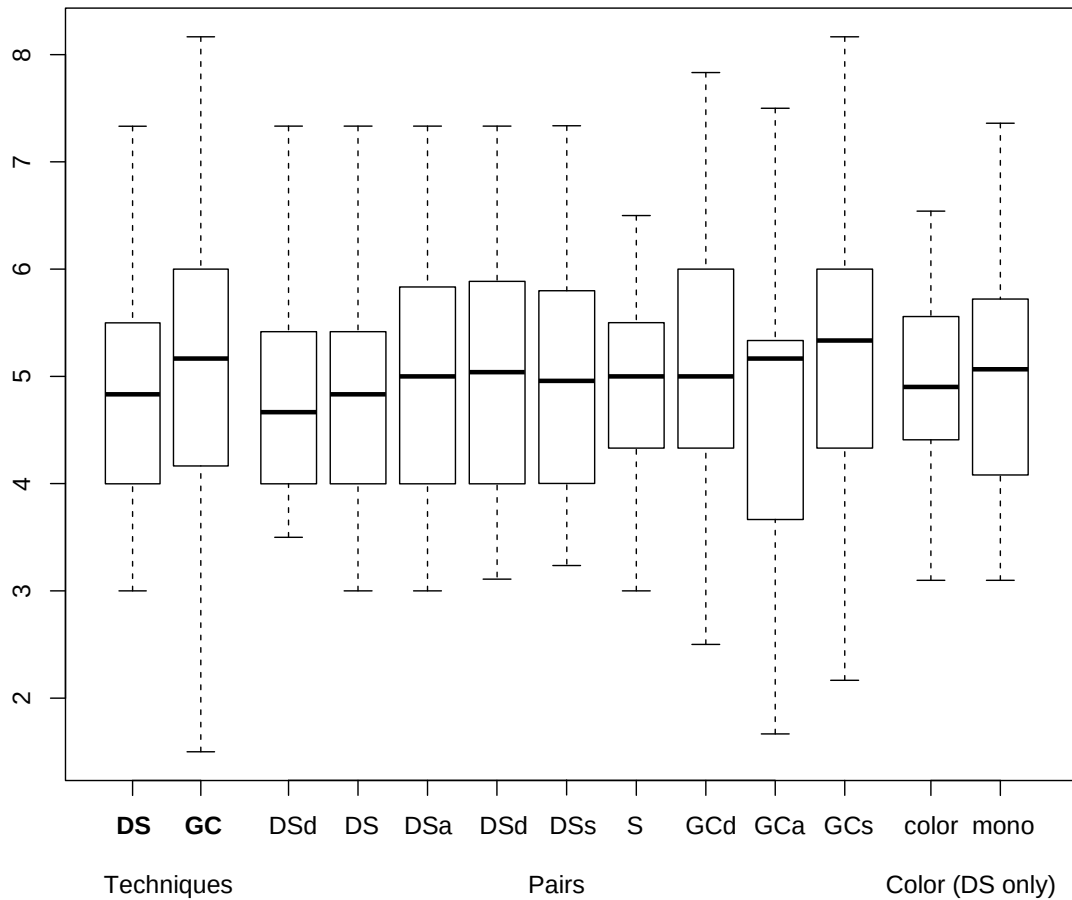


Figure 65 – HR tasks: Boxplot of TLX showing techniques (two left), pairs (central nine) and color modes (two right).

Table 12 – TLX analysis: significant differences (%) between pairs on HR tasks

	Mean	s.d.
S	4,93	0,97
DS	4,89	0,98
DSd	4,81	0,92
DSa	5,01	1,05
DSs	5,11	0,99
GCd	5,03	1,57
GCa	4,81	1,5
GCs	5,12	1,46

### 5.3 Task order influence assessment

Testing six modes requires two 6 x 6 latin squares (Table 13) - the second in inverse order - resulting in 12 groups of three participants. Therefore, each group tested a distinct order of system modes. The shape task presented no sharp variation of performance, with a mean percentage of correct answers around 31% and a 20% standard deviation (Table 14). Therefore execution order has not influenced performance on this task.

The order of other tasks was analyzed with ANOVAs on main scores A and B. The size

Table 13 – Sequence of execution order per group. Each letter refers to a system variation

Group	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
1	A	F	B	D	C	E
2	C	E	D	B	F	A
3	F	D	C	E	A	B
4	D	B	A	F	E	C
5	E	C	F	A	B	D
6	B	A	E	C	D	F

Table 14 – Shape factor task: performance per order group (% of correct answers). Mean = 31, S.d.= 6

1	1i	2	2i	3	3i	4	4i	5	5i	6	6i
39	27	30	24	36	23	33	34	42	31	25	30

task was analyzed using GLMM on score A and B, given the Weibull distribution of the responses. Although some differences can be spotted on boxplots (Figure 66), ANOVAs were not significant. HR task was analyzed using GLMM on score A, given the Weibull distribution, and the Aligned Rank Transform on score B, as responses did not conform to any known distribution. Once more, although slight differences can be spotted on boxplots (Figure 67), ANOVAs were not significant. Therefore task order elicited no effect on performance in any task.

## 5.4 Results summary

The following summary shows rankings of techniques and pairs (combinations of techniques with visualization modes) per score for each task, along with an overall ranking defined mainly after scores A, B, and strata analysis. Scores D and C are the first tiebreakers, followed by the sum A+B+D, then, if necessary, score C.

### 5.4.1 Size - Big

The final order of pairs is: DSd/DSa/DS/DSs (1), S (2), GCs (3),GCa (4),GCd (5). Therefore, the order of techniques is DS, S, and GC. Scores are mostly consistent (Table 15). Mono color excelled on GC.

### 5.4.2 Size - Small

With score A and B as the most critical metrics and adjusting with other scores, the final order of pairs is: GCs (1), GCa (2), DSs (3), GCd (4), DS, (5), S (6), DSd (7), DSa (8). Considering overall means and performance of pairs, the order of techniques is GC, DS / S, with score D favoring technique DS, as a tiebreaker with S. Except for score C, excelled by technique S, the other scores are mostly consistent (Table 16). Color excelled on DS.

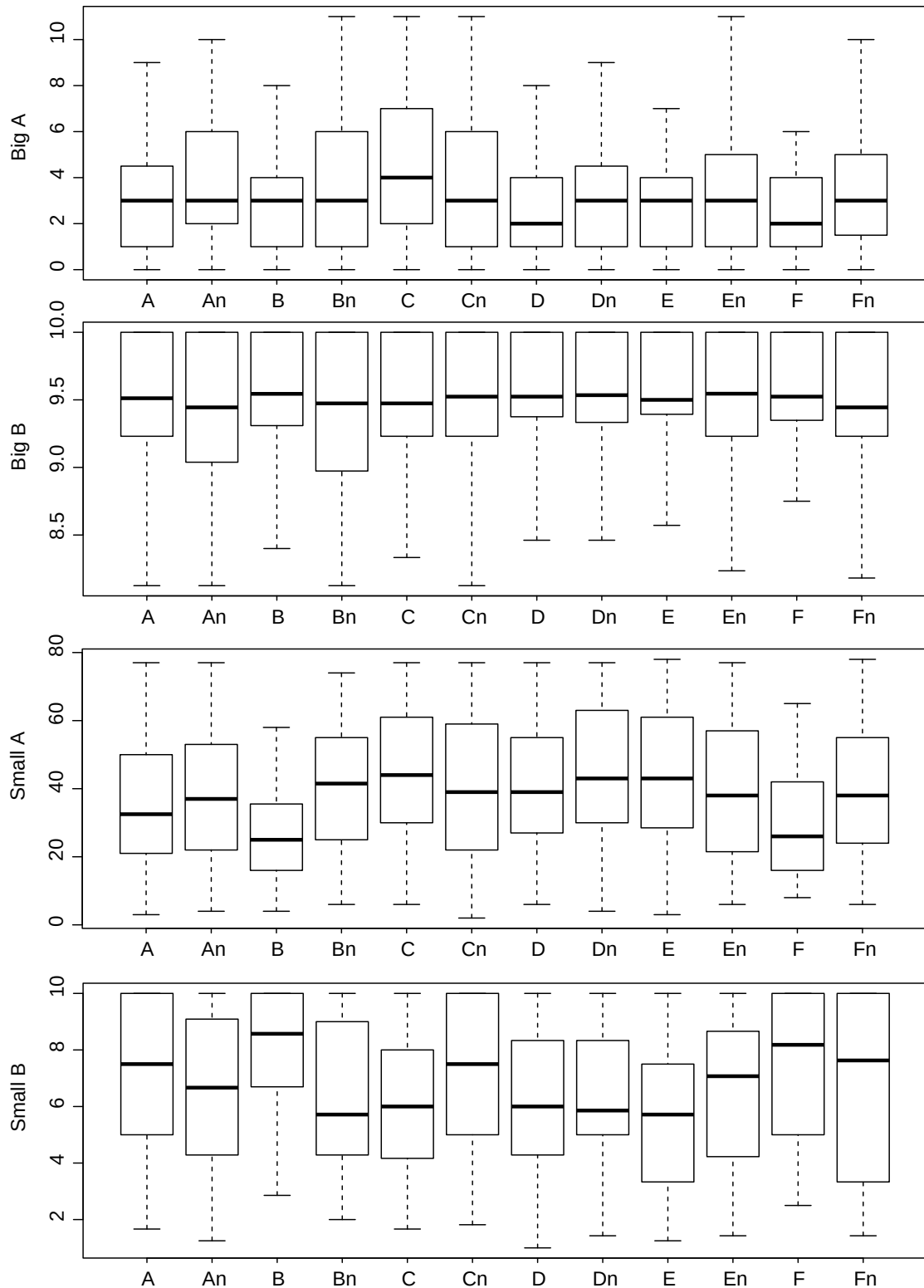


Figure 66 – Boxplots of scores A and B for Size tasks, by execution order.

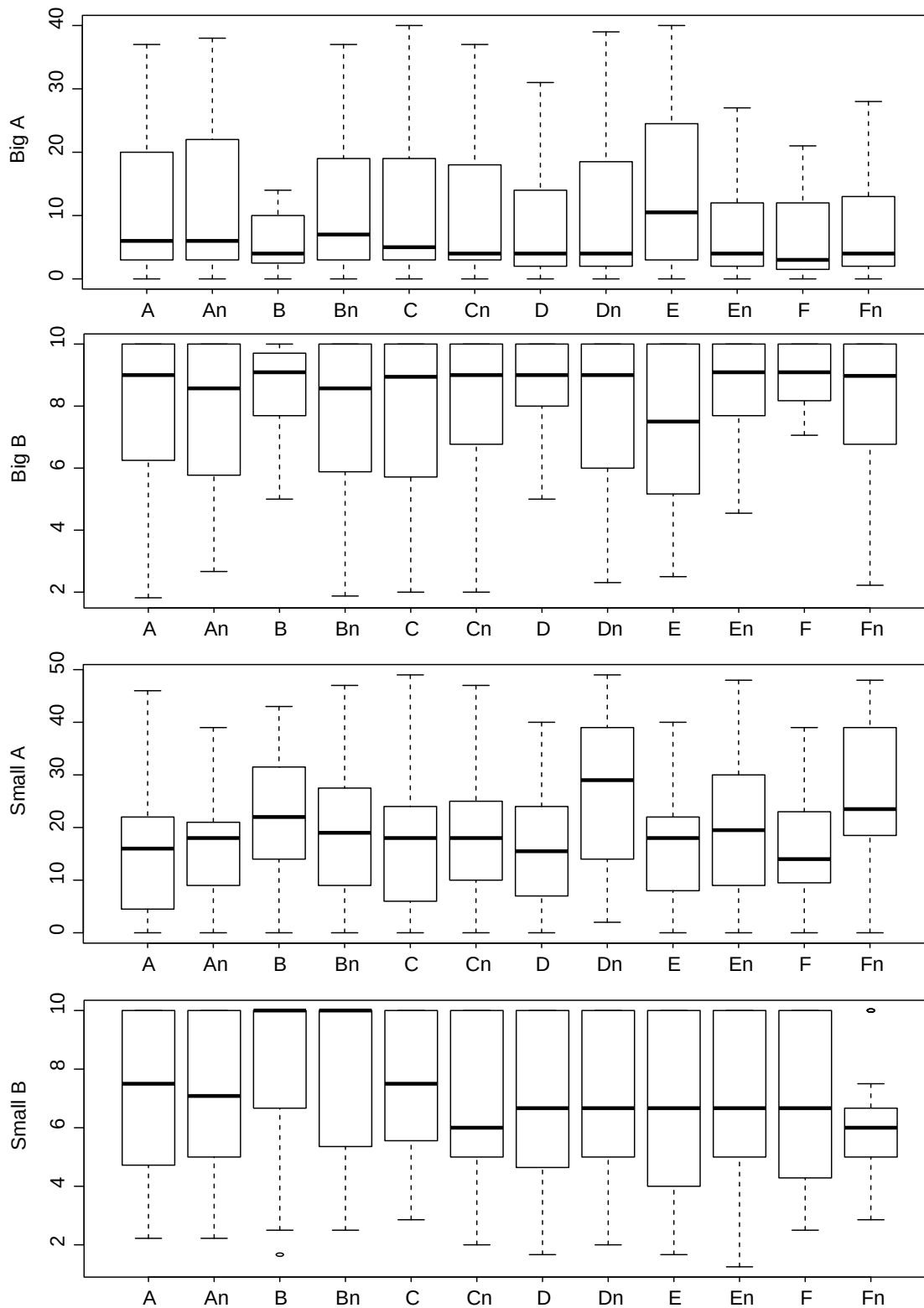


Figure 67 – Boxplots of scores A and B for HR tasks, by execution order.

Table 15 – Results summary of Sb task: ranks per score.

	Score A	Score B	Score C	Score D	A+B+D
Technique DS	1	1	2	2	1
Technique GC	2	2	3	1	2
Technique S	1	1	1	3	2
Pair DS 2D	1	1	2	1	1
Pair DS 2.5D	1	1	1	1	1
Pair DS 3D A	1	1	1	1	1
Pair DS 3D S	1	1	3	1	1
Pair GC 2.5D	3	4	4	4	4
Pair GC 3D A	4	3	3	3	3
Pair GC 3D S	2	3	1	2	2
Pair S	1	1	1	5	2

Table 16 – Results summary of Ss task: ranks per score

	Score A	Score B	Score C	Score D	A+B+D
Technique DS	2	2	2	2	2
Technique GC	1	1	3	1	1
Technique S	3	3	1	3	3
Pair DS 2D	2	3	2	5	5
Pair DS 2.5D	3	4	3	5	6
Pair DS 3D A	4	5	4	4	7
Pair DS 3D S	1	3	5	4	4
Pair GC 2.5D	2	2	7	3	3
Pair GC 3D A	2	2	6	2	2
Pair GC 3D S	1	1	5	1	1
Pair S	2	4	1	6	6

### 5.4.3 HRb - Bigger grain clustering

The final order of pairs is: DSa (1), DS (2), DSd (3), S (4), GCd (5), GCs/GCa (6), DSs (7). Therefore, the order of techniques is DS, S, GC. Score C was excelled by technique S and score D by technique GC (3D modes). Main scores A and B are consistent (Table 17).

Table 17 – Results summary of HRb: ranks per score

	Score A	Score B	Score C	Score D	A+B+D
Technique DS	1	1	2	2	1
Technique GC	3	3	3	1	2
Technique S	2	2	1	3	2
Pair DS 2D	1	2	2	4	2
Pair DS 2.5D	2	2	4	3	2
Pair DS 3D A	1	1	2	2	1
Pair DS 3D S	4	3	3	2	4
Pair GC 2.5D	3	2	5	3	3
Pair GC 3D A	3	4	7	1	3
Pair GC 3D S	3	4	6	1	3
Pair S	2	3	1	4	4

#### 5.4.4 HRs - Smaller grain clustering

With score A and B as the most relevant metrics and adjusting with other scores, the final order of pairs is: DS (1), DSd(2), DSs (3), DSa (4), S (5), GCa (6), GCd (7), GCs (8). Therefore, the order of techniques is DS, S, GC. Except for score D, excelled by technique GC, the other scores are consistent (Table 18).

Table 18 – Results summary of HRs task: ranks per score

	Score A	Score B	Score C	Score D	A+B+D
Technique DS	1	1	1	2	1
Technique GC	3	3	3	1	2
Technique S	2	2	2	3	2
Pair DS 2D	1	1	3	4	1
Pair DS 2.5D	1	2	2	4	2
Pair DS 3D A	2	2	1	3	2
Pair DS 3D S	2	1	4	3	1
Pair GC 2.5D	5	3	5	2	4
Pair GC 3D A	4	4	7	1	3
Pair GC 3D S	5	5	6	1	5
Pair S	3	2	2	5	4

#### 5.4.5 Shape task

Considering counts and time responses, GC was followed by DS, and pairs follow the order GCs (1), GCd (2), GCa (3), and all DS pairs come tied in last (Table 19).

Table 19 – Results summary of shape task: ranks per response

	Counts	Time	Sum
Technique DS	2	2	2
Technique GC	1	1	1
Pair DS 2D	3	4	4
Pair DS 2.5D	3	4	4
Pair DS 3D A	3	4	4
Pair DS 3D S	3	4	4
Pair GC 2.5D	1	2	2
Pair GC 3D A	2	3	3
Pair GC 3D S	1	1	1

#### 5.4.6 Final comments on results

Primary scores A and B are consistent on all tasks, while score D was consistent on Size tasks but not on HR tasks. Color played a role only in Size tasks.

Overall there is no clear dominant winner considering all tasks. Except for the Ss and shape tasks, the technique order found is DS, S, and GC. Regarding pairs, given the ranking positions in each task and tiebreakers, the overall order found is DS, DSd, DSa, DSs, S, GCs, GCa, GCd, with 2D modes excelling in general.

Ss task seems to follow a distinct rationale, where GC was superior, followed by DS and S. 3D modes of GC and split-screen stereo DS performed better.

GC dominated the shape task, particularly in split-screen stereo visualization mode.

Within GC, split-screen stereo excelled at size and shape tasks, while 2.5D excelled at HR tasks. Within DS, 2D and split-screen stereo were better at size tasks while 2D modes excelled at HR tasks.

Anaglyph 3D performance was inconsistent, and several cases of ocular fatigue and nausea were reported.

Notably, GC won all scores D, particularly with 3D modes. The S technique also excelled at most scores C.





## 6 Discussion

This project's scientific objective is to compare the efficiency of three visualization techniques in detecting relevant characteristics of grain or honeycomb models, considering particular learning objectives. Due to the high complexity of assessing learning improvements, it was decided to tackle a few fundamental features that could improve such structures' comprehension. The following characteristics were chosen: detection of extreme grain sizes (larger and smaller), heterogeneous regions (concentrating large or small grains), and form factor of isolated grains. Therefore, three tasks were designed to compare the performance of three techniques on formal experiments, namely Discrete Sections (S), Dynamic Sectioning (DS), and GrainCrawler (GC).

S simulates what is feasible to execute with real materials by extracting images of parallel sections at small distances, i.e., a visualization based on sequences of photos. DS is an improvement of S, where the user can interactively navigate the material through 'infinite' sections. A digital model allows navigating in any direction, so one can evaluate the shape of a grain by rotating around its center, which is not possible in S. Finally, GC was developed considering the model's peculiarities, perceptual potentials, and educational objectives. A multiscale visualization paradigm was chosen, where the representation of grains changes according to distance and perception objectives.

Except for grain shape analysis, the other tasks comprise identifying objects among distractors, which basically involves dimensional properties (minimum/maximum size). Such visual search tasks have been extensively studied in the HCI literature. Conventionally, the performance indicators are search quality and the response time, counted from the object's appearance to detection. Based on preliminary experiments, two main performance indicators were established: the object size ranking (overall performance) and the size ranking among the visible objects (transient performance). Two additional indicators were defined to analyze the 'good' choices in more detail: reaction time and a factor that determines whether the object was detected at the first appearance. In order to perform such measurements, it is necessary to determine the FOR (field of regard) continuously, i.e., which objects are visible at the moment.

Concerning the S technique, the view is restricted to the visible objects on the section plane. It is necessary to go through the sections and rely on memory to assess the size of an object or group. It is assumed that large objects can be detected instantly upon spotting a wide 2D section, even in peripheral vision, but small ones depend on a thorough inspection within the foveal region. So it should be straightforward to find big grains but rather hard to find smaller ones. However, if one suspects seeing a small grain, inspection should be

faster as examining two other sections might be enough. The FOR of this technique is composed of a set of two-dimensional sections.

DS technique presents two major improvements over S: higher spatial resolution and increased information density or FOR. The first is due to the ‘infinite sections’ allowed by the digital model. The second is brought by other visualization modes that allow seeing the remaining part of the sectioned grains (2.5D, 3D) beyond the section plane. The size may now be determined instantaneously or with less movement through other sections. If large objects can be detected by spotting a wide 2D section in S, in DS the confirmation may be quicker by leveraging depth information. In the case of the little grains, it is still necessary to spot them in the focal region, but confirmation is also potentially faster with depth information. The volumetric notion of 2.5D allows using the perceptual apparatus, which takes depth hints and stereo disparity into account. However, the vision is shallow (a single layer of grains), and the difference in depth may not be enough to bring positive effects, except maybe for very large grains. In fact, due to a slightly reduced FOV in stereoscopic modes, the information density is also reduced.

On GC, improvement comes as increased information density, brought by new features aiming to differentiate size extremes and by altering visualization logic on HR tasks, which pursued a scattering pattern detection rationale. In summary, in the case of grain size tasks, the main difference between techniques is the increased view depth achieved by changing the frontal grain layers to wireframe representations. This allows one to view up to four layers instantly. Larger grains can be detected by the absence or low edge density in a given area, while the smaller ones are characterized by a concentration of edges in a small region. GC is probably the only technique able to highlight small grains outside the foveal region. In the case of HR tasks, the FOR is almost complete, revealing almost the entire sample. The simultaneous use of wireframe and iconic representations of object centers also potentially improves the detection of clusters or empty space. The greater depth, in general, may also provide gains with 3D visualizations, as it should be possible to distinguish the depth planes better. However, the substantial increase in the FOR of this technique can potentially increase the reaction time (Score C) and the count of detections in the first appearance (Score D = 1), but it may also increase the relevance of global performance (Score A - ranking) and, mainly, of the transient performance (Score B), as more grains are visible at the same time. In fact, GC is the only technique that allows seeing a considerable portion of the sample at any time (from 30 to 100%), so the results must be judged from this perspective.

The analysis is more straightforward on the grain shape task. Rotation is not possible on the parallel sections of S technique. In DS’s case, being a sectioning technique, shape analysis requires rotation on the object’s central geometric axis. Then, deciding upon a shape factor requires finding at least two sections aligned to the sides of the grain’s

bounding box. On 2.5D modes, the possibility of looking beyond the section plane might ease finding those sides. 3D modes, however, may not provide an advantage due to the short depth on this task. Finally, as GC makes neighbors transparent, it is possible to visualize almost all the grain geometry from the outside by rotating around it, which should allow finding the representative sides of the shape faster. The longer observation distance of GC also increases view depth, which may yield positive effects on 3D modes.

Therefore, concerning tasks involving object size, all techniques except S aim to increase the FOR and the visual information density without overloading the perceptual system, while emphasizing the desired model properties. Therefore, the general objective was achieving the performance order  $GC > DS > S$  in every task, which aligns with the general scientific hypothesis, in view of the techniques' theoretical properties. Regarding visualization modes, a higher performance of 3D modes was expected, with the primacy of the split-screen mode, followed by 2.5D modes and finally by 2D modes. Among the 2.5D, AO mode was expected to excel due to the exploration of depth hints.

However, the results do not support this general hypothesis, which, in fact, can be considered very optimistic and perhaps simplistic. The exception was grain shape detection, where GC technique was much superior to DS, particularly in 3Ds mode. Probably the main success factors were the comprehensive view of the object, and the greater depth of view provided by 3Ds mode. Anaglyph was not better than 2.5D, so it seems that its three-dimensional quality is lower than 3Ds, not providing the same advantages. Another evidence that supports the advantages of GC is the difference in TLX results, which indicates less cognitive load. A more significant correlation with spatial skills was also expected, especially on the DS technique, due to the apparent need to perform mental operations. However, the correlation was not found. One possible explanation is the high average spatial skill shown by the public involved in the experiments, particularly in the MRT mental rotation test, with an average close to 8.5 (out of 10) and a minimum score above 5. Concerning other correlations, the strongest are related to movement and oscillation variables. This was expected as such variables are time-sensitive, thus, as the participant moves or oscillates more, the time response increases. Considering that GC's average time was considerably smaller than DS, the higher movement-time correlation on GC ( $0.5 \times 0.26$  in DS) probably means that GC induced more movement on average, which also reflected in the 0.2 correlation to oscillations, as more oscillations are expected with increased movement.

Concerning size tasks, the hypothesis was supported for small grain detection. GC was superior in general, especially in 3D modes, including anaglyph. The detection advantage is probably due to the visual emphasis given by the high density of edges, in addition to the greater differentiation between depth planes enabled by stereo vision. Anaglyph's limitations were not an issue as multiple colors do not influence much in

this scenario. The only negative impact was in score C, probably due to the higher FOR of the technique. However, the difference is just a few seconds and can be disregarded, considering that extreme performance is not a priority. Although the higher FOR may bring an artificial advantage to this technique in score D, it is still possible to compare modes within GC, placing 3Ds after 3Da. On the other hand, still considering the greater FOR, GC's superiority in score B over S and DS is even more remarkable since instantaneous detectability (spotting the best option at the moment) should be harder with more visible objects. Therefore, we can conclude that GC technique provided high detectability of small grains. Among DS modes, 3Ds stands out with the best main scores. A possible explanation is that the depth difference between a small grain and several larger ones in the same section plane may increase stereo's depth effect to create substantial highlight. Additionally, a subtle color influence was found in favor of DS in small grain detection, probably because color helps distinguishing small objects from each other - in some cases, monochrome may camouflage the difference between neighboring grains.

Considering large grains, however, DS was the best technique, followed closely by S, and GC comes in last. Score A in GC was eleven positions worse than DS and S, which is substantial. In fact, this supports the hypothesis that the detection of large grains might be determined mainly by the large section of the object within an even larger area. Additionally, the increased depth sense of 3D and 2.5D options of DS did not provide advantage over 2D. The superior score B of DS and S on this task also suggests that the broader FOV of the section view provides better comparison conditions by allowing simultaneous view of other large grains. On the other hand, the greater depth and the lower FOV of GC actually reduce the number of large grains simultaneously visible. Technique S was only worse to DS in score D, probably due to the larger jumps between sections, which can easily make a grain escape the first appearance. Still, within the GC technique, 3D versions were slightly superior to 2D in detectability as they highlight the volume of large grains. Another important fact is the indication of higher stress levels in 3D versions of GC, mainly anaglyph, probably due to ocular discomfort. Therefore, results of large grain detection are considered consistent.

The technique ranking  $DS > S > GC$  was also verified on HR tasks. In this case, GC has not performed as well as it did on the small grain task. The first notable difference of this task is the detection of groups of objects, not single objects. Probably the large-grain detection principle also applies to grain clusters, if they are interpreted as a single large object. Therefore, a cluster of small objects becomes a large composite object, and a cluster of large objects becomes an even larger composite object. Thus, in principle, DS and S techniques should excel in these cases. In fact, the order  $DS > S > GC$  is evident in the main scores. It should be noted, however, that GC has an almost complete FOR in this task, i.e., one sees almost the entire sample at all times, which might explain why score D is still high and higher than the other techniques overall, in addition to the worst

scores B and C. Given the large extent of regions, technique S ends up winning score C probably due to an agility advantage. S was worse than DS in score D, once more probably by virtue of the bigger jumps between sections, which can easily make the center of the region escape on the first appearance. In fact, a moderate negative correlation was found between movement/oscillations and score D in these tasks. The fundamental differences between small and bigger HR tasks are the mean difference in positions between GC and DS / S on score A (8 in HRb x 4 in HRs) and the lower performance of GC's 3D versions in score B. Despite the clear superiority of 2D here, DSa excels in these tasks. Still, 3D GC performed worse in score B.

Therefore, a 'large object' rationale seems to explain HR results, where section techniques proved to be superior to GC, and depth cues or stereoscopy did not present a positive advantage. In fact, when the composite object is smaller (small regions), the difference in the main score between GC and the others is reduced by half. In this matter, it is worth mentioning that GC performed particularly worse on score C of large regions. Experiment observation indicates that participants took a long time to find such regions. GC was indeed worse on all scores in this task, so this technique's perceptual scattering pattern is not efficient in highlighting large grain clusters.

Finally, ocular discomfort and VR sickness on 3D conditions deserve an in-depth discussion. It is known that movement in VR environments should ideally follow a natural paradigm, i.e., occur with minimal latency and concurrently with likewise body movements. However, participants remained seated and navigated the virtual scene using abstract keyboard commands, but VR sickness was seldom reported on 3Ds. A possible explanation, considering the GC technique, is the combination of higher immersion, a predominantly dark virtual environment, and low visual density. Anaglyph 3D, on the other hand, is less immersive, notably when participants performed tasks with smaller displays or in lighted rooms. This creates static visual anchors on the real environment, where 3D objects moving on the screen are more prone to cause motion sickness. The recommendation was executing tasks with a larger monitor in a very dark environment, which is perceptually similar to an HMD, where the dark virtual environment dissolves into the surrounding real environment. In contrast, the DS technique presents a bright virtual environment and the objects occupy all the field of view. However, the average 3D depth spans only a single grain layer, and the vision is similar to 2D versions, unlikely to cause issues. Therefore, in an immersive condition, the minor stereo disparity may not cause ocular stress or motion sickness. On the other hand, task execution with anaglyph goggles in lit real environments may still cause motion sickness.

## 6.1 Discussion summary

Summing up, the first relevant finding is that both sectioning and the use of wireframe, as advocated by many authors (ARTERO; OLIVEIRA, 2004; KREUSELER, 2000b) effectively reduced occlusion and cognitive overload by filtering information (ZHANG et al., 2012). TLX scores were between moderate and low on size and HR tasks and low on shape task. Participants were able to find small features in a relatively large sample, each representing less than 4% of the sample volume. HR demanded more effort, probably due to conjunctive search (PROCTOR; VU, 2009; WOLFE, 2000), as the participant should first detect clusters in the visual field and then decide which ones were either more (small grains) or less concentrated (big grains). In fact, score C or reaction time was exceptionally long for GC on those tasks.

Although the color factor was not tested on combinations of technique and visualization factors, the observed overall effect was minimal. Color affected mainly the size detection of small grains on technique DS, with a small advantage to the colored version, probably by assisting differentiation of small grains from their neighbors. Mono GC was slightly favored on scores B and C of bigger grains and C on smaller grains. No influence was found on HR tasks. Therefore, color suits best DS, while GC gains a slight advantage in monochromatic mode. This result also indicates that the chosen palette provided uniform intensity, avoiding attracting attention to stronger colors. Since color was strongly preferred by participants (84%), this palette can be adopted for similar tasks, particularly on sectioning visualization techniques, with possible performance gains on the detection of smaller objects.

Size was shown to be a better preattentive feature for highlighting bigger grains and clusters, given their relative size within the larger FOR of DS and S. However, all techniques allowed efficient detection of larger features, either grains or clusters, as suggested by studies (BILSKY; WOLFE, 1995; WOLFE, 2000). Large size features simply attract the eyes. This study verified this effect on bidimensional sections, tridimensional cells, and low-density open space. Given the dominance of DS and S over GC in larger features, it seems that the section was better than perspective viewing for spotting large features - the user can see much more grains at the same time in the section, whereas one can get lost in the void of a big grain in GC. However, considering heterogeneous regions, GC's difference in score A to DS/S was smaller (4 positions in the ranking) on small grain clustering, showing that a reduction in feature size is positive for GC in general. Therefore, it seems that such large features should not be too large relative to the FOV, as the size reference is lost. This global spatial reference seems to be more relevant when it is visible at the sides rather than at depth. In GC's case, the largest the feature, the larger the score C, going from 112 % on size task (large grains) to 460 on HRs, to 1870 % on HRb. However, despite the worse performance in those particular tasks, one should note that the larger FOR of this technique may suit other types of tasks and analyses not evaluated

in this work. Within the broader educational perspective, where absolute performance in specific tasks is not a priority, GC probably yields better overall spatial comprehension of this type of structure as the student can see the model to a larger extent and still detect the features studied in this work with acceptable performance. So it is argued that GC probably yields better long-range spatial comprehension, however, this should be further studied.

On the other hand, considering the detection of smaller grains, better depth perception afforded by preattentive features such as stereopsis (NAKAYAMA; SILVERMAN, 1986), depth cues (ANDERSEN, 1990; DOWNING; PINKER, 1985), perspective, lighting, and hierarchical structure (WOLFE, 1994), seems to provide superior feature highlighting. GC performed better, allowing finding grains four positions smaller on average. The technique also excelled at detectability with a superior score B even in a larger FOR situation, showing that small grains are effectively highlighted by this technique, especially in 3D visualization modes. Sectioning techniques were surprisingly even, owing probably to the distance between sections on S, which was small enough to find even the smaller grains. DS in 3D split screen also performed well for finding small grains, probably due to adequate stereo disparity between smaller grains in the frontal plane and the deeper surfaces of larger neighbors (HOLLIDAY; BRADDICK, 1991; HE; NAKAYAMA, 1992; HE; NAKAYAMA, 1994a; HE; NAKAYAMA, 1994b). Interestingly, in several situations, 2.5D and 3Da performed similarly, reinforcing other studies findings (LIVATINO *et al.*, 2015; SANFTMANN; WEISKOPF, 2011) that anaglyph lacks overall depth perception quality when compared to split view stereo. Typical anaglyph issues such as ghosting and discomfort (SANFTMANN; WEISKOPF, 2011) probably hindered the expected superiority over 2D and 2.5D renderings. Indeed, nausea and ocular discomfort were reported by participants on 3Da. Surprisingly, on the other hand, the expected VR sickness in 3Ds was not detected, despite the unnatural keyboard-based movement adopted on the experimental system, indicating that this simple and highly accessible interaction method can be used effectively. Accessibility is critical in educational contexts and, in this case, any cardboard-type adapter equipped with a standard smartphone would be enough to benefit from similar 3D visualizations.

The known influence of spatial ability on 3D search performance (CARROLL, 1993) was not detected in the main study, perhaps due to the absence of rotation on most tasks, no need for object position recalling, or to high participants' ability level - engineering students are expected to possess high spatial abilities (HUK, 2006; ALIAS, 2000b). Furthermore, tutorials guaranteed navigation proficiency and contact with the type of spatial structures experienced on the experiments and trained participants on detection criteria. Although preliminary experiments suggested a role of spatial ability on performance, MRT and SBSOD were not applied in those occasions and the experimental groups were more heterogeneous, comprising participants from distinct areas. In contrast,

on main experiments, they came mostly from the same engineering course and had similar backgrounds. Tasks and execution details were also distinct in preliminary experiments. By fixing the rotation on both size and HR tasks of the main experiment, most spatial awareness issues (GLUECK et al., 2009; MCCRAE et al., 2010b; ZHANG, 2008; ZHANG, 2009) are likely to be reduced as the sample's boundaries location remains fixed. Spatial ability seemed to play a minor role even at the shape task, that required rotation and imagining a bounding box around the inspected grain. Therefore, the influence of spatial ability should be further studied with experiments comprising participants with both high and low levels.

Finally, the effort devoted to visual design and aesthetics, as defended by some authors (BYRNE, 1996; RHYNE et al., 1999), probably helped to engage participants, channel attention to the tasks, and contribute to improved immersion (THON, 2008), and studies indicate that motivation can contribute to ease the apparent cognitive effort on such tasks (SZALMA; HANCOCK, 2009). In this matter, data showed a strong preference of 2.5D over 2D and also of color over monochromatic versions.



## 7 Conclusions

Three interactive tools were created to improve comprehension of grain microstructures and similar honeycomb structures. These tools were based on cognitive principles and the state of the art in interactive systems. A survey was performed on prospective interactive technologies, and WebGL was chosen for system development. A systematic review was carried out on multiscale visualization, considered an ideal paradigm in this work.

In a second stage, tool development focused on investigating perceptual performance. Instead of conducting the long-term experiments required to properly assess learning improvements, the approach focused on features deemed more influential to spatial grain structure comprehension, such as identifying size extremes, regions with localized size heterogeneities, and the shape of individual grains.

Tools were adapted for user studies, centered on three experimental tasks. Based on Human-computer interaction knowledge and the outcomes of two preliminary experiments involving more than 40 participants, four dependent variables were defined to attest performance, and experimental logistics were detailed, including forms and auxiliary tests, in addition to appropriate statistical methods to analyze mixed factorial experiments results. Two formal experiments were carried out with two groups of 30 participants each, comprising members from the intended target audience. Pre-tests indicated homogeneity of spatial skills between the groups. Participants were recruited among graduate and undergraduate engineering students of Escola Politecnica at the University of Sao Paulo, and experiments were approved by EACH (University of Sao Paulo) ethics committee (project code CAAE - 04436918.8.0000.5390).

The tools are Discrete Sections (S), Dynamic Sectioning (DS), and GrainCrawler (GC). S simulates a visualization based on sequences of parallel sections, performed over short distances on real materials. DS improves upon S on spatial resolution, and the user can cut through the model interactively in any direction. GC was developed after the model's peculiarities, perceptual potentials, and educational objectives. A multiscale visualization paradigm was chosen, where the representation of grains changes according to distance and perceptual objectives. Visualization modes were developed after known depth perception principles, leveraging the model's three-dimensional nature to improve search performance.

Essentially, the main improvements across techniques are the increase in visual information density or FOR and features to filter or highlight specific objects. Visual search theories better explain the results, and performance seems determined by a relationship between movement, FOR, FOV, and the depth perception mechanisms. The main attention elements are feature size, depth cues, stereopsis, and color may have influenced in

some cases. On tasks involving large grains and regions, the main preattentive factor was feature size, and performance was largely dependent on the broader FOV of S and DS, which allows visual size references in the same section even for the largest elements. The perspective view of GC does not provide such references, where large grains or regions may occupy the whole view. Despite the greater FOR of GC on small grain task, the increased depth perception influenced performance substantially, especially on 3D modes. GC also dominated the shape task, indicating that a comprehensive view of the object and a better sense of depth are important factors.

This research supported many findings of visual attention research, contributing to the particular case of space-filling structures. The main objective and hypothesis was that GC would surpass DS and especially S and that 3D modes would overcome 2D. However, the hypothesis was shown simplistic, and results were surprising in a positive way - all tools performed well overall and are deemed effective to improve spatial understanding of this type of structure. Interactive sectioning and multiscale visualization were successfully applied in detecting extreme sizes and heterogeneous regions in a space-filling model containing over a thousand objects, without explicit visual aids. However, each technique had specific strengths. S can be applied to real materials, given that sections are not widely spaced. This procedure can be readily executed with the current technology available in most universities' materials laboratories. The technique allowed agility on large grains and regions detection, which can be quite useful in visual analytics of such structures in industrial and academic environments, not requiring special software, sophisticated modeling or digital reconstructions. Additionally, other structures can be readily visualized and highlighted with standard image analysis software. DS is ideal for digital models, being the most balanced and flexible technique as it allows rotation of the section plane. Finally, GC excelled in small grain detection and shape analysis, characterized by high complexity and cognitive loads. Considering the broader educational objectives of this project, it is argued that GC probably yields better long-range spatial comprehension, despite the lower observed performance, particularly on regions tasks; however, this should be further studied. Additionally, the multiscale paradigm allows combining each tool's strengths according to the analytical needs of the intended application.

## 7.1 Limitations and future work

This work has several limitations. Microstructures have numerous features, but the scope was restricted to the grain structure, and even grains are simplified as polyhedral objects. Despite leaving the neighbor scale out of scope, HR results indicated that it is possible to detect spatial neighboring relationships and inspect a given grain's surroundings with interactive rotation. Other studies may assess spatial comprehension more thoroughly, particularly on the broader model scale.

This work was restricted to visualization aspects, while interaction factors were also mentioned, within an immersive continuum theoretical framework. A factorial study involving visualization and interaction requires many participants to obtain an adequate volume of data to conduct proper statistical analysis. Therefore, this continuum's exploration is suggested as future work, either focusing on interaction impacts or as an extensive factorial study. Such studies may employ modern VR systems to evaluate the impacts of fully natural interaction on spatial comprehension of similar space-filling models.

Further research can also evaluate the introduction of similar tools in materials science education. Stimulating features of gamification may also impact learning in educational settings. The same factors (sizes, heterogeneities, shape) can also be assessed on other individual elements of the grain structures, such as crystalline defects. Other analyses were also considered, such as feature dispersions and neighboring relationships. Finally, better tools can be developed and evaluated, combining the features studied here.

Other prospective multiscale visualization features mentioned in the manuscript deserve more studies, particularly spatial references and guided animations. Spatial references will probably increase performance and reduce cognitive load, especially for students with lower spatial abilities. Guided animation tools promise to empower teachers, leveraging their expert knowledge on didactic narratives that will guide students through specific parts of the model, building tailored educational experiences. The tool may provide control of the camera's properties and other objects, along with complementary textual information. Such tools may help teachers to readily experiment with the proposed visualizations by purposefully directing attention to specific aspects.

Finally, another aspect worth exploring on the HCI experimental field is adopting single or small N-subject designs on perceptual studies, potentially reducing participant needs to detect large effects and simplifying experimental logistics and results analysis. Gamification may also help improving participants' engagement in experiments.



# Bibliography

ABBASCHIAN, R.; REED-HILL, R. *Physical Metallurgy Principles*. Cengage Learning, 2008. ISBN 9780495082545. Disponível em: <<https://books.google.com.br/books?id=wh4v6UWjYdIC>>. Cited on page 6.

ALIAS, M. Spatial visualisation ability and problem solving in civil engineering. In: . [S.l.: s.n.], 2000. Cited 2 times on pages 3 and 32.

ALIAS, M. *Spatial Visualisation Ability and Problem Solving in Civil Engineering*. Tese (Doutorado) — University of Surrey, The address of the publisher, 1 2000. Thesis submitted for the degree of Doctor of Philosophy, School of Educational Studies, University of Surrey. Disponível em: <<http://epubs.surrey.ac.uk/id/eprint/957>>. Cited on page 119.

Alice - Geometry and Light. *Graphite*. 2020. Disponível em: <<http://alice.loria.fr/index.php/software/3-platform/22-graphite.html>>. Acesso em: 15 abr. 2020. Cited on page 24.

ANDERSEN, G. J. Focused attention in three-dimensional space. *Perception & Psychophysics*, v. 47, n. 2, p. 112–120, Mar 1990. ISSN 1532-5962. Disponível em: <<https://doi.org/10.3758/BF03205975>>. Cited 2 times on pages 12 and 119.

ARTERO, A. O.; OLIVEIRA, M. C. F. de. Viz3d: effective exploratory visualization of large multidimensional data sets. In: *Proceedings. 17th Brazilian Symposium on Computer Graphics and Image Processing*. [S.l.: s.n.], 2004. p. 340–347. Cited 2 times on pages 23 and 118.

BACIM, F.; BOWMAN, D.; PINHO, M. Wayfinding techniques for multiscale virtual environments. In: *2009 IEEE Symposium on 3D User Interfaces*. [S.l.: s.n.], 2009. p. 67–74. Cited 2 times on pages 17 and 22.

BARBOSA, J. F. V. *Ambientes urbanos virtuais para a web*. Dissertação (Mestrado) — Universidade do Porto, 7 2013. Disponível em: <<https://repositorio-aberto.up.pt/handle/10216/68494>>. Cited on page 26.

BARDELLA, F.; RODRIGUES, A. M.; Leal Neto, R. M. CrystalWalk: crystal structures, step by step. *Journal of Applied Crystallography*, v. 50, n. 3, p. 949–950, Jun 2017. Disponível em: <<https://doi.org/10.1107/S160057671700560X>>. Cited on page 5.

BARTLE, R. *Designing Virtual Worlds*. New Riders, 2004. (New Riders Games Series). ISBN 9780131018167. Disponível em: <<https://books.google.com.br/books?id=z3VP7MYKqaIC>>. Cited on page 46.

BILSKY, A. B.; WOLFE, J. M. Part—whole information is useful in visual search for size size but not orientation orientation conjunctions. *Perception & Psychophysics*, v. 57, n. 6, p. 749–760, Jan 1995. ISSN 1532-5962. Disponível em: <<https://doi.org/10.3758/BF03206791>>. Cited 2 times on pages 12 and 118.

BODNER, G. M.; GUAY, R. B. The purdue visualization of rotations test. *The Chemical Educator*, v. 2, n. 4, p. 1–17, Oct 1997. ISSN 1430-4171. Disponível em:

- <https://doi.org/10.1007/s00897970138a>. Cited 6 times on pages 11, 5, 9, 10, 32, and 88.
- BOSCA, A.; BONINO, D. Ontosphere3d: A multidimensional visualization tool for ontologies. In: *17th International Workshop on Database and Expert Systems Applications (DEXA '06)*. [S.l.: s.n.], 2006. p. 339–343. ISSN 2378-3915. Cited 2 times on pages 17 and 22.
- BOWMAN, D. A. et al. Information-rich virtual environments: Theory, tools, and research agenda. In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. New York, NY, USA: Association for Computing Machinery, 2003. (VRST '03), p. 81–90. ISBN 1581135696. Disponível em: <https://doi.org/10.1145/1008653.1008669>. Cited 2 times on pages 21 and 23.
- BREIVOLD, H. P. et al. Architecting for the cloud: A systematic review. In: *2014 IEEE 17th International Conference on Computational Science and Engineering*. [S.l.: s.n.], 2014. p. 312–318. Cited on page 26.
- BRYSON, S. Direct manipulation in virtual reality. *Visualization Handbook*, p. 413–430, 12 2005. Cited on page 23.
- BYRNE, C. M. *Water on Tap: The Use of Virtual Reality as an Educational Tool*. Tese (Doutorado) — University of Washington, 1996. Cited 4 times on pages 3, 7, 54, and 120.
- CARROLL, J. et al. *Minimalism Beyond the Nurnberg Funnel*. Cambridge, Mass., 1998. (Digital Communication Series). ISBN 9780262032490. Disponível em: <https://books.google.com.br/books?id=LvXiZJEUJjAC>. Cited on page 38.
- CARROLL, J. B. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. [S.l.]: Cambridge University Press, 1993. Cited 2 times on pages 9 and 119.
- CHANDLER, P.; SWELLER, J. Cognitive load theory and the format of instruction. *Cognition and Instruction*, Routledge, v. 8, n. 4, p. 293–332, 1991. Disponível em: [https://doi.org/10.1207/s1532690xc0804\\_2](https://doi.org/10.1207/s1532690xc0804_2). Cited on page 8.
- CHANG, R. et al. Legible cities: Focus-dependent multi-resolution visualization of urban relationships. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, USA, v. 13, n. 6, p. 1169–1175, nov. 2007. ISSN 1077-2626. Disponível em: <https://doi.org/10.1109/TVCG.2007.70574>. Cited on page 21.
- COCKBURN, A.; KARLSON, A.; BEDERSON, B. B. A review of overview+detail, zooming, and focus+context interfaces. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 41, n. 1, jan. 2009. ISSN 0360-0300. Disponível em: <https://doi.org/10.1145/1456650.1456652>. Cited on page 21.
- CORDEIL, M. et al. Assessing and improving 3d rotation transition in dense visualizations. In: *Proceedings of the 27th International BCS Human Computer Interaction Conference*. Swindon, GBR: BCS Learning & Development Ltd., 2013. (BCS-HCI '13). Cited 2 times on pages 21 and 22.
- CRASSIN, C. et al. Gigavoxels: ray-guided streaming for efficient and detailed voxel rendering. In: *In I3D '09: Proceedings of the 2009 symposium on Interactive 3D graphics and games*. [S.l.]: ACM, 2009. p. 15–22. Cited on page 1.

- DEHOFF, R. T. Microstructure: The realistic and quantitative description of microstructural evolution. *Metallography, Microstructure, and Analysis*, v. 1, n. 5, p. 244–258, Oct 2012. ISSN 2192-9270. Disponível em: <<https://doi.org/10.1007/s13632-012-0042-8>>. Cited 5 times on pages 3, 5, 6, 30, and 55.
- DENNIS, B. M.; HEALEY, C. G. Assisted navigation for large information spaces. In: *Proceedings of the Conference on Visualization '02*. USA: IEEE Computer Society, 2002. (VIS '02), p. 419–426. ISBN 0780374983. Cited on page 17.
- DOWNING, C.; PINKER, S. *The Spatial Structure of Visual Attention*. MIT, 1985. (MIT Industrial Liaison Program report). Disponível em: <<https://books.google.com.br/books?id=KB15GwAACAAJ>>. Cited 2 times on pages 12 and 119.
- DOŠKÁŘ, M.; NOVÁK, J. A jigsaw puzzle framework for homogenization of high porosity foams. *Computers & Structures*, v. 166, p. 33 – 41, 2016. ISSN 0045-7949. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0045794916000043>>. Cited on page 13.
- FISCHER, M. et al. 3d metallography of multiphase steels. In: \_\_\_\_\_. *Characterization of Minerals, Metals, and Materials*. John Wiley & Sons, Ltd, 2012. p. 175–184. ISBN 9781118371305. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118371305.ch22>>. Cited 2 times on pages 1 and 2.
- GLUECK, M. et al. Multiscale 3d reference visualization. In: *Proceedings of the 2009 Symposium on Interactive 3D Graphics and Games*. New York, NY, USA: ACM, 2009. (I3D '09), p. 225–232. ISBN 978-1-60558-429-4. Disponível em: <<http://doi.acm.org/10.1145/1507149.1507186>>. Cited 5 times on pages 17, 21, 22, 31, and 120.
- GODDARD, T. D.; HUANG, C. C.; FERRIN, T. E. Software extensions to ucsf chimera for interactive visualization of large molecular assemblies. *Structure*, v. 13, n. 3, p. 473 – 482, 2005. ISSN 0969-2126. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0969212605000572>>. Cited on page 17.
- GRIEBEL, M. et al. Flow field clustering via algebraic multigrid. In: *IEEE Visualization 2004*. [S.l.: s.n.], 2004. p. 35–42. Cited on page 21.
- GROEBER, M. A.; JACKSON, M. A. Dream.3d: A digital representation environment for the analysis of microstructure in 3d. *Integrating Materials and Manufacturing Innovation*, v. 3, n. 1, p. 56–72, Dec 2014. ISSN 2193-9772. Disponível em: <<https://doi.org/10.1186/2193-9772-3-5>>. Cited 3 times on pages 11, 24, and 25.
- GUIARD, Y. et al. View size and pointing difficulty in multi-scale navigation. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. New York, NY, USA: Association for Computing Machinery, 2004. (AVI '04), p. 117–124. ISBN 1581138679. Disponível em: <<https://doi.org/10.1145/989863.989881>>. Cited 2 times on pages 17 and 22.
- GYULASSY, A. et al. Visualization corner: Robust topology-based multiscale analysis of scientific data. *Computing in Science Engineering*, v. 11, n. 5, p. 88–95, Sep. 2009. ISSN 1558-366X. Cited on page 22.

HAMARI, J.; KOIVISTO, J.; SARSA, H. Does gamification work? – a literature review of empirical studies on gamification. In: *2014 47th Hawaii International Conference on System Sciences*. [S.l.: s.n.], 2014. p. 3025–3034. Cited 3 times on pages 46, 47, and 48.

HART, S. G.; STAVELAND, L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In: HANCOCK, P. A.; MESHKATI, N. (Ed.). *Human Mental Workload*. North-Holland, 1988, (Advances in Psychology, v. 52). p. 139 – 183. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0166411508623869>>. Cited 3 times on pages 80, 89, and 241.

HE, Z. J.; NAKAYAMA, K. Surfaces versus features in visual search. *Nature*, v. 359, n. 6392, p. 231–233, 1992. ISSN 1476-4687. Disponível em: <<https://doi.org/10.1038/359231a0>>. Cited 2 times on pages 12 and 119.

HE, Z. J.; NAKAYAMA, K. Perceived surface shape not features determines correspondence strength in apparent motion. *Vision Research*, v. 34, n. 16, p. 2125 – 2135, 1994. ISSN 0042-6989. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0042698994903220>>. Cited 2 times on pages 12 and 119.

HE, Z. J.; NAKAYAMA, K. Perceiving textures: Beyond filtering. *Vision Research*, v. 34, n. 2, p. 151 – 162, 1994. ISSN 0042-6989. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S004269899490328X>>. Cited 2 times on pages 12 and 119.

HEGARTY, M. et al. Development of a self-report measure of environmental spatial ability. *Intelligence*, v. 30, n. 5, p. 425 – 447, 2002. ISSN 0160-2896. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0160289602001162>>. Cited 3 times on pages 80, 88, and 237.

HENRY, J. A.; POLYS, N. F. The effects of immersion and navigation on the acquisition of spatial knowledge of abstract data networks. *Procedia Computer Science*, v. 1, n. 1, p. 1737 – 1746, 2010. ISSN 1877-0509. ICCS 2010. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1877050910001961>>. Cited on page 21.

HERTZBERG, R.; VINCI, R.; HERTZBERG, J. *Deformation and Fracture Mechanics of Engineering Materials, 5th Edition*. Wiley, 2012. ISBN 9781118213391. Disponível em: <<https://books.google.com.br/books?id=8d8bAAAAQBAJ>>. Cited 2 times on pages 5 and 6.

HILDEBRANDT, D.; TIMM, R. An assisting, constrained 3d navigation technique for multiscale virtual 3d city models. *GeoInformatica*, v. 18, n. 3, p. 537–567, Jul 2014. ISSN 1573-7624. Disponível em: <<https://doi.org/10.1007/s10707-013-0189-8>>. Cited 3 times on pages 17, 21, and 22.

HIRVASOJA, M. *Effects of computer game playing on spatial skills*. Dissertação (Mestrado) — University of Jyväskylä, 8 2004. Disponível em: <<http://urn.fi/URN:NBN:fi:ju-2004950711>>. Cited on page 10.

HOLLIDAY, I. E.; BRADDICK, O. J. Pre-attentive detection of a target defined by stereoscopic slant. *Perception*, v. 20, n. 3, p. 355–362, 1991. PMID: 1762877. Disponível em: <<https://doi.org/10.1068/p200355>>. Cited 2 times on pages 12 and 119.



- HORNBOGEN, E. A systematic description of microstructure. *Journal of Materials Science*, v. 21, n. 11, p. 3737–3747, Nov 1986. ISSN 1573-4803. Disponível em: <<https://doi.org/10.1007/BF00553424>>. Cited 5 times on pages 3, 5, 6, 30, and 55.
- HOUTGAST, E.; PFEIFFER, O.; POST, F. Navigation and interaction in a multi-scale stereoscopic environment. In: *Proceedings of the 2005 IEEE Conference 2005 on Virtual Reality*. Washington, DC, USA: IEEE Computer Society, 2005. (VR '05), p. 275–276. ISBN 0-7803-8929-8. Disponível em: <<https://doi.org/10.1109/VR.2005.56>>. Cited 3 times on pages 22, 23, and 31.
- HSU, W.-H.; MA, K.-L.; CORREA, C. A rendering framework for multiscale views of 3d models. *ACM Trans. Graph.*, Association for Computing Machinery, New York, NY, USA, v. 30, n. 6, p. 1–10, dez. 2011. ISSN 0730-0301. Disponível em: <<https://doi.org/10.1145/2070781.2024165>>. Cited on page 21.
- HUK, T. Who benefits from learning with 3d models? the case of spatial ability. *Journal of Computer Assisted Learning*, v. 22, n. 6, p. 392–404, 2006. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2729.2006.00180.x>>. Cited 3 times on pages 3, 32, and 119.
- IKEDA, M.; TAKEUCHI, T. Influence of foveal load on the functional visual field. *Perception & Psychophysics*, v. 18, n. 4, p. 255–260, Jul 1975. ISSN 1532-5962. Disponível em: <<https://doi.org/10.3758/BF03199371>>. Cited on page 12.
- INSLEY, J. A.; GRINBERG, L.; PAPKA, M. E. Visualizing multiscale, multiphysics simulation data: Brain blood flow. In: *2011 IEEE Symposium on Large Data Analysis and Visualization*. [S.l.: s.n.], 2011. p. 3–7. Cited on page 21.
- JOY, K. I. Massive data visualization: A survey. In: \_\_\_\_\_. *Mathematical Foundations of Scientific Visualization, Computer Graphics, and Massive Data Exploration*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 285–302. ISBN 978-3-540-49926-8. Disponível em: <[https://doi.org/10.1007/b106657\\_15](https://doi.org/10.1007/b106657_15)>. Cited on page 13.
- KAUFMAN, A.; MUELLER, K. 7 overview of volume rendering. In: \_\_\_\_\_. *Visualization Handbook*. [s.n.], 2005. p. 127–xi. Disponível em: <<https://app.dimensions.ai/details/publication/pub.1043361602>>. Cited on page 1.
- KIM, A. J. *Game Thinking: Innovate smarter & drive deep engagement with design techniques from hit games*. Burlingame: gamethinking.io, 2018. ISBN 9780999788547. Cited on page 46.
- KITCHENHAM, B.; CHARTERS, S. *Guidelines for performing systematic literature reviews in software engineering*. Staffordshire, 2007. 57 p. Cited on page 13.
- KOPPER, R. et al. Design and evaluation of navigation techniques for multiscale virtual environments. In: *IEEE Virtual Reality Conference (VR 2006)*. [S.l.: s.n.], 2006. p. 175–182. ISSN 2375-5334. Cited 2 times on pages 17 and 21.
- KRAUSE, S. et al. Identifying student misconceptions in introductory materials engineering classes. In: *ASEE Annual Conference Proceedings*. [S.l.: s.n.], 2003. p. 3753–3760. Cited on page 32.

KREUSELER, M. Visualization of geographically related multidimensional data in virtual 3d scenes. *Comput. Geosci.*, Pergamon Press, Inc., USA, v. 26, n. 1, p. 101–108, fev. 2000. ISSN 0098-3004. Disponível em: <[https://doi.org/10.1016/S0098-3004\(99\)00036-9](https://doi.org/10.1016/S0098-3004(99)00036-9)>. Cited on page 21.

KREUSELER, M. Visualization of geographically related multidimensional data in virtual 3d scenes. *Comput. Geosci.*, Pergamon Press, Inc., Tarrytown, NY, USA, v. 26, n. 1, p. 101–108, fev. 2000. ISSN 0098-3004. Disponível em: <[http://dx.doi.org/10.1016/S0098-3004\(99\)00036-9](http://dx.doi.org/10.1016/S0098-3004(99)00036-9)>. Cited on page 118.

KURI, N. P.; SILVA, A. N. R. d.; PEREIRA, M. d. A. Estilos de aprendizagem e recursos da hipermídia aplicados no ensino de planejamento de transportes. *Revista Portuguesa de Educação*, Centro de Investigação em Educação (CIEd), Instituto de Educação da Universidade Minho, v. 19, n. 2, p. 111–137, 2006. ISSN 0871-9187. Cited on page 32.

LANZAGORTA, M. et al. Three-dimensional visualization of microstructures. In: *Proceedings Visualization '98 (Cat. No.98CB36276)*. [S.l.: s.n.], 1998. p. 487–490. Cited 8 times on pages 11, 1, 2, 3, 7, 24, 39, and 72.

LAVIOLA, J. J. et al. Hands-free multi-scale navigation in virtual environments. In: *Proceedings of the 2001 Symposium on Interactive 3D Graphics*. New York, NY, USA: Association for Computing Machinery, 2001. (I3D '01), p. 9–15. ISBN 1581132921. Disponível em: <<https://doi.org/10.1145/364338.364339>>. Cited on page 22.

LAW, N. Scaffolding scientific conceptualization: multiple representation and multilevel visualization using an iconic modeling tool. In: *International Conference on Computers in Education, 2002. Proceedings*. [S.l.: s.n.], 2002. p. 257–261 vol.1. Cited on page 21.

LI, Y.; FU, C.; HANSON, A. Scalable wim: Effective exploration in large-scale astrophysical environments. *IEEE Transactions on Visualization and Computer Graphics*, v. 12, n. 5, p. 1005–1012, Sep. 2006. Cited 2 times on pages 17 and 21.

LIVATINO, S. et al. Stereoscopic visualization and 3-d technologies in medical endoscopic teleoperation. *IEEE Transactions on Industrial Electronics*, v. 62, n. 1, p. 525–535, Jan 2015. Cited 2 times on pages 81 and 119.

LUBOSCHIK, M. et al. Heterogeneity-based guidance for exploring multiscale data in systems biology. In: *2012 IEEE Symposium on Biological Data Visualization (BioVis)*. [S.l.: s.n.], 2012. p. 33–40. Cited 2 times on pages 21 and 22.

MARTIN, D. *Doing Psychology Experiments*. Cengage Learning, 2007. ISBN 9780495115779. Disponível em: <<https://books.google.com.br/books?id=QdWPr0UJ8gC>>. Cited 7 times on pages 55, 76, 83, 84, 89, 91, and 95.

MCCRAE, J. et al. Exploring the design space of multiscale 3d orientation. In: *Proceedings of the International Conference on Advanced Visual Interfaces*. New York, NY, USA: ACM, 2010. (AVI '10), p. 81–88. ISBN 978-1-4503-0076-6. Disponível em: <<http://doi.acm.org/10.1145/1842993.1843008>>. Cited 2 times on pages 21 and 23.

MCCRAE, J. et al. Exploring the design space of multiscale 3d orientation. In: *Proceedings of the International Conference on Advanced Visual Interfaces*. New York, NY, USA: ACM, 2010. (AVI '10), p. 81–88. ISBN 978-1-4503-0076-6. Disponível em: <<http://doi.acm.org/10.1145/1842993.1843008>>. Cited on page 120.

- MCCRAE, J. et al. Multiscale 3d navigation. In: *Proceedings of the 2009 Symposium on Interactive 3D Graphics and Games*. New York, NY, USA: Association for Computing Machinery, 2009. (I3D '09), p. 7–14. ISBN 9781605584294. Disponível em: <<https://doi.org/10.1145/1507149.1507151>>. Cited 4 times on pages 11, 13, 17, and 21.
- MCFARLANE, N. J. B. et al. 3d multiscale visualisation for medical datasets. In: *2008 Fifth International Conference BioMedical Visualization: Information Visualization in Medical and Biomedical Informatics*. [S.l.: s.n.], 2008. p. 47–52. Cited 2 times on pages 17 and 23.
- MCFARLANE, N. J. B. et al. A survey and classification of visualisation in multiscale biomedical applications. In: *2012 16th International Conference on Information Visualisation*. [S.l.: s.n.], 2012. p. 561–566. ISSN 1550-6037. Cited 4 times on pages 14, 17, 22, and 23.
- MCMAHAN, R. P. *Exploring the Effects of Higher-Fidelity Display and Interaction for Virtual Reality Games*. Tese (Doutorado) — Virginia Tech, 12 2011. Disponível em: <<http://hdl.handle.net/10919/30123>>. Cited on page 53.
- MEEHAN, M. et al. Effect of latency on presence in stressful virtual environments. In: *IEEE Virtual Reality, 2003. Proceedings*. [S.l.: s.n.], 2003. p. 141–148. Cited on page 81.
- MIAO, H. et al. Multiscale visualization and scale-adaptive modification of dna nanostructures. *IEEE Transactions on Visualization and Computer Graphics*, v. 24, n. 1, p. 1014–1024, Jan 2018. ISSN 2160-9306. Cited on page 16.
- MIAO, H. et al. Multiscale molecular visualization. *Journal of Molecular Biology*, v. 431, n. 6, p. 1049 – 1070, 2019. ISSN 0022-2836. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022283618310490>>. Cited on page 16.
- NAKAYAMA, K.; SILVERMAN, G. H. Serial and parallel processing of visual feature conjunctions. *Nature*, v. 320, n. 6059, p. 264–265, 1986. ISSN 1476-4687. Disponível em: <<https://doi.org/10.1038/320264a0>>. Cited 2 times on pages 12 and 119.
- NEWELL, A.; ROSENBLOOM, P. S. The soar papers (vol. 1). In: ROSENBLOOM, P. S.; LAIRD, J. E.; NEWELL, A. (Ed.). Cambridge, MA, USA: MIT Press, 1993. cap. Mechanisms of Skill Acquisition and the Law of Practice, p. 81–135. ISBN 0-262-68071-8. Disponível em: <<http://dl.acm.org/citation.cfm?id=162580.162586>>. Cited 2 times on pages 8 and 95.
- NUNES, E. P. d. S. *Um método para avaliar a aquisição de conhecimento em Ambientes Virtuais de Aprendizagem Tridimensionais Interativos*. Tese (Doutorado) — Universidade de São Paulo, 3 2014. Disponível em: <[https://teses.usp.br/teses/disponiveis/3/3141/tde-26122014-154342/publico/Tese\\_Eunice\\_Nunes.pdf](https://teses.usp.br/teses/disponiveis/3/3141/tde-26122014-154342/publico/Tese_Eunice_Nunes.pdf)>. Cited on page 7.
- PARISI, T. *WebGL: Up and Running: Building 3D Graphics for the Web*. O'Reilly Media, 2012. ISBN 9781449326517. Disponível em: <<https://books.google.com.br/books?id=Eip9ZrdGtEIC>>. Cited on page 26.
- PASEWALDT, S.; TRAPP, M.; DÖLLNER, J. Multiscale visualization of 3d geovirtual environments using view-dependent multi-perspective views. *Journal of WSCG*, v. 19, p. 111–118, 01 2011. Cited 2 times on pages 17 and 23.

- PATTRASITIDECHA, A. *Comparison and evaluation of 3D mobile game engines*. Dissertação (Mestrado) — University of Gothenburg, 2014. Disponível em: <https://hdl.handle.net/20.500.12380/193979>. Cited on page 25.
- PETRIDIS, P. et al. Game engines selection framework for high-fidelity serious applications. *International Journal of Interactive Worlds*, IBIMA Publishing, v. 2012, p. 418638, 2012. ISSN 2165-9508. Cited on page 25.
- PIETRIGA, E.; APPERT, C.; BEAUDOUIN-LAFON, M. Pointing and beyond: An operationalization and preliminary evaluation of multi-scale searching. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2007. (CHI '07), p. 1215–1224. ISBN 978-1-59593-593-9. Disponível em: <http://doi.acm.org/10.1145/1240624.1240808>. Cited on page 22.
- PINDAT, C. et al. Jellylens: Content-aware adaptive lenses. In: *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*. New York, NY, USA: ACM, 2012. (UIST '12), p. 261–270. ISBN 978-1-4503-1580-7. Disponível em: <http://doi.acm.org/10.1145/2380116.2380150>. Cited on page 21.
- PROCTOR, R. W.; VU, K.-P. L. Human information processing: An overview for human–computer interaction. In: SEARS, A.; JACKO, J. (Ed.). *Human-Computer Interaction Fundamentals*. New York: Erlbaum Taylor and Francis Group, 2009. p. 35–81. Disponível em: <https://doi.org/10.1201/b10368>. Cited 5 times on pages 7, 8, 11, 12, and 118.
- PURVES, D. *Neuroscience*. Oxford University Press, 2012. ISBN 9780878936953. Disponível em: <https://books.google.com.br/books?id=B5YXRAAACAAJ>. Cited 2 times on pages 8 and 11.
- QUEY, R.; DAWSON, P.; BARBE, F. Large-scale 3d random polycrystals for the finite element method: Generation, meshing and remeshing. *Computer Methods in Applied Mechanics and Engineering*, v. 200, n. 17-20, p. 1729–1745, 2011. Cited 3 times on pages 11, 24, and 25.
- RHYNE, T.-M. et al. Visualization needs more visual design! (panel session). In: SWAN II, J. E. (Ed.). *Proceedings of the Conference on Visualization '99: Celebrating Ten Years*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1999. (VIS '99), p. 485–490. ISBN 0-7803-5897-X. Disponível em: <http://dl.acm.org/citation.cfm?id=319351.319461>. Cited 3 times on pages 2, 7, and 120.
- RICARDO CABELLO. *three.js – JavaScript 3D library*. 2020. Disponível em: <https://threejs.org/>. Acesso em: 15 abr. 2020. Cited on page 33.
- RODRIGUES, A. M. et al. Integrated approach for geometric modeling and interactive visual analysis of grain structures. *Computer-Aided Design*, v. 97, p. 1 – 14, 2018. ISSN 0010-4485. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0010448517302014>. Cited 3 times on pages 12, 33, and 51.
- RODRIGUES, A. M. et al. Batman beyond: Natural 3d manipulation with the batwand. In: *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. [S.l.: s.n.], 2017. p. 258–259. Cited 7 times on pages 3, 5, 13, 32, 33, 40, and 94.

SANFTMANN, H.; WEISKOPF, D. Anaglyph stereo without ghosting. In: *Proceedings of the Twenty-second Eurographics Conference on Rendering*. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2011. (EGSR '11), p. 1251–1259. Disponível em: <<http://dx.doi.org/10.1111/j.1467-8659.2011.01984.x>>. Cited 2 times on pages 35 and 119.

SESTER, M.; BRENNER, C. A vocabulary for a multiscale process description for fast transmission and continuous visualization of spatial data. *Comput. Geosci.*, Pergamon Press, Inc., USA, v. 35, n. 11, p. 2177–2184, nov. 2009. ISSN 0098-3004. Disponível em: <<https://doi.org/10.1016/j.cageo.2008.11.003>>. Cited on page 17.

SHARMA, A. et al. Large multidimensional data visualization for materials science. *Computing in Science Engineering*, v. 5, n. 2, p. 26–33, March 2003. ISSN 1558-366X. Cited on page 21.

SMELIK, R. M. et al. A survey on procedural modelling for virtual worlds. *Computer Graphics Forum*, v. 33, n. 6, p. 31–50, 2014. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12276>>. Cited on page 13.

SOBOYEJO, W. *Mechanical Properties of Engineered Materials*. CRC Press, 2002. (Mechanical engineering). ISBN 9780203910399. Disponível em: <[https://books.google.com.br/books?id=jcD\\\_K-WOS1kC](https://books.google.com.br/books?id=jcD\_K-WOS1kC)>. Cited 2 times on pages 5 and 32.

STOLTE, C.; TANG, D.; HANRAHAN, P. Multiscale visualization using data cubes. *IEEE Transactions on Visualization and Computer Graphics*, v. 9, n. 2, p. 176–187, April 2003. ISSN 2160-9306. Cited on page 22.

SZALMA, J. L.; HANCOCK, P. A. Task loading and stress in human–computer interaction: Theoretical frameworks and mitigation strategies. In: SEARS, A.; JACKO, J. (Ed.). *Human-Computer Interaction Fundamentals*. New York: Erlbaum Taylor and Francis Group, 2009. p. 92–110. Disponível em: <<https://doi.org/10.1201/b10368>>. Cited 2 times on pages 9 and 120.

TAGKOPOULOS, L. Evevis: A multi-scale visualization system for dense evolutionary data. In: *Proceedings of the 2011 IEEE Symposium on Biological Data Visualization*. USA: IEEE Computer Society, 2011. (BIOVIS '11), p. 143–150. ISBN 9781467300032. Disponível em: <<https://doi.org/10.1109/BioVis.2011.6094059>>. Cited on page 17.

THON, J.-N. Immersion revisited: On the value of a contested concept. In: LEINO, O.; WIRMAN, H.; FERNANDEZ, A. (Ed.). *Extending Experiences: Structure, Analysis and Design of Computer Game Player Experience*. Lapland: Lapland University Press, 2008. p. 29–43. ISBN 9789524841979. Disponível em: <<https://books.google.com.br/books?id=n4bfwAEACAAJ>>. Cited 2 times on pages 53 and 120.

TREISMAN, A. M.; GELADE, G. A feature-integration theory of attention. *Cognitive Psychology*, v. 12, n. 1, p. 97 – 136, 1980. ISSN 0010-0285. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0010028580900055>>. Cited on page 12.

TRINDADE, D. R.; RAPOSO, A. B. Improving 3d navigation techniques in multiscale environments: A cubemap-based approach. *Multimedia Tools Appl.*, Kluwer Academic

Publishers, USA, v. 73, n. 2, p. 939–959, nov. 2014. ISSN 1380-7501. Disponível em: <<https://doi.org/10.1007/s11042-012-1127-8>>. Cited 3 times on pages 17, 21, and 23.

ULLAH, A. et al. Optimal approach of three-dimensional microstructure reconstructions and visualizations. *Materials Express*, v. 3, n. 2, p. 109–118, 2013. ISSN 2158-5849. Disponível em: <<https://www.ingentaconnect.com/content/asp/me/2013/00000003/00000002/art00002>>. Cited on page 1.

UNDERWOOD, E. E. Quantitative stereology for microstructural analysis. In: \_\_\_\_\_. *Microstructural Analysis: Tools and Techniques*. Boston, MA: Springer US, 1973. p. 35–66. ISBN 978-1-4615-8693-7. Disponível em: <[https://doi.org/10.1007/978-1-4615-8693-7\\_3](https://doi.org/10.1007/978-1-4615-8693-7_3)>. Cited on page 1.

VALIATI, E. R. A.; PIMENTA, M. S.; FREITAS, C. M. D. S. A taxonomy of tasks for guiding the evaluation of multidimensional visualizations. In: *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*. New York, NY, USA: Association for Computing Machinery, 2006. (BELIV '06), p. 1–6. ISBN 1595935622. Disponível em: <<https://doi.org/10.1145/1168149.1168169>>. Cited on page 23.

VANACKEN, L.; GROSSMAN, T.; CONINX, K. Exploring the effects of environment density and target visibility on object selection in 3d virtual environments. In: *2007 IEEE Symposium on 3D User Interfaces*. [S.l.: s.n.], 2007. Cited on page 22.

VANDENBERG, S. G.; KUSE, A. R. Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, v. 47, n. 2, p. 599–604, 1978. PMID: 724398. Disponível em: <<https://doi.org/10.2466/pms.1978.47.2.599>>. Cited 3 times on pages 80, 88, and 219.

VICECONTI, M. et al. Multimodal fusion of biomedical data at different temporal and dimensional scales. *Computer Methods and Programs in Biomedicine*, v. 102, n. 3, p. 227 – 237, 2011. ISSN 0169-2607. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0169260710001148>>. Cited on page 17.

WAERNÉR, K. *3D graphics technologies for web applications: An evaluation from the perspective of a real world application*. Dissertação (Mestrado) — Linköping University, 5 2012. Disponível em: <<http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Aliu%3Adiva-78726>>. Cited on page 25.

WILLIAMS, L. J. Tunnel vision induced by a foveal load manipulation. *Human Factors*, v. 27, n. 2, p. 221–227, 1985. PMID: 4018814. Disponível em: <<https://doi.org/10.1177/001872088502700209>>. Cited on page 12.

WILLIAMS, L. J. Foveal load affects the functional field of view. *Human Performance*, Routledge, v. 2, n. 1, p. 1–28, 1989. Disponível em: <[https://doi.org/10.1207/s15327043hup0201\\_1](https://doi.org/10.1207/s15327043hup0201_1)>. Cited on page 12.

WOBBROCK, J. O.; KAY, M. Nonparametric statistics in human–computer interaction. In: \_\_\_\_\_. *Modern Statistical Methods for HCI*. Cham: Springer International Publishing, 2016. p. 135–170. ISBN 978-3-319-26633-6. Disponível em: <[https://doi.org/10.1007/978-3-319-26633-6\\_7](https://doi.org/10.1007/978-3-319-26633-6_7)>. Cited 2 times on pages 92 and 93.

- WOLFE, J. M. Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, v. 1, n. 2, p. 202–238, Jun 1994. ISSN 1531-5320. Disponível em: <<https://doi.org/10.3758/BF03200774>>. Cited 2 times on pages 12 and 119.
- WOLFE, J. M. Chapter 8 - visual attention. In: VALOIS, K. K. D. (Ed.). *Seeing*. San Diego: Academic Press, 2000, (Handbook of Perception and Cognition). p. 335 – 386. ISBN 978-0-12-443760-9. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B9780124437609500106>>. Cited 3 times on pages 11, 12, and 118.
- ZHANG, X. Space-scale animation: enhancing cross-scale understanding of multiscale structures in multiple views. In: *Coordinated and Multiple Views in Exploratory Visualization (CMV'05)*. [S.l.: s.n.], 2005. p. 109–120. Cited 2 times on pages 17 and 21.
- ZHANG, X. A multiscale progressive model on virtual navigation. *International Journal of Human-Computer Studies*, v. 66, n. 4, p. 243 – 256, 2008. ISSN 1071-5819. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1071581907001383>>. Cited 4 times on pages 17, 21, 23, and 120.
- ZHANG, X. L. Multiscale traveling: Crossing the boundary between space and scale. *Virtual Real.*, Springer-Verlag, Berlin, Heidelberg, v. 13, n. 2, p. 101–115, jun. 2009. ISSN 1359-4338. Disponível em: <<https://doi.org/10.1007/s10055-009-0114-5>>. Cited 2 times on pages 17 and 120.
- ZHANG, X. L.; FURNAS, G. W. Multiscale space and place. In: \_\_\_\_\_. *Spaces, Spatiality and Technology*. Dordrecht: Springer Netherlands, 2005. p. 261–280. ISBN 978-1-4020-3273-8. Disponível em: <[https://doi.org/10.1007/1-4020-3273-0\\_18](https://doi.org/10.1007/1-4020-3273-0_18)>. Cited on page 22.
- ZHANG, X. L. et al. Supporting knowledge exploration and discovery in multi-dimensional data with interactive multiscale visualisation. *Journal of Engineering Design*, Taylor & Francis, v. 23, n. 1, p. 23–47, 2012. Disponível em: <<https://doi.org/10.1080/09544828.2010.487260>>. Cited 2 times on pages 22 and 118.





# Appendix



## APPENDIX A – Detailed Results

Each task analysis is divided into four scores (A, B, C, D). Score analysis follows this sequence:

- characterization of the experimental group with a table, showing experimental group size per technique, mode and outliers;
- the profile of responses distribution and the chosen statistical analysis method (survival analysis, GLMM or ART). Plots of distributions are provided (overall, DS technique and GC technique);
- results of statistical tests, listing significant overall and pairwise comparisons ( $p < 0.05$ );
- comparisons between techniques, showing relative differences in percentage (table);
- comparisons between color modes, showing relative differences in percentage (table);
- pairwise comparisons between pairs (combinations of techniques and visualization modes), showing relative differences in percentage (table);
- supplementary analysis, specific to each score class: score A counts (%) by score A stratum; Counts (% of B=10) per technique, visualization mode, and color mode; score C mean by score A stratum; Counts D=1 (%) per technique, visualization modes and per score A stratum. Refer to section [4.6.9](#) for more details about score A strata;
- supplementary correlation analysis (Pearson coefficient) of the score with selected variables: selection time, movement, and oscillations before selection (table);
- supplementary correlation analysis (Pearson coefficient) of the score with spatial ability tests (MRT and SBSOD) and TLX task load test (table).

It is important to notice that relative percentage differences are only shown for significant comparisons (tables) and includes both calculations (A/B and B/A). All comparisons are assisted by boxplots. For the sake of simplicity, the S technique was analysed as a pair or visualization mode. The term DS represents the 2D flat visualization mode on pairwise comparison of pairs.

## A.1 Size task analysis: bigger grains (Sb)

### A.1.1 Score A

#### A.1.1.1 Distributions and outliers

The sample size is 1058 choices for DS (34 per participant) and 1256 for GC (39 per participant), with a total of 197 outliers found (8.5%, see table 20). Concerning modes, sample sizes are almost equal per technique (232 per mode on DS and 297 on GC), differing about 22%, which is not a concern for statistical analysis.

Data distributions clearly deviate from normality and appears to follow the Weibull distribution (Figure 68, top). Therefore this score was analyzed using a GLMM, adopting the Weibull distribution family.

Table 20 – Sb: group size and outliers on score A

Technique	N	N (per mode)	Outliers
DS	1058	232	130
GC	1256	297	67

#### A.1.1.2 Boxplots, ANOVAs and comparisons

GLMM ANOVA results were significant for techniques ( $p < 0.01$ ) and pairs ( $p < 0.01$ ). Indeed, boxplots (Figure 68, middle) show noticeable differences between Techniques (Table 21) and between GC and DS Pairs. Inspecting pairwise comparisons (Table 23), DS and S are almost equally dissimilar than all GC pairs (about 76 % less), and differences among GC pairs are negligible (4-8 %). Differences in Color modes on DS and GC were not significant, and the omnibus difference in color is only due to the difference between techniques (Table 22).

Therefore we can conclude that DS and S have a significantly smaller score A than GC, and color played no role in this score.

Table 21 – Sb: significant differences (%) per technique on score A

Technique	Mean	s.d.	Difference
DS	3,45	2,86	-76,5
GC	14,73	12,02	325,9

#### A.1.1.3 Strata analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing how choices are distributed between nine strata (Score A < 2, between 2 and 4, 4-10, 10-20, 20-30, 30-40, 40-50, 50-60, higher than 60). In this case, most choices should be within the first strata ( $A < 10$ ). By examining techniques (Figure 69, top), S and DS

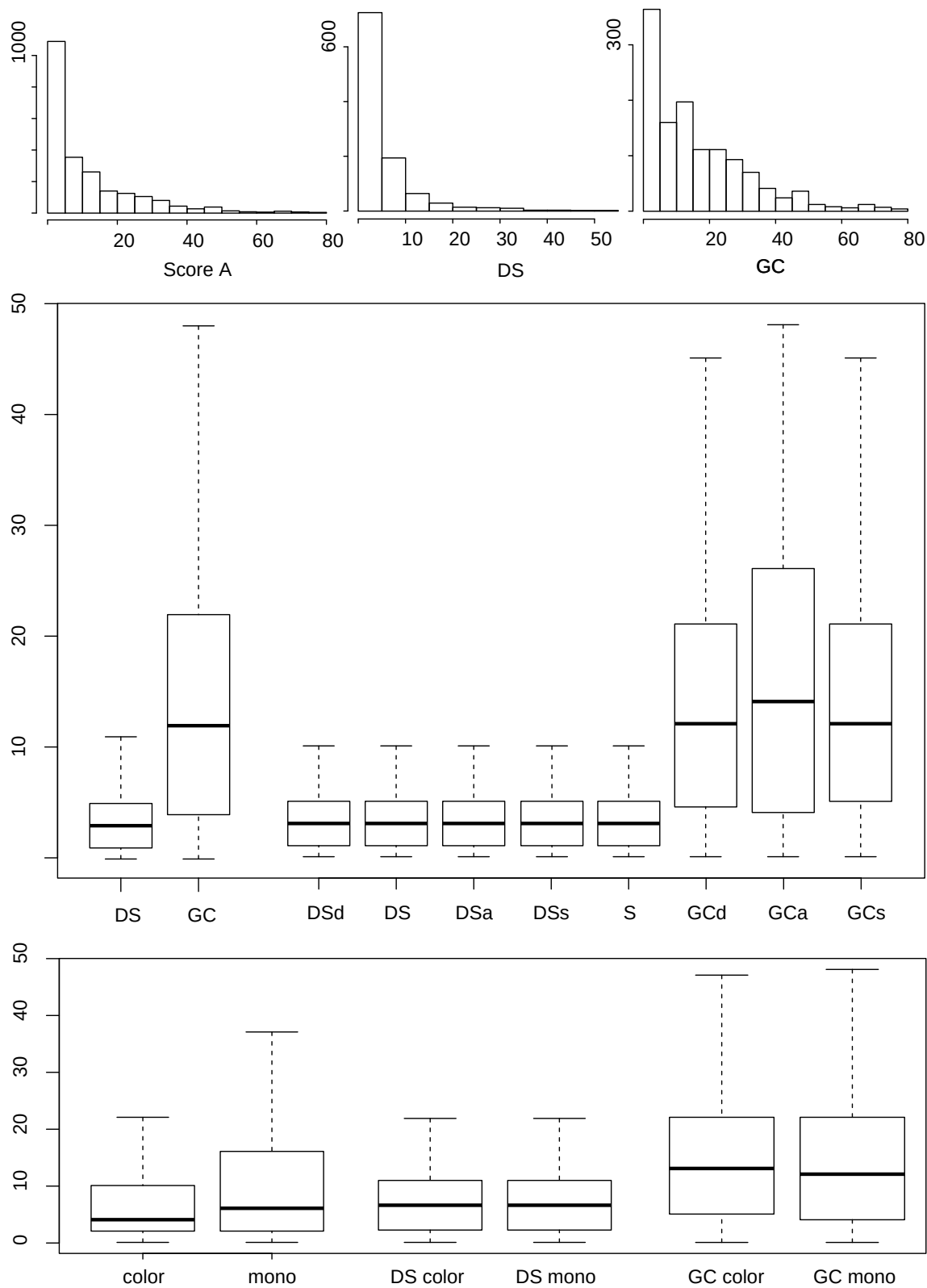


Figure 68 – Score A. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

Table 22 – Sb: significant differences (%) per Color mode on score A

Color mode	Mean	s.d.	Difference
DS Color	3,45	2,88	
DS Mono	3,46	2,85	
GC Color	15,22	12,09	
GC Mono	14,53	12	
Color	8,15	9,82	-25,1
Mono	10,88	11,24	33,5

Table 23 – Sb: significant differences (%) between pairs on score A

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	3,36	2,87	0	-	-	-	-	-	358,6	321,5
DS	3,4	2,66	-	0	-	-	-	-	352,3	315,7
DSd	3,57	3,02	-	-	0	-	-	-	330,9	296
DSa	3,41	2,82	-	-	-	0	-	-	351,4	314,9
DSs	3,61	2,94	-	-	-	-	0	-	-	-
GCd	14,71	11,97	-	-	-	-	-	0	4,78	-3,68
GCa	15,42	12,53	-78,1	-77,9	-76,8	-77,8	-	-4,56	0	-8,08
GCs	14,17	11,72	-76,2	-75,9	-74,7	-75,9	-	3,82	8,79	0

remain quite similar, while GC is less efficient precisely at the first three strata ( $A < 10$ ). Visualization modes performed equally on the DS technique (Figure 69, middle) while on GC, GCs was better on 4-10 stratum then GCd on the 10-20 stratum.

#### A.1.1.4 Correlations

Table 24 shows that only a few weak correlations ( $r < 0.2$ ) were found between score A and selected variables (Time, Movement, Oscillations) or pre/post tests scores (MRT, SBSOD and TLX).

Table 24 – Sb: correlations of other variables with score A

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	-	0.1	0.1	0.05	-	0.06
DS	-0.08	-0.1	-	-	-	-
GC	-	-	-	-	-	-
S	-	-	-	-	-	-

##### A.1.1.4.1 Score A summary

Considering all the evidence about score A, the following order (descending performance) is suggested: DS and S > GCs >= GCd >= GCa.

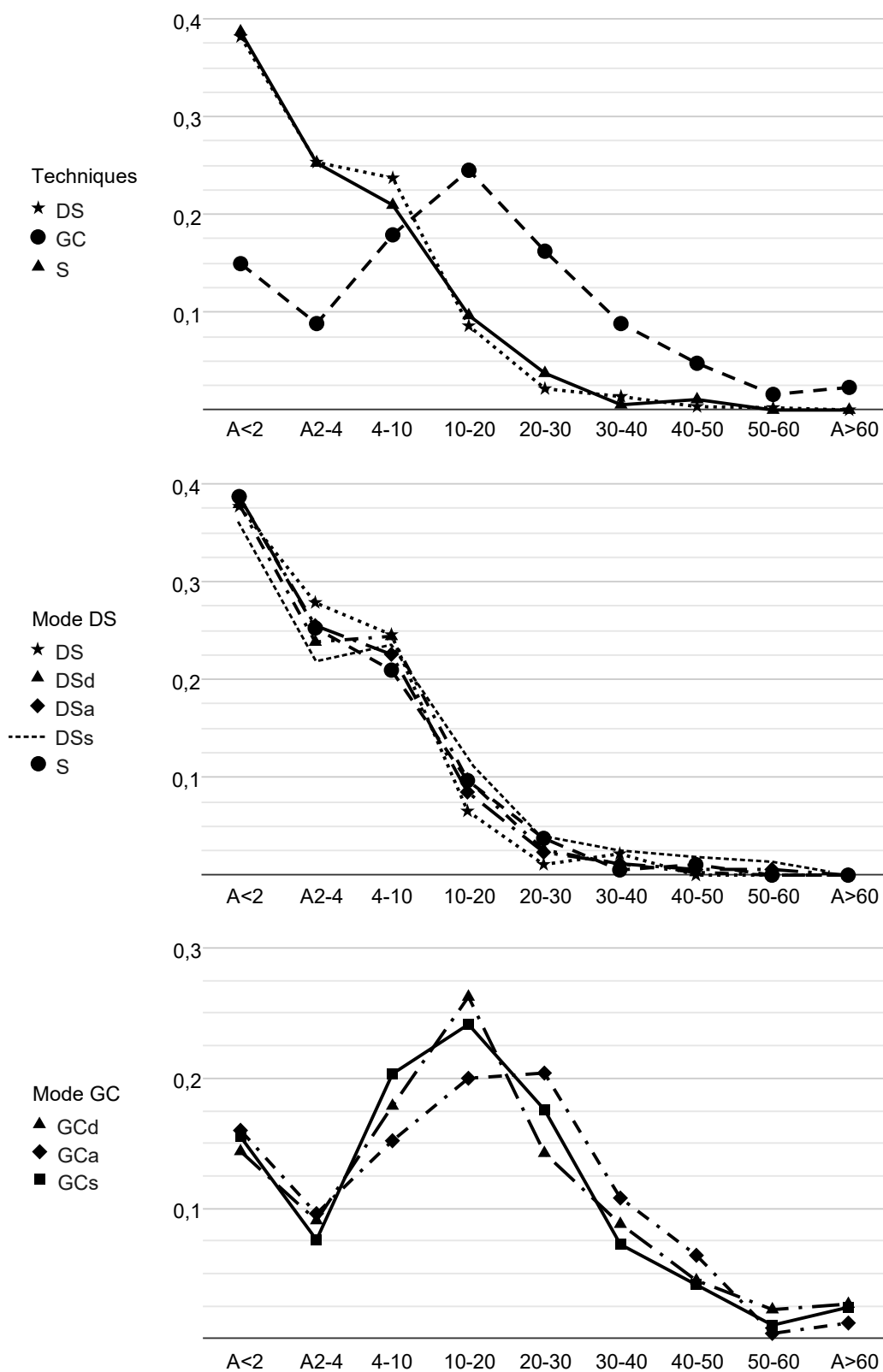


Figure 69 – Score A counts by stratum. Top: techniques. Center: DS pairs. Bottom: GC pairs.

## A.1.2 Score B

### A.1.2.1 Distributions and outliers

Sample size is 1047 choices for DS (30.4 per participant) and 1232 for GC (38.5 per participant), with a total of 35 outliers found (1.5%, see table 25). Concerning modes, sample sizes are similar per technique (349 per mode on DS and 410 on GC), differing about 15%, which is not a concern for statistical analysis.

Data distributions clearly deviate from normality and appear to follow a Beta distribution (Figure 70, top). For the sake of simplicity, as GLMMs for such distributions are not well documented, the distribution was inverted to a Weibull distribution, and this score was analyzed adopting the Weibull distribution family.

Table 25 – Sb: group size and outliers on score B

Technique	N	N (per mode)	Outliers
DS	1047	349	11
GC	1232	410	24

### A.1.2.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results were significant for techniques ( $p < 0.01$ ) and pairs ( $p < 0.01$ ). Indeed, boxplots (Figure 70, middle) show noticeable differences between Techniques (Table 26) and between GC and DS Pairs. Inspecting pairwise comparisons (Table 28) DS and S are almost equally dissimilar than all GC pairs (about 17-22 % higher), and no significant differences were found among GC pairs. Differences in Color modes was significant only on GC ( $p < 0.01$ ) where Mono had a 4.7 % higher score than Color (Table 27).

Therefore we can conclude that DS and S have a slightly higher score B than GC, and color played a minor role in GC.

Table 26 – Sb: significant differences (%) per technique on score B

Technique	Mean	s.d.	Difference
DS	9,51	0,5	20,07
GC	7,92	2,01	-16,71

Table 27 – Sb: significant differences (%) per Color mode on score B

Color mode	Mean	s.d.	Difference
DS Color	9,49	0,51	-
DS Mono	9,53	0,49	-
GC Color	7,66	2,02	-4,48
GC Mono	8,02	1,99	4,7
Color	8,76	1,61	-
Mono	8,51	1,81	-



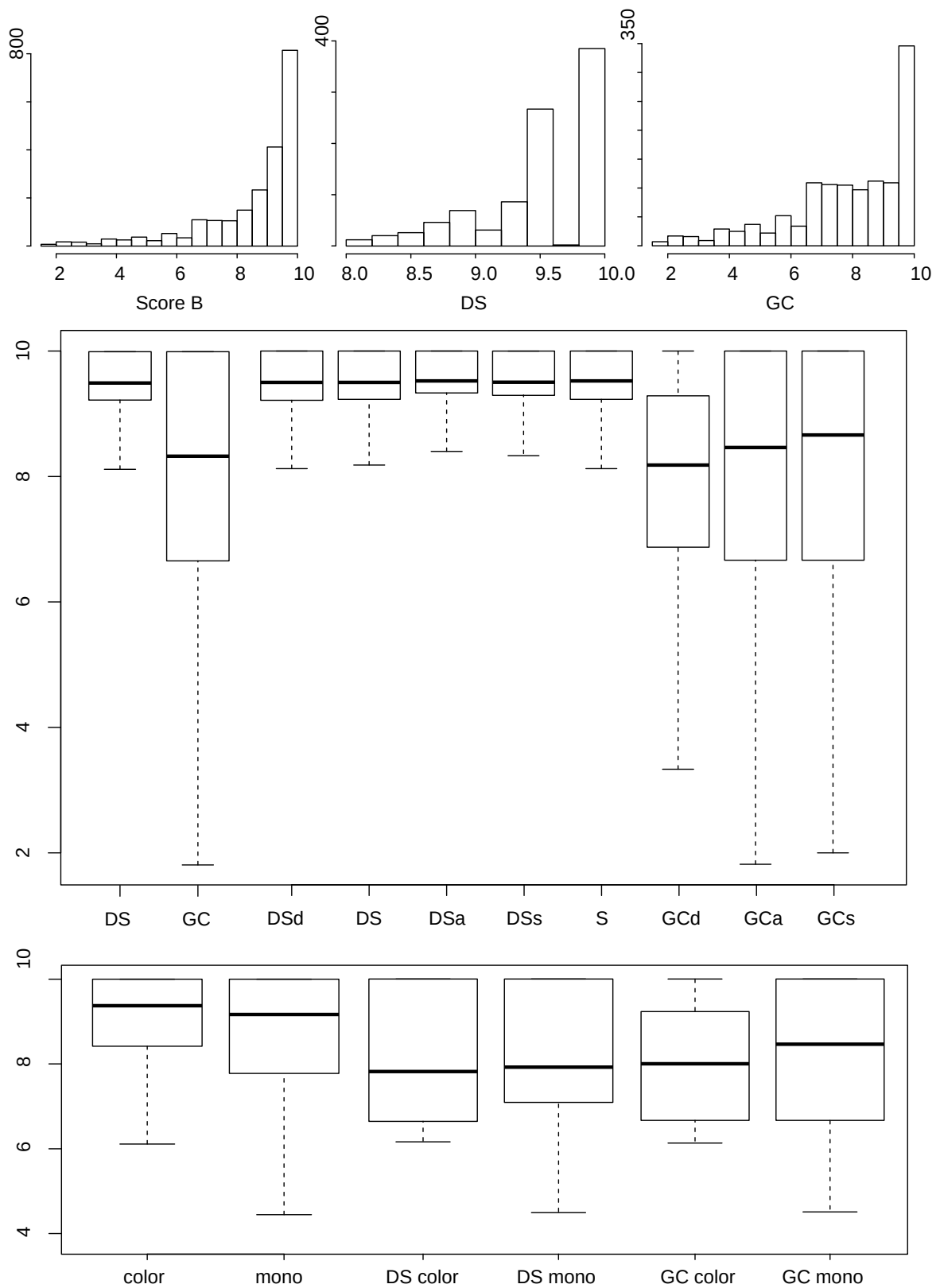


Figure 70 – Score B. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

Table 28 – Sb: significant differences (%) between pairs on score B

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	9,53	0,49	0	-	-	-	-	-18,2	-15,0	-15,0
DS	9,5	0,5	-	0	-	-	-	-18,0	-14,7	-14,8
DSd	9,47	0,5	-	-	0	-	-	-17,8	-14,5	-14,5
DSa	9,54	0,49	-	-	-	0	-	-18,3	-15,1	-15,1
DSs	9,62	0,49	-	-	-	-	0	-	-	-
GCd	7,78	1,99	22,38	22,06	21,71	22,51	-	0	-	-
GCa	8,1	1,94	17,66	17,35	17,02	17,79	-	-	0	-
GCs	8,09	2,09	17,73	17,42	17,08	17,86	-	-	-	0

### A.1.2.3 Analysis of score B = 10

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing the dynamics of choices with maximum score B, as in this special case, B = 10 means that the best visible grain in the moment was chosen. This was done by counting the proportion of B = 10 per technique (Figure 71, top), pairs (Figure 71, middle) and color modes (Figure 71, bottom).

By examining techniques, S and DS remain slightly better than GC in this score. Concerning visualization modes, one can see that 3D on GC technique has a significantly higher proportion (about 50 %) of max scores than 2.5D and that 3D modes are slightly favored by DS, considering this score. Finally, concerning color mode, monochrome GC was 41 % better than GC Color, reinforcing the small distinction found by the ANOVA analysis.

### A.1.2.4 Correlations

Table 29 shows that only a few weak correlations ( $r < 0.2$ ) were found between score B and selected variables (Time, Movement, Oscillations) or pre/post tests scores (MRT, SBSOD, and TLX).

Table 29 – Sb: correlations of other variables with score B

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	-	-0.09	-0.08	-0.05	-	-0.06
DS	-	-	-	-	-	-
GC	-	-	-	-	-	-
S	-	-	-	-	-	-

#### A.1.2.4.1 Score B summary

Considering all the evidence gathered about score B (ANOVA, counts of B = 10, and correlations), the following order (descending performance) is suggested: DS and S > GCs and GCa > GCd.

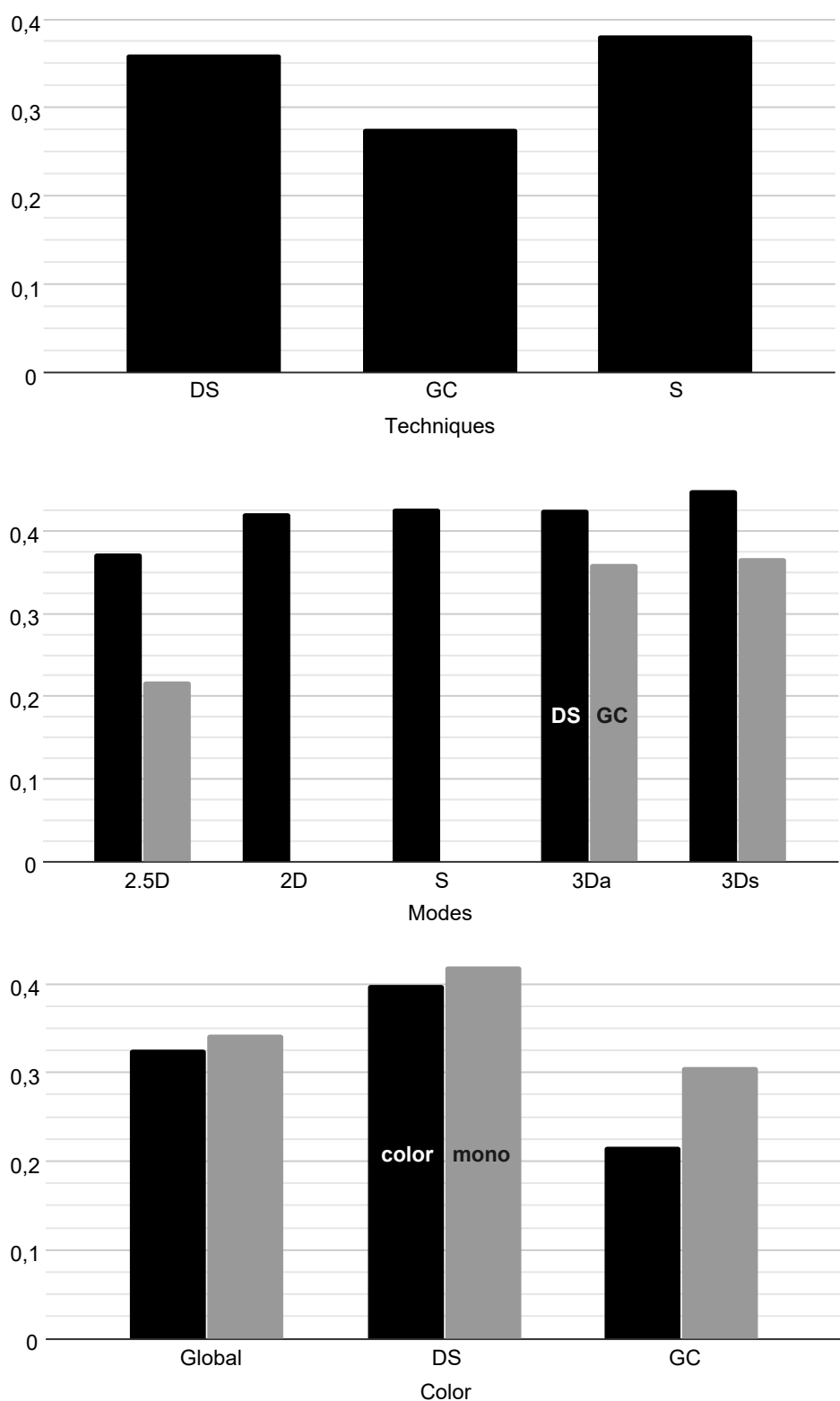


Figure 71 – Score B = 10 counts (%). Top: techniques. Center: DS and GC. Bottom: color modes.

### A.1.3 Score C

#### A.1.3.1 Distributions and outliers

Being a secondary score, this analysis was only performed for good choices or the top 10 choices (Score A  $\leq 10$ ). The sample size is 862 choices for DS (28 per participant) and 919 for GC (29 per participant), with a total of 93 outliers found (5.2%, see table 30). Concerning modes, sample sizes are similar per technique (215 per mode on DS and 229 on GC), differing about 6%, which is not a concern for statistical analysis.

Data distributions clearly deviate from normality and appear to follow a Gamma distribution (Figure 72, top). Therefore this score was analyzed using a GLMM, adopting the Gamma distribution family.

Table 30 – Sb: group size and outliers on score C

Technique	N	N (per mode)	Outliers
DS	862	215	41
GC	919	229	52

#### A.1.3.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results were significant for techniques ( $p < 0.01$ ) and pairs ( $p < 0.01$ ). Indeed, boxplots (Figure 72, middle) show noticeable differences between Techniques (Table 31) however, it seems to be due to the GCd mode. Inspecting the numerous pairwise comparisons (Table 33) DS and S are almost equally dissimilar than GCd (about 57-70 % lower). Within DS, DSd, DSa, and S form a group, DSs and DS form another, differing between 20 and 40 %. The difference is quite notorious within GC, where GCs, GCa, and GCd stand in an almost linear crescent, with radical differences among them, higher than 90%. Differences in Color modes was significant on GC ( $p < 0.01$ ) where Mono had a 44 % lower score than Color (Table 32). Overall color mode differences, despite significant, was only due to the GC technique.

Therefore we can conclude that GCd ends up unbalancing the technique analysis, and the numerous differences between pairs can be organized in three groups: GCs, S, DSa, DSd; DS, DSs, GCa; GCd. Color played a substantial role in GC.

Table 31 – Sb: significant differences (%) per technique on score C

Technique	Mean	s.d.	Difference
DS	156,34	89	-52,94
GC	332,26	225,6	112,52

#### A.1.3.3 Strata analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing the dynamics of score C along the 9 strata of score A, which may show if any

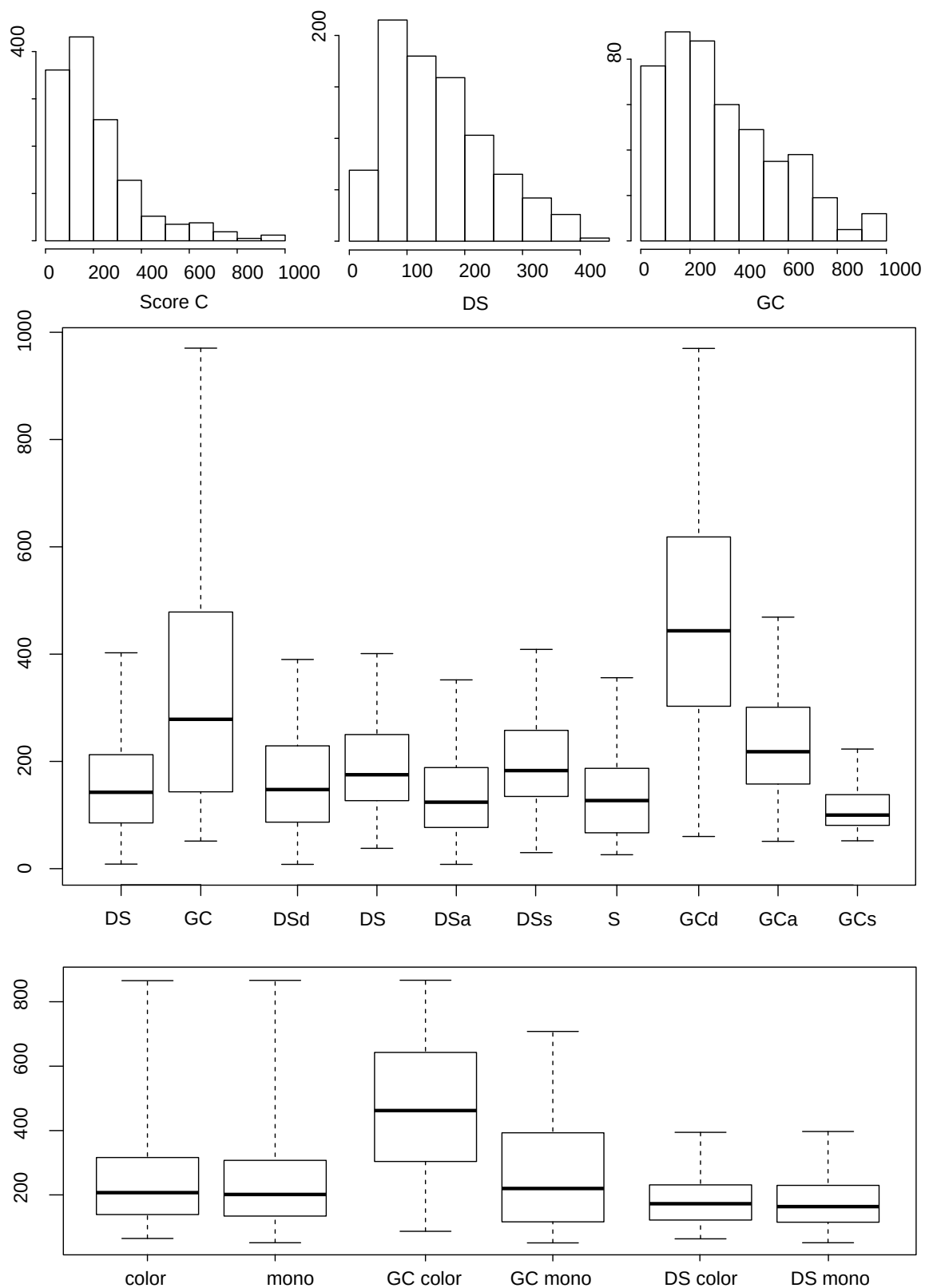


Figure 72 – Score C. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

Table 32 – Sb: significant differences (%) per Color mode on score C

Color mode	Mean	s.d.	Difference
DS Color	158	88	-
DS Mono	154	90	-
GC Color	495	226	78,7
GC Mono	277	197	-44,04
Color	226	186	6,1
Mono	213	163	-5,75

Table 33 – Sb: significant differences (%) between pairs on score C

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	141	86	0	34	-	-	42	234	72	-
DS	190	87	-25	0	-16	-24	6	148	-	-35
DSd	158	90	-	20	0	-	27	198	53	-22
DSa	144	85	-	32	-	0	40	228	69	-
DSs	202	88	-29	-5,74	-21	-28	0	134	-	-39
GCd	474	205	-70	-59	-66	-69	-57	0	-48	-73
GCa	244	129	-41	-	-34	-40	-	94	0	-49
GCs	123	77	-	54	28	-	64	284	97	0

condition provides faster feature detection on the first three strata, i.e., of top 10 bigger grains. Overall, each condition was deemed very consistent across those strata. Analysis of techniques (Figure 73, top) reinforces the differences found between GC and the others, and how S and DS performance are pretty close. Within DS (Figure 73, middle), 2D is slightly worse than the others, while within GC (Figure 73, bottom), the analysis only reinforces the difference between the three modes.

#### A.1.3.4 Correlations

Table 34 shows mostly weak correlations ( $r < 0.2$ ) between score C and selected variables (Time, Movement, Oscillations) or pre/post tests scores (MRT, SBSOD and TLX). Score C's Relationships with Time, Movement, and Oscillations are expected as all variables are related to time spent, and score C is essentially a time-related variable. The weak negative relationship with Movement on S makes sense, as with the larger jumps, a participant may reduce the choice time. Curiously, a weak to medium relationship was found with spatial ability on S technique, i.e., more spatial ability would be associated with longer reaction/choice time, which is counter-intuitive and will be further discussed.

Table 34 – Sb: correlations of other variables with score C

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	0,13	0,22	0,17	-	0,05	-
DS	-	-	0,08	0,1	0,16	-
GC	0,11	0,13	0,06	0,07	-	-0,14
S	-	-0,17	-	0,19	0,38	-

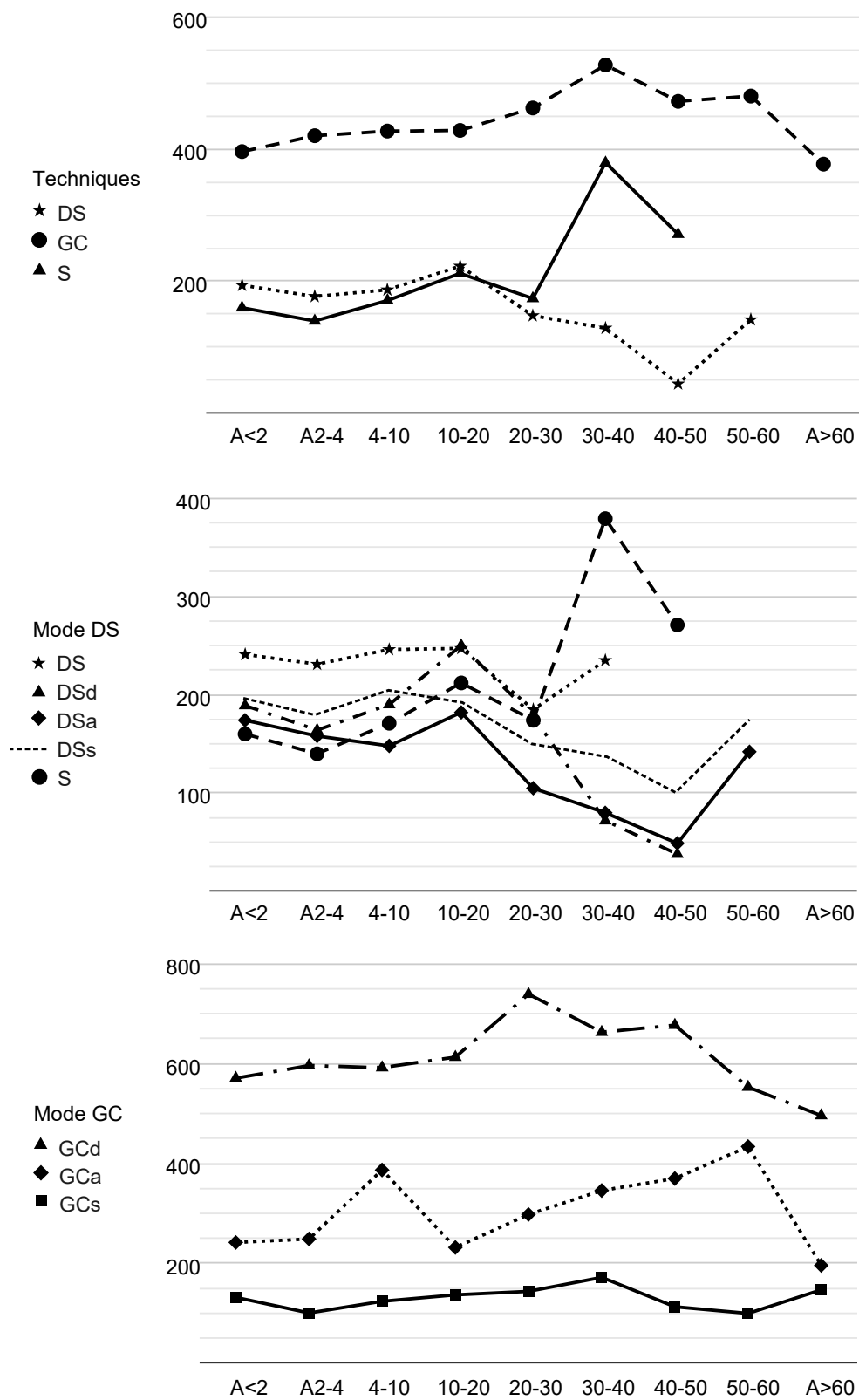


Figure 73 – Score C means by score A stratum. Top: techniques. Center: DS. Bottom: GC.

#### A.1.3.4.1 Score C summary

Considering all the evidence gathered about score C, the following order (descending performance) is suggested: GCs and S and DSa and DSd > DS >= DSs and GCa > GCd. At first sight, this order does not show a clear pattern. However, score C was measured in logic tics, not in seconds. If we exclude GCd, the maximum difference is about two seconds (GCa x GCs), which one can consider negligible. When GCd is included, this difference rises to six seconds. Therefore it seems reasonable to find why GCd differed so much from the others.

### A.1.4 Score D

#### A.1.4.1 Distributions and outliers

Being a secondary score, this analysis was only performed for good choices or top 10 choices (Score A <= 10). As D = 1 has a distinct meaning, those scores were also removed from the data sample. The remaining sample size has 529 choices for DS (17 per participant) and 625 for GC (19 per participant), with no outliers found (see table 35). Concerning modes, sample sizes are comparable per technique (132 per mode on DS and 157 on GC), differing about 16%, which is not too large for statistical analysis.

Data distributions clearly deviate from normality and appear to follow a Gamma distribution (Figure 74, top). Therefore this score was analyzed using a GLMM, adopting the Gamma distribution family.

Table 35 – Sb: group size and outliers on score D

Technique	N	N (per mode)	Outliers
DS	529	132	0
GC	625	157	0

#### A.1.4.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results was significant for techniques ( $p < 0.01$ ) and pairs ( $p < 0.01$ ). Boxplots (Figure 74, middle) shows clear differences between Techniques (Table 36) and between GC and DS/S pairs. Indeed, pairwise comparisons (Table 38) are only allowed among DS / S and GC pairs, where the former group have lower scores (21-37 %). Differences in Color modes was significant only on the omnibus test ( $p < 0.05$ ) where Color had a 14 % lower score than Mono (Table 37).

Table 36 – Sb: significant differences (%) per technique on score D

Technique	Mean	s.d.	Difference
DS	0,41	0,22	-22,64
GC	0,53	0,24	29,26



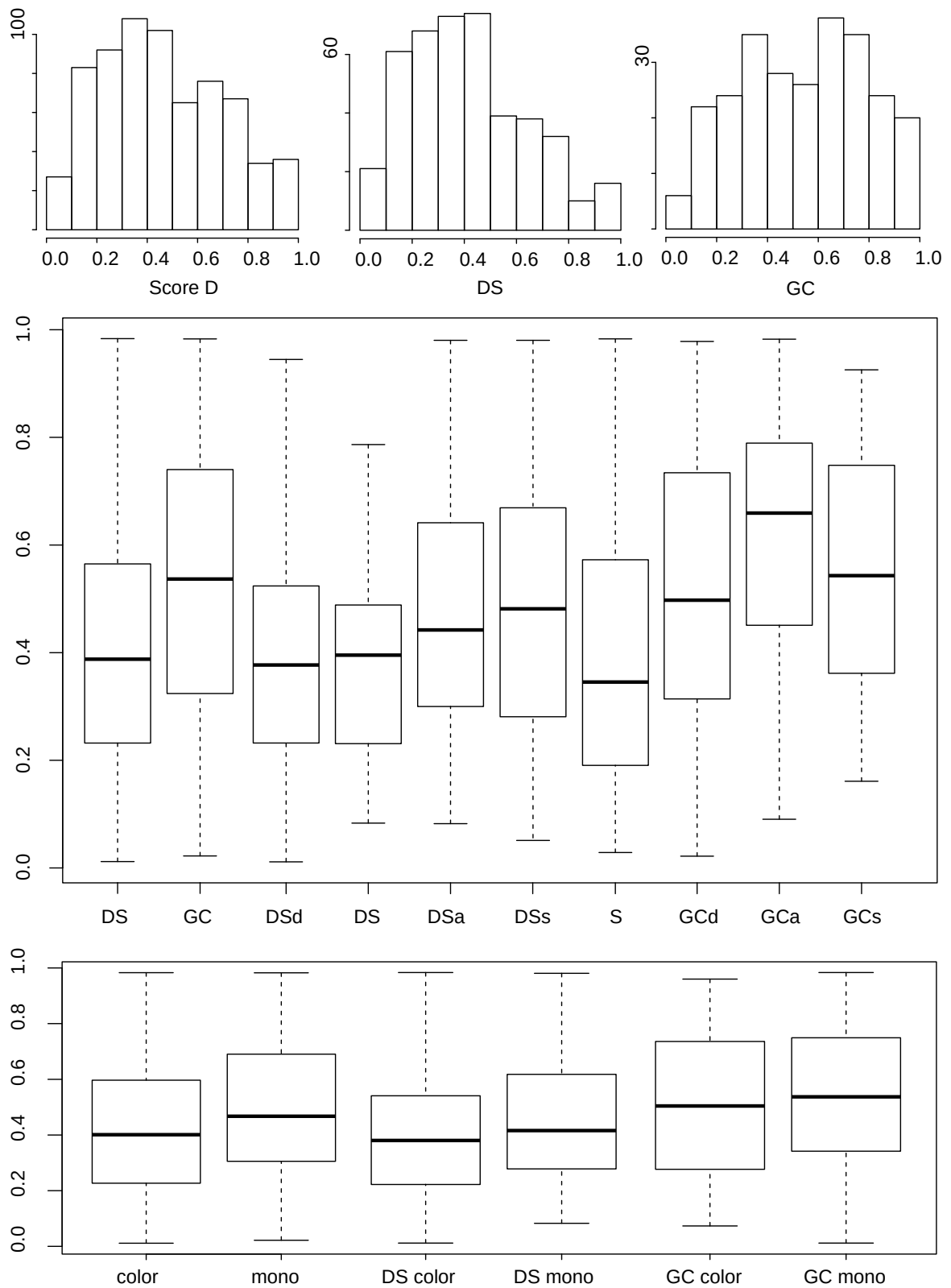


Figure 74 – Score D. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

Table 37 – Sb: significant differences (%) per Color mode on score D

Color mode	Mean	s.d.	Difference
DS Color	0,39	0,22	-
DS Mono	0,44	0,22	-
GC Color	0,51	0,25	-
GC Mono	0,54	0,24	-
Color	0,42	0,23	-14,28
Mono	0,49	0,23	16,66

Table 38 – Sb: significant differences (%) between pairs on score D

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	0,39	0,25	0	-	-	-	-	27,6	55,4	37,1
DS	0,39	0,2	-	0	-	-	-	-	56,5	38,1
DSd	0,38	0,2	-	-	0	-	-	31,6	60,2	41,4
DSa	0,46	0,22	-	-	-	0	-	-	33,5	-
DSs	0,48	0,23	-	-	-	-	0	-	29,4	-
GCd	0,51	0,25	-21,6	-	-24	-	-	0	-	-
GCa	0,62	0,24	-35,6	-36,1	-37,6	-25,1	-22,7	-	0	-
GCs	0,54	0,22	-27	-27,5	-29,2	-	-	-	-	0

#### A.1.4.3 Strata analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing the dynamics of choices with maximum score D, as in this special case,  $D = 1$  means that the selected grain was chosen in its first appearance. This was done by counting the proportion of  $D = 1$  per technique (Figure 75, top), pairs (Figure 75, middle) and per score A strata (Figure 75, bottom).

The distinction between techniques are clear ( $S 53\% < GC 25\% < DS$ ). Concerning visualization modes, all DS modes are balanced, however, in GC, the order  $GCs > GCa > GCd$  is suggested. Strata analysis shows a relative consistency in the first three strata ( $A < 10$ ), reinforcing previous findings.

#### A.1.4.4 Correlations

Table 39 shows weak (around 0.2) and moderate correlations (around 0.5) between score D and selected variables (Time, Movement, Oscillations) or pre/post tests scores (MRT, SBSOD and TLX). It is worth noting that the stronger correlations are all negative and related to the first three variables. This was expected as score D is time-sensitive, thus, as the participant moves or oscillates more, the score gets smaller. The table shows that DS is more sensitive to this phenomenon.

##### A.1.4.4.1 Score D summary

Considering all the evidence gathered about score D, the following order (descending performance) is suggested:  $DS > GCs \geq GCa \geq GCd > S$ .

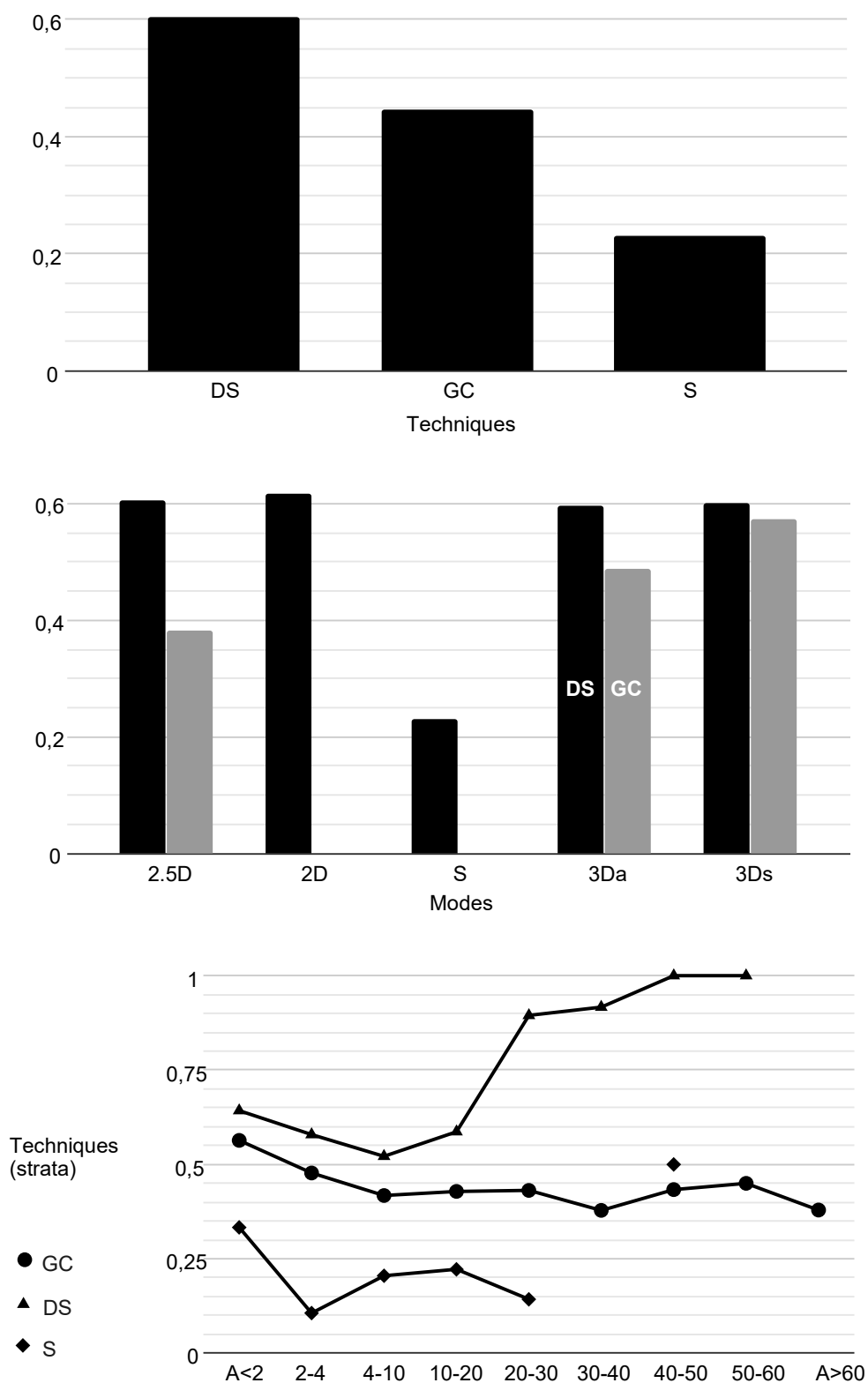


Figure 75 – Score D. Top: counts D = 1 per technique (%). Center: per DS and GC pairs. Bottom: per score A stratum.

Table 39 – Sb: correlations of other variables with score D

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	-0,22	-0,25	-0,2	-	-	0,1
DS	-0,31	-0,53	-0,46	0,12	0,23	-
GC	-0,22	-0,34	-0,33	-0,1	-0,1	0,19
S	-0,34	-0,34	-0,29	0,17	0,17	-

## A.2 Size task analysis: smaller grains (Ss)

### A.2.1 Score A

#### A.2.1.1 Distributions and outliers

The sample size is 1168 choices for DS (38 per participant) and 1121 for GC (35 per participant), without outliers. Concerning modes, sample sizes are almost equal per technique (see table 40). Data distributions clearly deviate from normality and appear to conform to a Gamma distribution (Figure 76, top). Therefore this score was analyzed using a GLMM, adopting the Gamma distribution family.

Table 40 – Ss: group size and outliers on score A

Technique	N	N (per mode)	Outliers
DS	1168	233	0
GC	1121	224	0

#### A.2.1.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results were significant for techniques ( $p < 0.01$ ) and pairs ( $p < 0.01$ ). Boxplots (Figure 76, middle) show minor differences between Techniques (Table 41), no apparent differences between GC modes and more distinct differences among DS levels. Technique GC presented a score about 11 % smaller (Table 41). Inspecting pairwise comparisons (Table 43), DSs and DSa are the more distinct among DS and S techniques, with the lowest and highest scores, respectively. Considering GC, only GCa and GCs are comparable, and the latter presented a 13 % smaller score. Color mode was significant only on DS ( $p < 0.01$ ), with only a 8 % difference in favor of Color (Table 42).

Therefore we can conclude that split-screen 3D (GCs and DSs) had the lowest scores, DSa and DSd the highest, and DS, S, GCd and GCa are in between. Color had a minor impact.

Table 41 – Ss: significant differences (%) per technique on score A

Technique	Mean	s.d.	Difference
DS	34,94	19,86	12,43
GC	31,07	20,75	-11,06

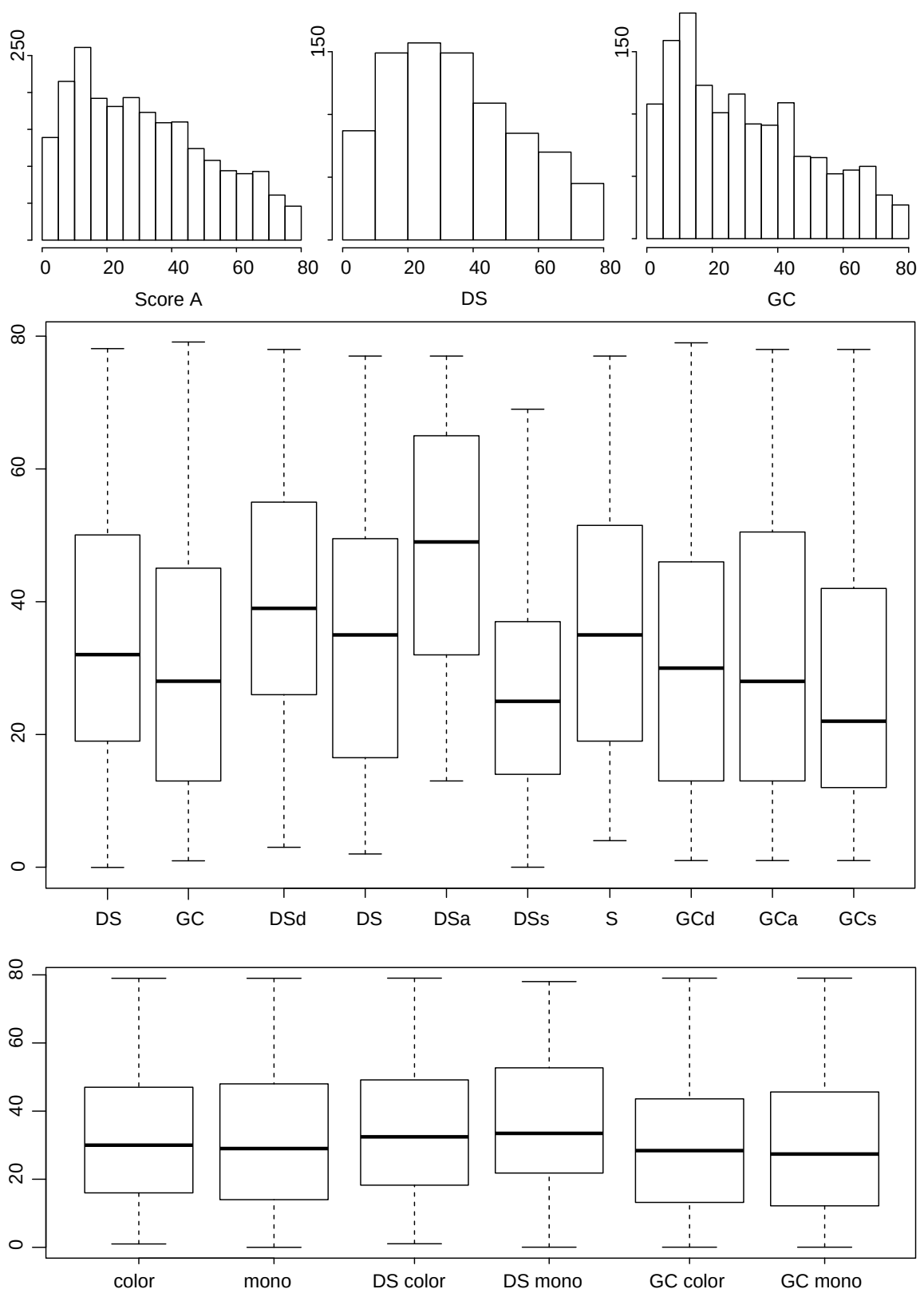


Figure 76 – Score A. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

Table 42 – Ss: significant differences (%) per Color mode on score A

Color mode	Mean	s.d.	Difference
DS Color	33,77	19,67	-7,99
DS Mono	36,7	20,03	8,68
GC Color	31,2	20,8	-
GC Mono	31	20,7	-
Color	32,7	20,2	-
Mono	32,4	20,7	-

Table 43 – Ss: significant differences (%) between pairs on score A

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	35,87	19,65	0	-	-	34,4	-23,8	-	-	-20,3
DS	34,99	20,29	-	0	-	37,8	-21,9	-	-	-18,3
DSd	40,87	19,34	-	-	0	-	-33,1	-23,2	-19,6	-30
DSa	48,22	18,28	-25,6	-27,4	-	0	-43,3	-34,9	-31,8	-40,7
DSs	27,3	17,07	31,3	28,1	49,6	76,6	0	14,8	20,3	-
GCd	31,35	20,4	-	-	30,3	53,8	-12,9	0	-	-
GCa	32,85	21,62	-	-	24,3	46,7	-16,8	-	0	-12,9
GCs	28,58	20,59	25,4	22,3	42,9	68,6	-	-	14,9	0

### A.2.1.3 Strata Analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing how choices are distributed between nine strata (Score A < 2, between 2 and 4, 4-10, 10-20, 20-30, 30-40, 40-50, 50-60, higher than 60). In this case, most choices should be within the first strata (A<10). All detected differences were small and can be considered negligible. By examining techniques (Figure 77, top), S and DS remain quite similar, while GC is more efficient precisely at the 4-10 strata. Visualization modes on the DS technique (Figure 77, middle) and GC confirmed the same performances found by ANOVA analysis. The fundamental difference between GC levels occurred between scores 4-20.

### A.2.1.4 Correlations

Table 44 shows that only a few weak correlations ( $r < 0.2$ ) were found between score A and pre/post tests scores (MRT, SBSOD, and TLX).

Table 44 – Ss: correlations of other variables with score A

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	-	-	-	-0,05	-	-
DS	-	-	-	-	-	-
GC	-	-	-	-0,06	-	0,06
S	-	-	-	-	-	-

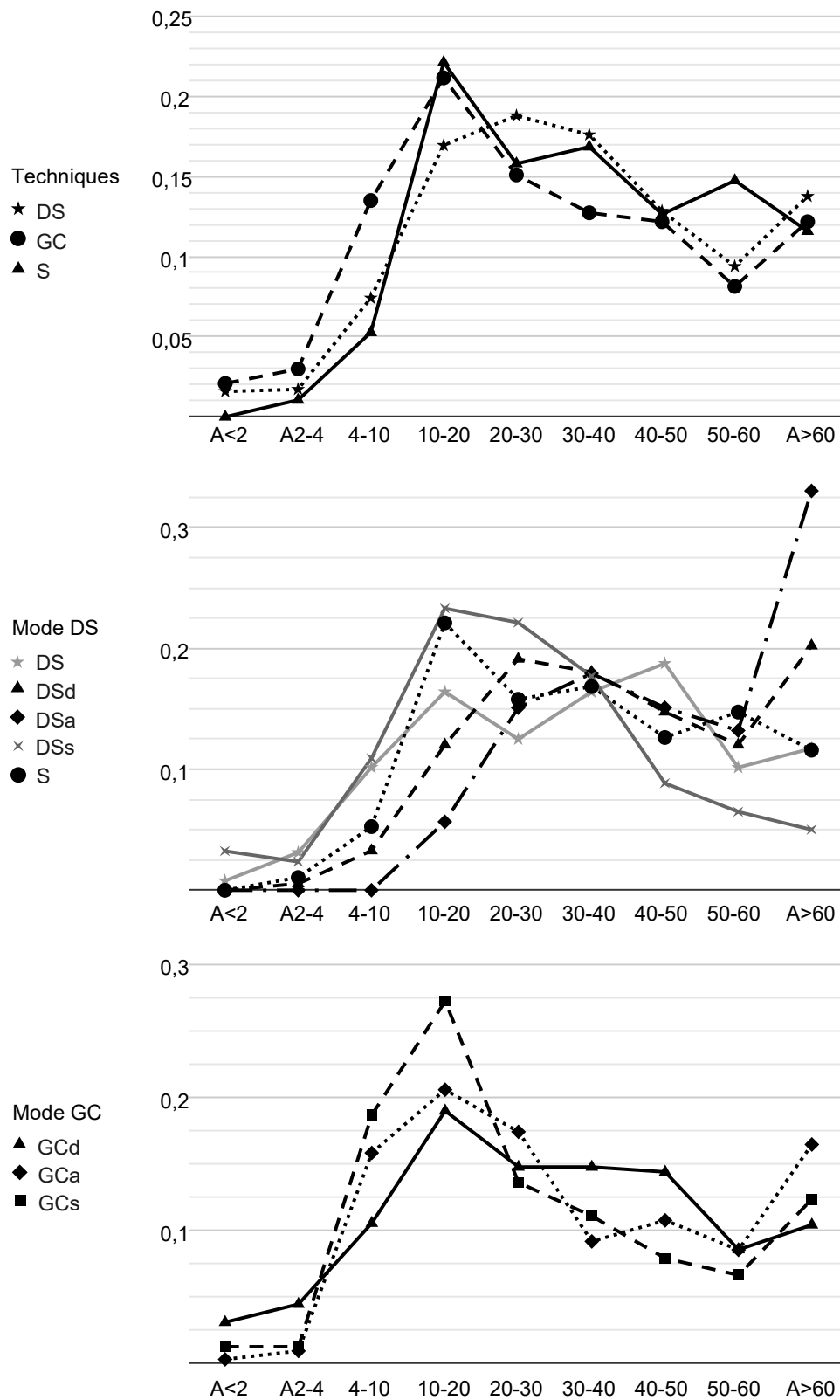


Figure 77 – Score A counts by stratum. Top: techniques. Center: DS pairs. Bottom: GC pairs.

#### A.2.1.4.1 Score A summary

Considering all the evidence about score A, the following order (descending performance) is suggested: GCs and DSs > DS, S, GCd, and GCa > DSd >= DSa.

### A.2.2 Score B

#### A.2.2.1 Distributions and outliers

Sample size is 1168 choices for DS (38 per participant) and 1121 for GC (35 per participant), with no outliers found. Concerning modes, sample sizes are almost equal per technique (Table 45).

Data distributions showed a distinct peak on  $B = 10$ , which is already known to have a different meaning. Therefore, this stratum was removed, and the rest clearly deviate from normality, not following any specific distribution (Figure 78, top). In this case, the Aligned Rank Transform is more suited.

Table 45 – Ss: group size and outliers on score B

Technique	N	N (per mode)	Outliers
DS	1168	233	0
GC	1121	224	0

#### A.2.2.2 Boxplots, ANOVAs and comparisons

GLMM ANOVA results were significant for techniques ( $p < 0.01$ ) and pairs ( $p < 0.01$ ). Boxplots (Figure 78, middle) shows minor differences between Techniques and between GC Pairs, while in DS they seem more pronounced, following a similar pattern as found in score A boxplots. However, pairwise comparisons (Table 48) show fewer differences, mostly related to DSa (all others 21-39 % higher) and GCs (all others 9-20 % lower), while the two presented the larger difference (51 % in favor of the GCs). Techniques differed only by 8 % in favor of GC (Table 46) and once more only DS shown significant differences in Color mode ( $p < 0.01$ ), with a mild 6 % difference in favor of Color (Table 47).

Therefore we can conclude that scores are very close amongst most conditions and GCs has the highest score B while DSa has the lowest. Color played a minor role in DS. This is in agreement with score A results.

Table 46 – Ss: significant differences (%) per technique on score B

Technique	Mean	s.d.	Difference
DS	6,93	2,59	-7,8
GC	7,51	2,55	8,47



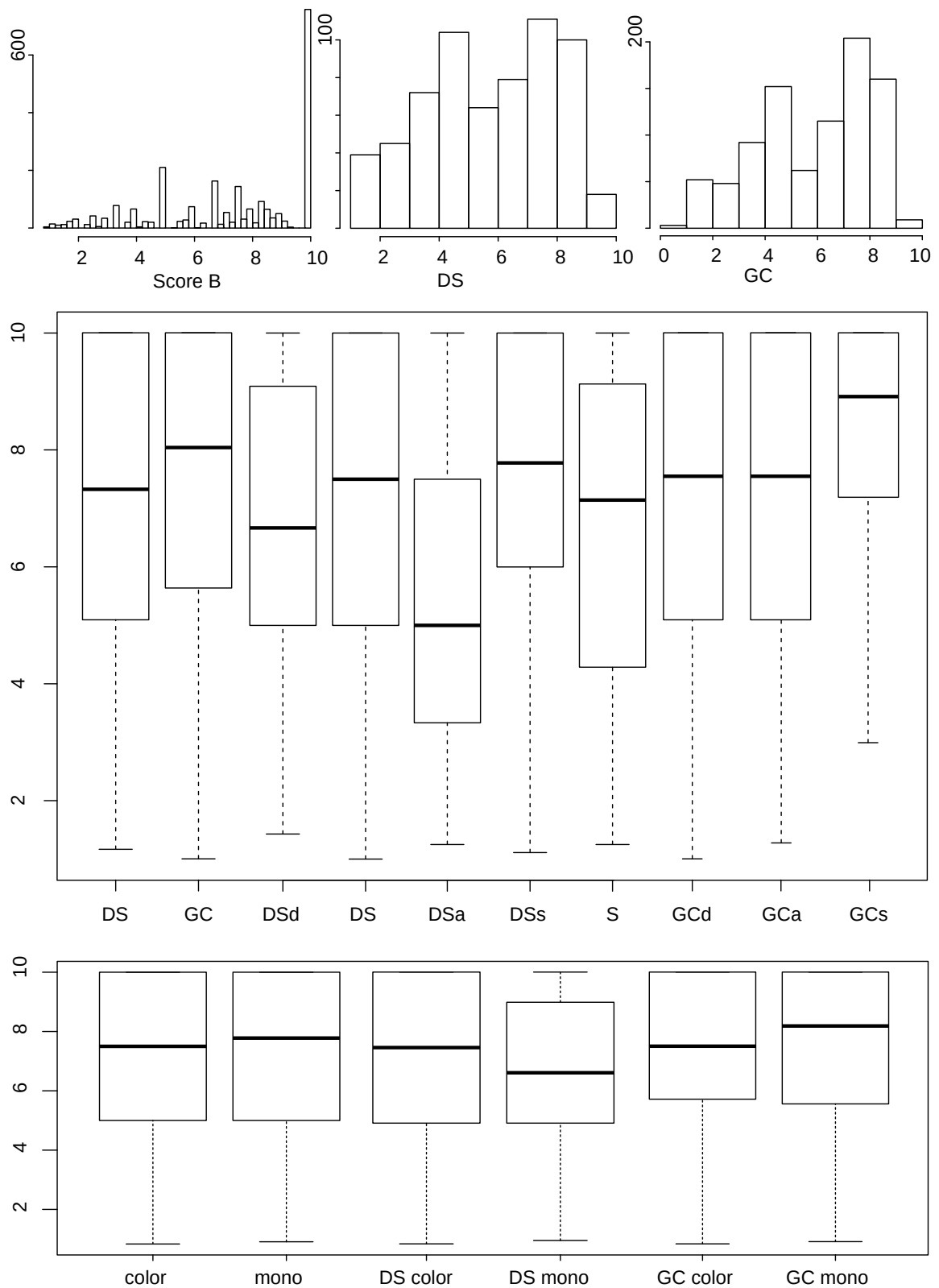


Figure 78 – Score B. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

Table 47 – Ss: significant differences (%) per Color mode on score B

Color mode	Mean	s.d.	Difference
DS Color	7,09	2,57	6,3
DS Mono	6,67	2,6	-5,92
GC Color	7,37	2,64	-
GC Mono	7,57	2,52	-
Color	7,16	2,61	-
Mono	7,39	2,56	-

Table 48 – Ss: significant differences (%) between pairs on score B

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	6,58	2,71	0	-	-	-17,5	-	-	-	24,9
DS	7,09	2,63	-	0	-	-23,4	-	-	-	15,9
DSd	6,68	2,64	-	6	0	-18,8	13,2	-	-	22,8
DSa	5,42	2,57	21,2	30,6	23,2	0	39,5	33,7	37,8	51,4
DSs	7,57	2,29	-	-	-11,6	-28,3	0	-	-	8,5
GCd	7,26	2,7	-	-	-	-25,2	-	0	-	13,2
GCa	7,47	2,36	-	-	-	-27,4	-	-	0	9,8
GCs	8,21	2,2	-19,9	-13,7	-18,6	-33,9	-7,8	-11,6	-9	0

### A.2.2.3 Strata analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing the dynamics of choices with maximum score B, as in this special case, B = 10 means that the best visible grain at the moment was chosen. This was done by counting the proportion of B = 10 per technique (Figure 79, top), pairs (Figure 79, middle) and color modes (Figure 79, bottom).

By examining techniques, GC remains slightly better, and DS is above S. Concerning visualization modes, one can see that 3D GCs on GC technique have a significantly higher proportion (about 28 %) of max scores than 2.5D and 3Da. On the DS technique, DSs and DS are superior to DSd (25 %), which surpasses DSa by 61 %. 3D modes are slightly favored by DS, considering this score. Finally, concerning color mode, GC appears to show a larger difference in favor of color, not found by the ANOVA analysis, and contrarily here, DS mono is slightly better than DS color.

### A.2.2.4 Correlations

No correlations were found between score B and selected variables (Time, Movement, Oscillations) or pre/post-test scores (MRT, SBSOD, and TLX).

#### A.2.2.4.1 Score B summary

Considering all the evidence gathered about score B, the following order (descending performance) is suggested: GCs > DSs and GCa and DS and GCd > S and DSd > DSa.

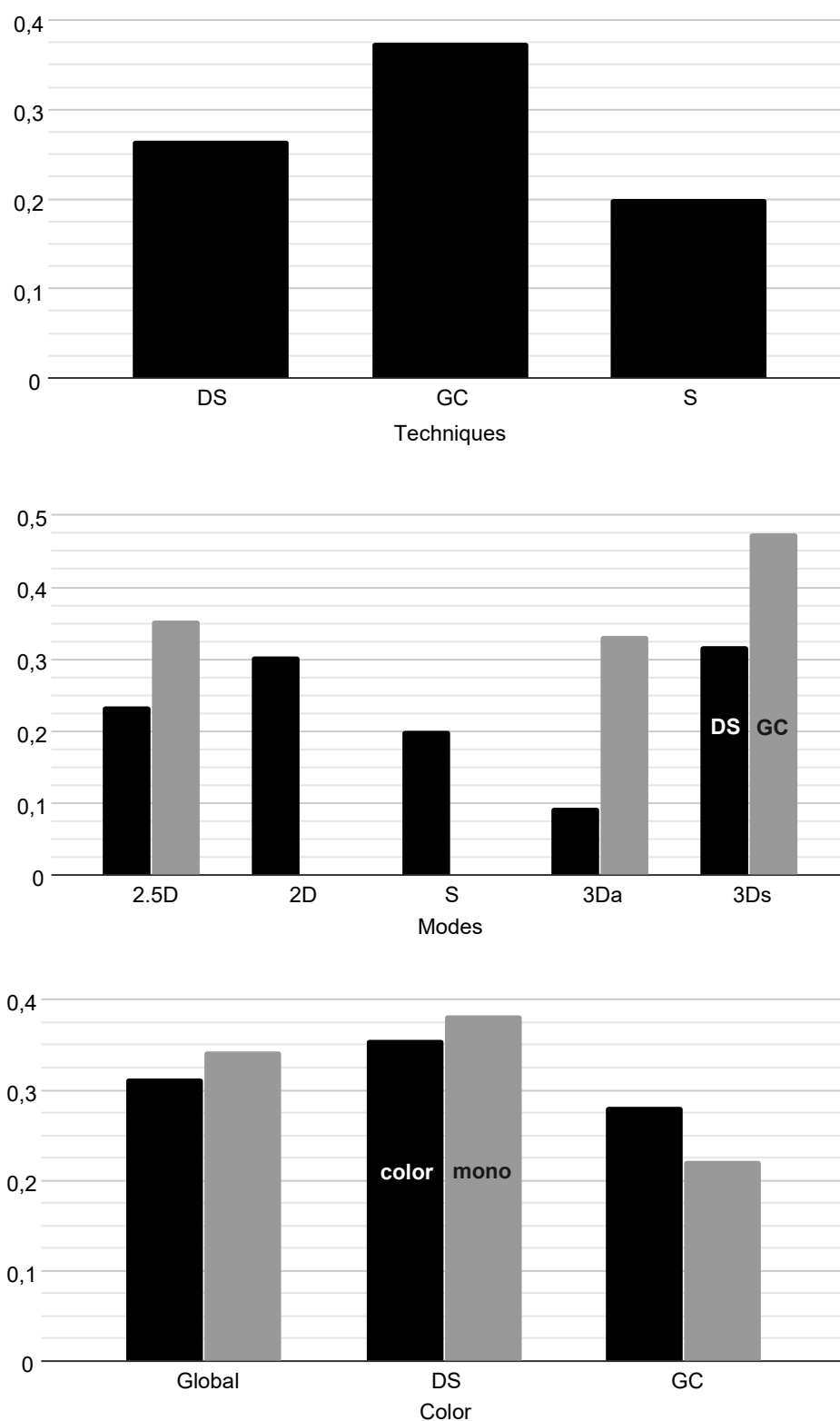


Figure 79 – Score B = 10 counts (%) by score A stratum. Top: techniques. Center: DS and GC pairs. Bottom: color modes.

## A.2.3 Score C

### A.2.3.1 Distributions and outliers

Being a secondary score, this analysis was only performed for good choices or top 10 choices (Score A  $\leq$  10). The sample size is 693 choices for DS (22 per participant) and 780 for GC (24 per participant), with a total of 95 outliers found (6.4%). Concerning modes, sample sizes are similar per technique (138 per mode on DS and 156 on GC), differing about 12%, which is not a concern for statistical analysis (Table 49).

Data distributions clearly deviate from normality and appear to conform to a Gamma distribution (Figure 80, top). Therefore this score was analyzed using a GLMM, adopting the Gamma distribution family.

Table 49 – Ss: group size and outliers on score C

Technique	N	N (per mode)	Outliers
DS	693	138	19
GC	780	156	76

### A.2.3.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results were significant for techniques ( $p < 0.01$ ) and pairs ( $p < 0.01$ ). Indeed, boxplots (Figure 80, middle) show noticeable differences between Techniques (Table 50), however, it seems to be due to the GCd mode. Inspecting the numerous pairwise comparisons (Table 52), DS and S are almost equally dissimilar than GCd (about 68-83 % lower), except for DSs. Within DS, DSd, DSa, and DS form a group, and DSs stand out with the highest score (between 60 and 104 % higher). The difference is quite notorious within GC, where GCs, GCa, and GCd stand in a crescent, with substantial differences among them, around 33%. Differences in Color modes (Figure 80, bottom) was significant on GC ( $p < 0.01$ ) and DS ( $p < 0.05$ ) and an interaction seems to take place, as Color had the lowest score on DS (23 % less) but it was Mono in GC (31 % less, see table 51).

Therefore we can conclude that GCd ends up unbalancing the technique analysis, and the numerous differences between pairs can be grouped this way: S (lower score); DS, DSd, and DSa; GCs and DSs; GCa; GCd (higher score). Color played a substantial role both in GC and DS, with opposite effects.

Table 50 – Ss: significant differences (%) per technique on score C

Technique	Mean	s.d.	Difference
DS	148,7	86,11	-51,06
GC	303,86	189,32	104,34

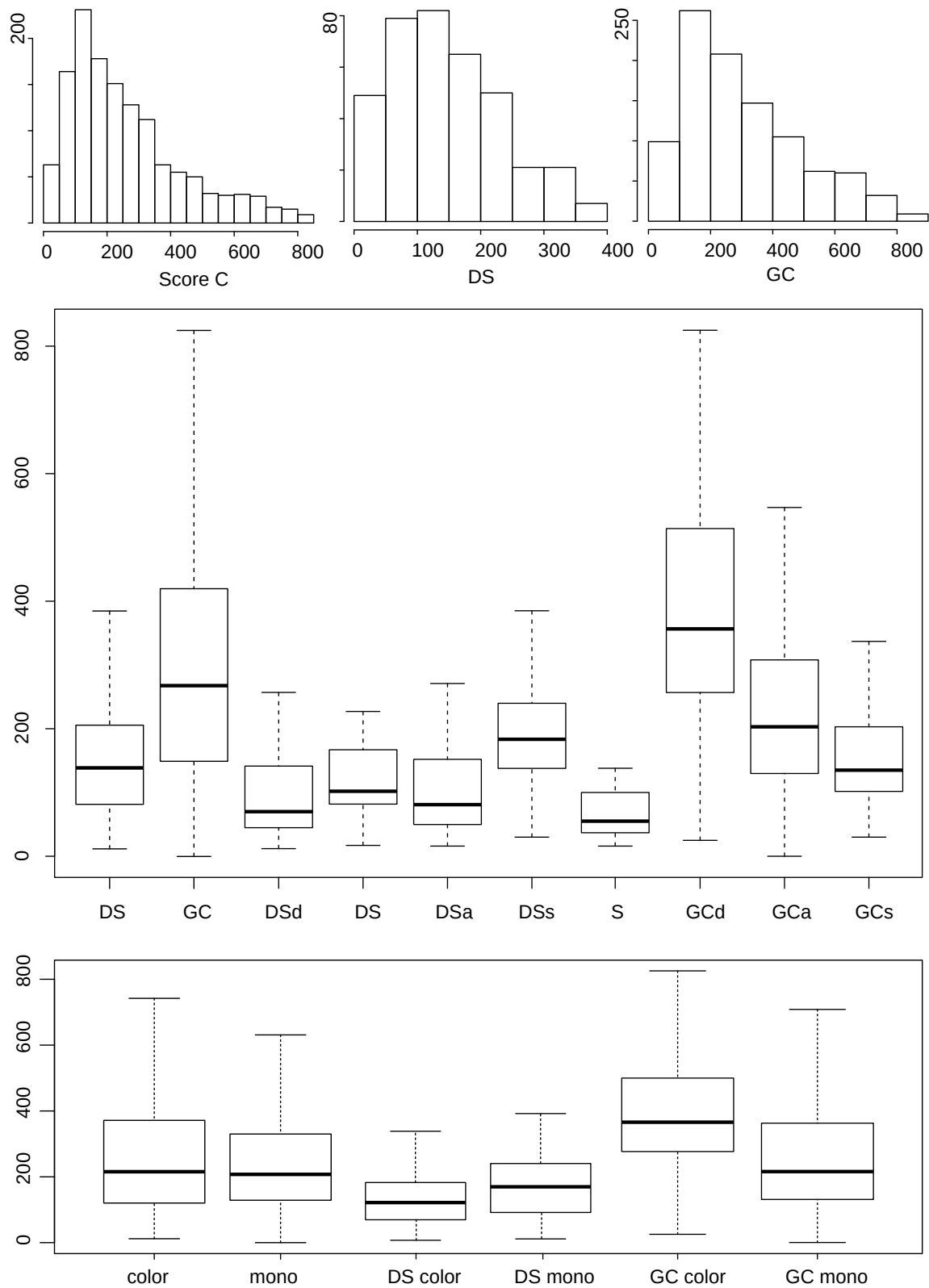


Figure 80 – Score C. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

Table 51 – Ss: significant differences (%) per Color mode on score C

Color mode	Mean	s.d.	Difference
DS Color	133,86	78,35	-23,1
DS Mono	174,08	92,91	30,04
GC Color	397,35	177,17	45,17
GC Mono	273,71	183,28	-31,11
Color	266,7	190,3	3,33
Mono	258,1	176	-3,22

Table 52 – Ss: significant differences (%) between pairs on score C

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	66,42	35,18	0	83,8	44,5	54,5	195,2	489,4	271,1	160,9
DS	122,13	61,64	-45,6	0	-	-	60,5	220,6	101,8	41,9
DSd	96,03	68,3	-30,8	-	0	-	104,2	307,7	156,7	80,5
DSa	102,68	75,38	-35,3	-	-	0	90,9	281,3	140,1	68,8
DSs	196,1	76,53	-66,1	-37,7	-51	-47,6	0	99,6	25,7	-
GCd	391,59	180,1	-83	-68,8	-75,4	-73,7	-49,9	0	-37	-55,7
GCa	246,54	164,02	-73	-50,4	-61	-58,3	-20,4	58,8	0	-29,6
GCs	173,36	123,07	-61,6	-29,5	-44,6	-40,7	-	125,8	42,2	0

### A.2.3.3 Strata analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing the dynamics of score C along the nine strata of score A, which may show if any condition provides faster feature detection on the first three strata, i.e., of top 10 bigger grains. Overall, conditions vary widely across those strata. S and DS techniques (Figure 81, top) seem consistent across strata, and S has the lowest score, although it has no choice in the first stratum ( $A < 2$ ). GC presents a rising score from 4-10 stratum to  $A < 2$ , although this seems to be due to GCd. Within DS (Figure 81, middle), 2D seems slightly better than the others, presenting choices in the first strata, followed by DSd and DSs. The average score of DSa is lower than DSs, but it has no choices in the first three strata. Within GC modes (Figure 81, bottom), the analysis only reinforces the difference between the three modes, where GCd remains clearly distinct, and GCs seems slightly better than GCa, although mostly due to lesser relevant strata.

### A.2.3.4 Correlations

Table 53 shows mostly weak correlations ( $r < 0.2$ ) between score C and selected variables (Time, Movement, Oscillations) or pre/post tests scores (MRT, SBSOD and TLX).

Score C's Relationships with Time, Movement and Oscillations are expected as all variables are related to time spent and score C is essentially a time-related variable. The weak negative relationship with Movement on DS may indicate that movement may reduce this technique's choice time. The weak to moderate correlation with time on GC may be related to the seemingly anomalous measurements on GCd - for some reason, users took longer to make choices in this condition than in 3D versions of GC. Finally, a weak

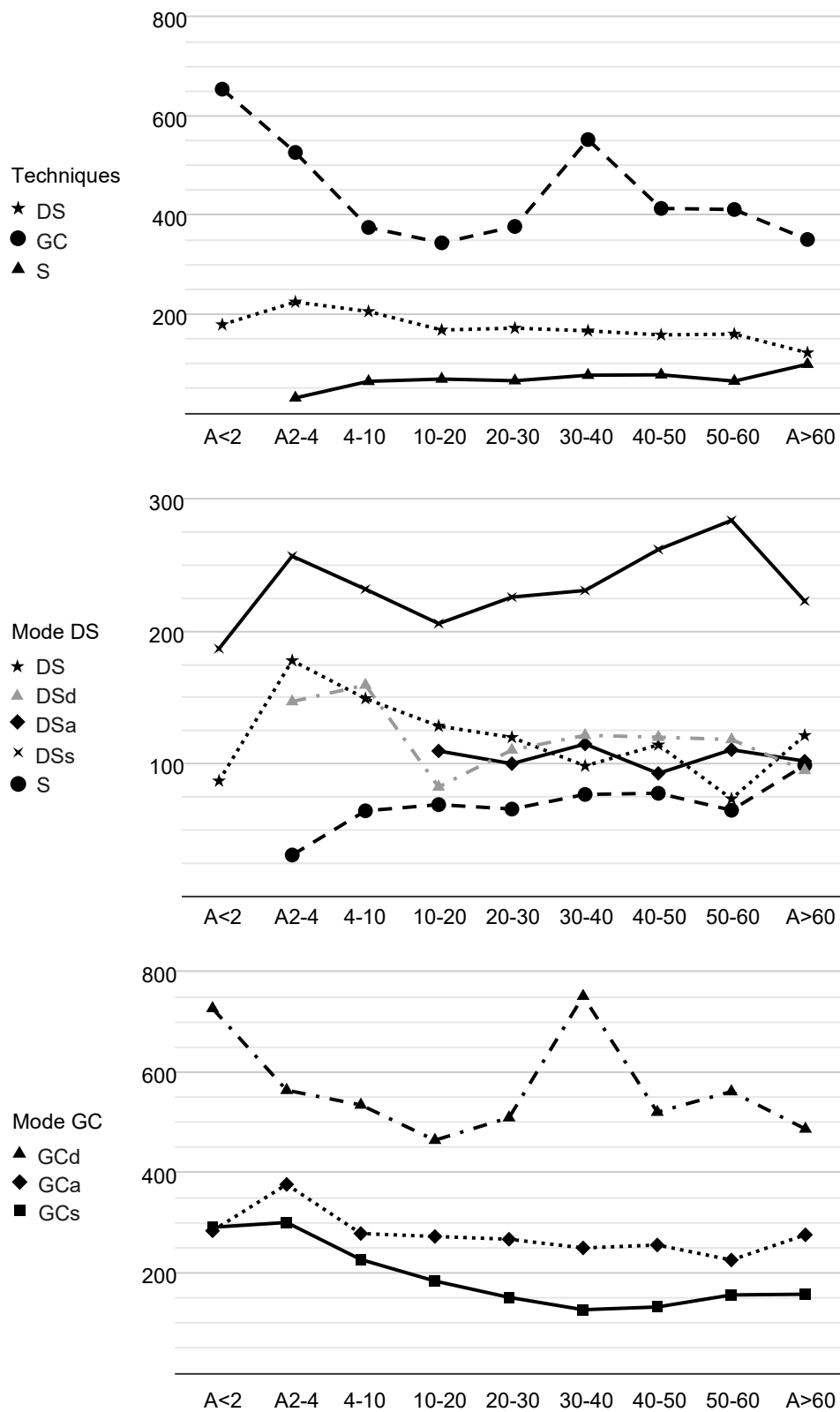


Figure 81 – Score C means by score A stratum. Top: techniques. Center: DS. Bottom: GC.

relationship was found with spatial ability on S technique, i.e., more spatial ability would be associated with longer reaction/choice time, which is counter-intuitive and will be further discussed.

Table 53 – Ss: correlations of other variables with score C

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	0,09	0,06	-	-	-	-
DS	-	-0,14	-	-	0,14	-
GC	0,27	-	-	-	-	-
S	-	-	-	-	0,25	0,22

#### A.2.3.4.1 Score C summary

Considering all the evidence gathered about score C, the following order (descending performance) is suggested:  $S > DS \geq DSd \geq DSa > DSs$  and  $GCs > GCa > GCd$ . This order shows a clear pattern, at least technique-wise, with the order  $S > DS > GC$ . However, the order  $2D > 2.5D > 3Da > 3Dst$  in DS was inverted for GC. One should note that score C was measured in logic tics, not in seconds. If we exclude GCd, the maximum difference is about three seconds ( $GCa \times S$ ), which one can consider negligible. When GCd is included, this difference rises to about five seconds. Therefore it seems reasonable to find why GCd differed so much from the others.

## A.2.4 Score D

### A.2.4.1 Distributions and outliers

Being a secondary score, this analysis was only performed for good choices or top 10 choices (Score A  $\leq$  10). As  $D = 1$  has a distinct meaning, those scores were also removed from the data sample. The remaining sample size has 1168 choices for DS (38 per participant) and 1121 for GC (35 per participant), without outliers. Concerning modes, sample sizes are almost equal per technique (Table 54).

Data distributions clearly deviate from normality and appear to follow a Gamma distribution (Figure 82, top). Therefore this score was analyzed using a GLMM, adopting the Gamma distribution family.

Table 54 – Ss: group size and outliers on score D

Technique	N	N (per mode)	Outliers
DS	1168	233	0
GC	1121	224	0



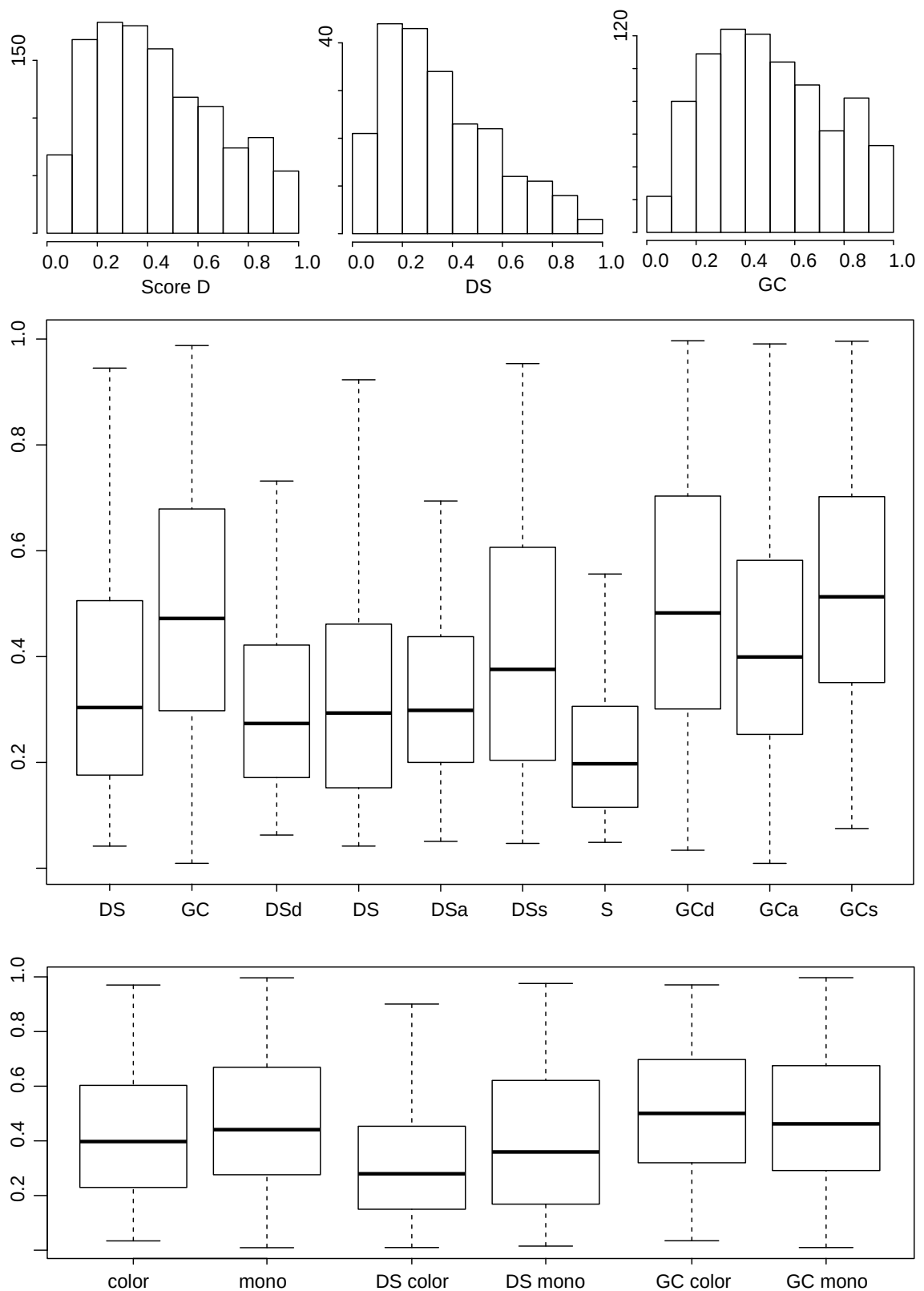


Figure 82 – Score D. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

#### A.2.4.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results were significant for techniques ( $p < 0.01$ ) and pairs ( $p < 0.01$ ). Boxplots (Figure 82, middle) show clear differences between Techniques and between GC, DS and S pairs. DS and GC variations seem not to be very different. Pairwise comparisons (Table 57) confirm that DSs stand out within DS (about 33 % higher), while GCa is about 20 % lower within GC. S has the lowest score, about 30 % lower than DS, while GC technique has a 35 % higher score than DS (Table 55). Differences in Color modes were significant both on the omnibus test and DS ( $p < 0.05$ ). DS mono had a 23 % higher score than Color (Table 56) and Mono surpasses Color by 11 % overall.

Table 55 – Ss: significant differences (%) per technique on score D

Technique	Mean	s.d.	Difference
DS	0,36	0,22	-26,3
GC	0,49	0,24	35,69

Table 56 – Ss: significant differences (%) per Color mode on score D

Color mode	Mean	s.d.	Difference
DS Color	0,33	0,21	-18,75
DS Mono	0,41	0,25	23,08
GC Color	0,51	0,24	-
GC Mono	0,49	0,24	-
Color	0,42	0,24	-10,37
Mono	0,47	0,24	11,57

Table 57 – Ss: significant differences (%) between pairs on score D

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	0,24	0,15	0	45,1	-	-	73,1	110	80,9	121,1
DS	0,34	0,24	-31,1	0	-	-	-	44,6	-	52,3
DSd	0,31	0,17	-	-	0	-	33,2	61,6	39,2	70,1
DSa	0,33	0,17	-	-	-	0	-	52	-	60
DSs	0,41	0,24	-42,2	-	-24,9	-	0	21,2	-	27,6
GCd	0,5	0,24	-52,3	-30,8	-38,1	-34,2	-17,5	0	-	-
GCa	0,43	0,23	-44,7	-	-28,1	-	-	-	0	22,2
GCs	0,53	0,23	-54,7	-34,3	-41,2	-37,5	-21,6	-	-18,1	0

#### A.2.4.3 Strata analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing the dynamics of choices with maximum score D, as in this special case,  $D = 1$  means that the selected grain was chosen in its first appearance. This was done by counting the proportion of  $D = 1$  per technique (Figure 83, top), pairs (Figure 83, middle) and per score A strata (Figure 83, bottom). By examining techniques, one can see a clear distinction between GC and DS/S (700 % higher). Concerning visualization modes, all DS modes are balanced, however in GC, the order  $GCs > GCa > GCd$  is suggested. Strata analysis shows that the distinction between DS and S lies in the 4-10 stratum.

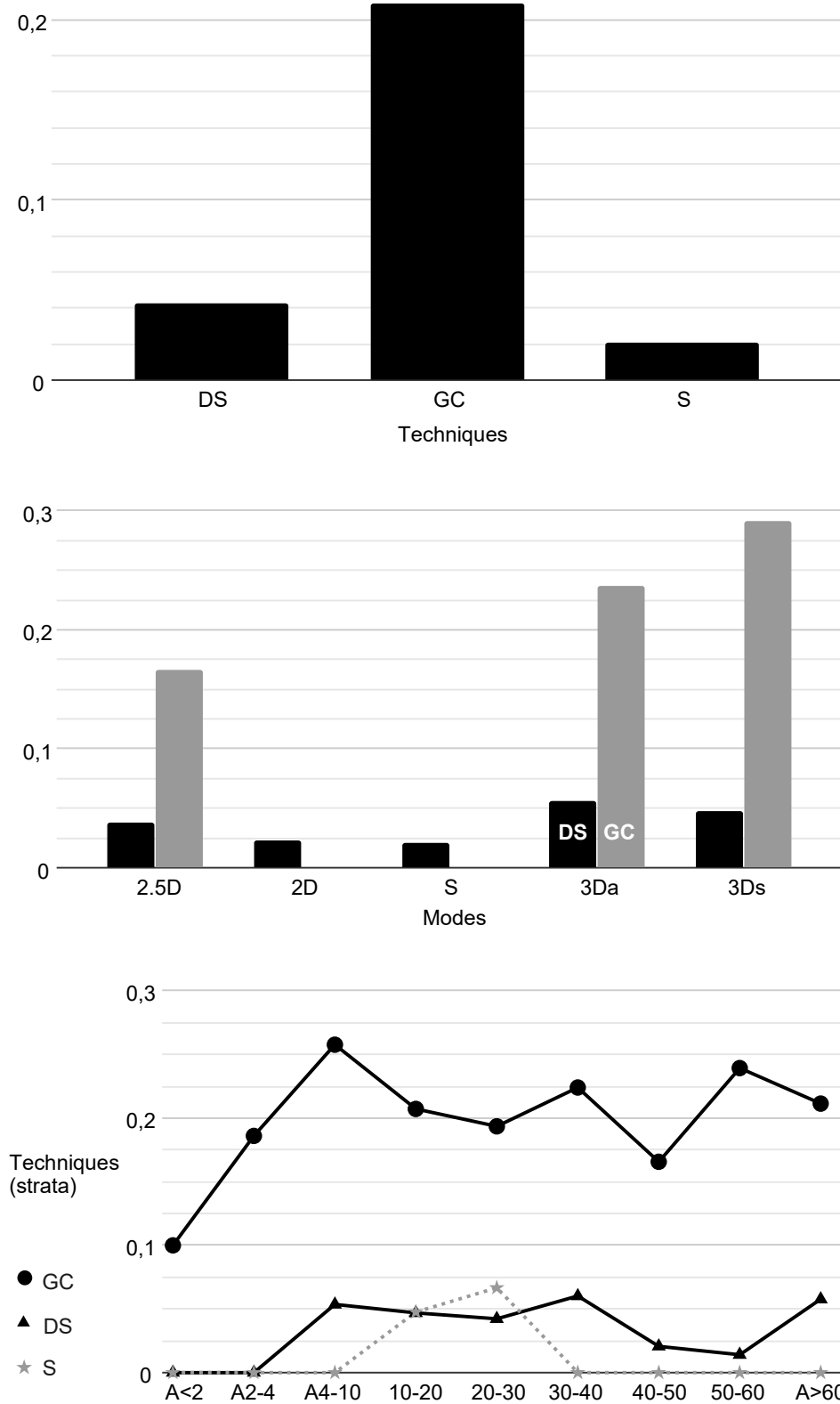


Figure 83 – Score D. Top: counts D = 1 per technique (%). Center: per DS and GC pairs. Bottom: per score A stratum.

#### A.2.4.4 Correlations

Table 58 shows weak (around 0.2) correlations between score D and selected variables (Time, Movement, Oscillations) or pre/post tests scores (MRT, SBSOD and TLX). It is worth noting that most correlations are all negative and related to movement and oscillations variables. This was expected as score D is time-sensitive, thus, as the participant moves or oscillates more, the score gets smaller.

Table 58 – Sb: correlations of other variables with score D

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	-	-	-	-	-	0,12
DS	-	-0,17	-0,11	-	-	0,16
GC	-	-0,12	-0,14	-	-	0,06
S	-	-	-0,24	-	-	-

##### A.2.4.4.1 Score D summary

Considering all the evidence gathered about score D, the following order (descending performance) is suggested: GCs > GCa > GCd > DSs and DSa >= DSd and DS > S

### A.3 Heterogeneous regions task: bigger grains clustering

#### A.3.1 Score A

##### A.3.1.1 Distributions and outliers

The sample size is 690 choices for DS (22 per participant) and 836 for GC (26 per participant), without outliers. Concerning modes, sample sizes are almost equal per technique (Table 59), differing about 17%, which is not a concern for statistical analysis.

Data distributions clearly deviate from normality and appear to follow no particular distribution (Figure 84, top). Therefore this score was analyzed using an Aligned Rank Transform ANOVA.

Table 59 – HRb: group size and outliers on score A

Technique	N	N (per mode)	Outliers
DS	690	138	0
GC	836	167,2	0

##### A.3.1.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results were significant for techniques ( $p < 0.01$ ) and pairs ( $p < 0.01$ ), however, modes cannot be compared within each technique, i.e., DS modes can only be compared to GC modes. Indeed, boxplots (Figure 84, middle) show noticeable differences

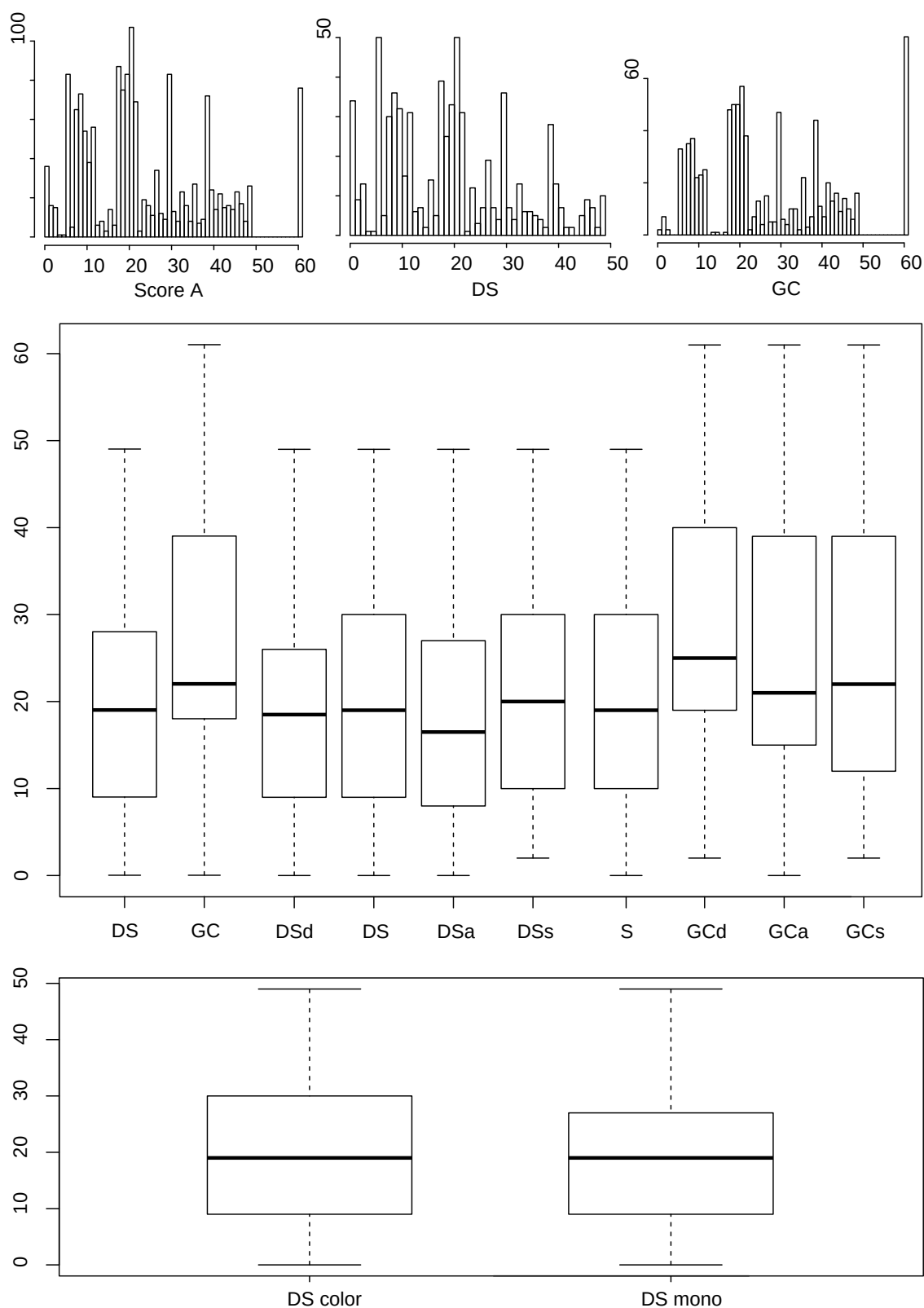


Figure 84 – Score A. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

between Techniques (Table 60) and between GC and DS Pairs. Inspecting pairwise comparisons (Table 62) DS and S differences to GC pairs range from 24 to 37 % less. Therefore, the main difference in this score is due to techniques, where DS has a score about 30 % lower. Differences in Color modes were not significant (Table 61).

Therefore we can conclude that DS and S have a significantly lower score A than GC, and color played no role in this score.

Table 60 – HRb: significant differences (%) per technique on score A

Technique	Mean	s.d.	Difference
DS	19,72	12,44	-29,03
GC	27,79	15,98	40,91

Table 61 – HRb: significant differences (%) per Color mode, score A

Color mode	Mean	s.d.	Difference
DS Color	20,13	12,68	-
DS Mono	19,18	12,13	-

Table 62 – HRb: significant differences (%) between pairs, score A

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	20,1	12,7	0	-	-	-	-	47,2	-	32,3
DS	19,96	13,6	-	0	-	-	-	48,2	32,9	33,1
DSd	18,68	11,59	-	-	0	-	-	58,4	42	42,3
DSa	17,79	12,41	-	-	-	0	-	66,3	49,1	49,4
DSs	21,39	12,32	-	-	-	-	0	38,4	24	-
GCd	29,6	16,22	-32,1	-32,5	-36,8	-39,8	-27,7	0	-	-
GCa	26,54	16,26	-24,2	-24,7	-29,5	-32,9	-19,4	-	0	-
GCs	26,59	15,25	-24,4	-24,9	-29,7	-33	-	-	-	0

### A.3.1.3 Strata analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing how choices are distributed between nine strata (Score A < 2, between 2 and 4, 4-10, 10-20, 20-30, 30-40, 40-50, 50-60, higher than 60). In this case, most choices should be within the first strata (A<10).

By examining techniques (Figure 85, top), the previous analysis is held, and no further distinctions are evident. Considering visualization modes on the DS technique (Figure 85, middle), DS and DSa seem to be slightly better at the first stratum (A<2), followed by DSd and S. In GC, no particular distinction is notable.

### A.3.1.4 Correlations

Table 63 shows that only a few weak correlations ( $r < 0.2$ ) were found between score A and selected variables (Time, Movement, Oscillations) or pre/post tests scores (MRT, SBSOD, and TLX) and the magnitude found here is too small to draw any conclusion.

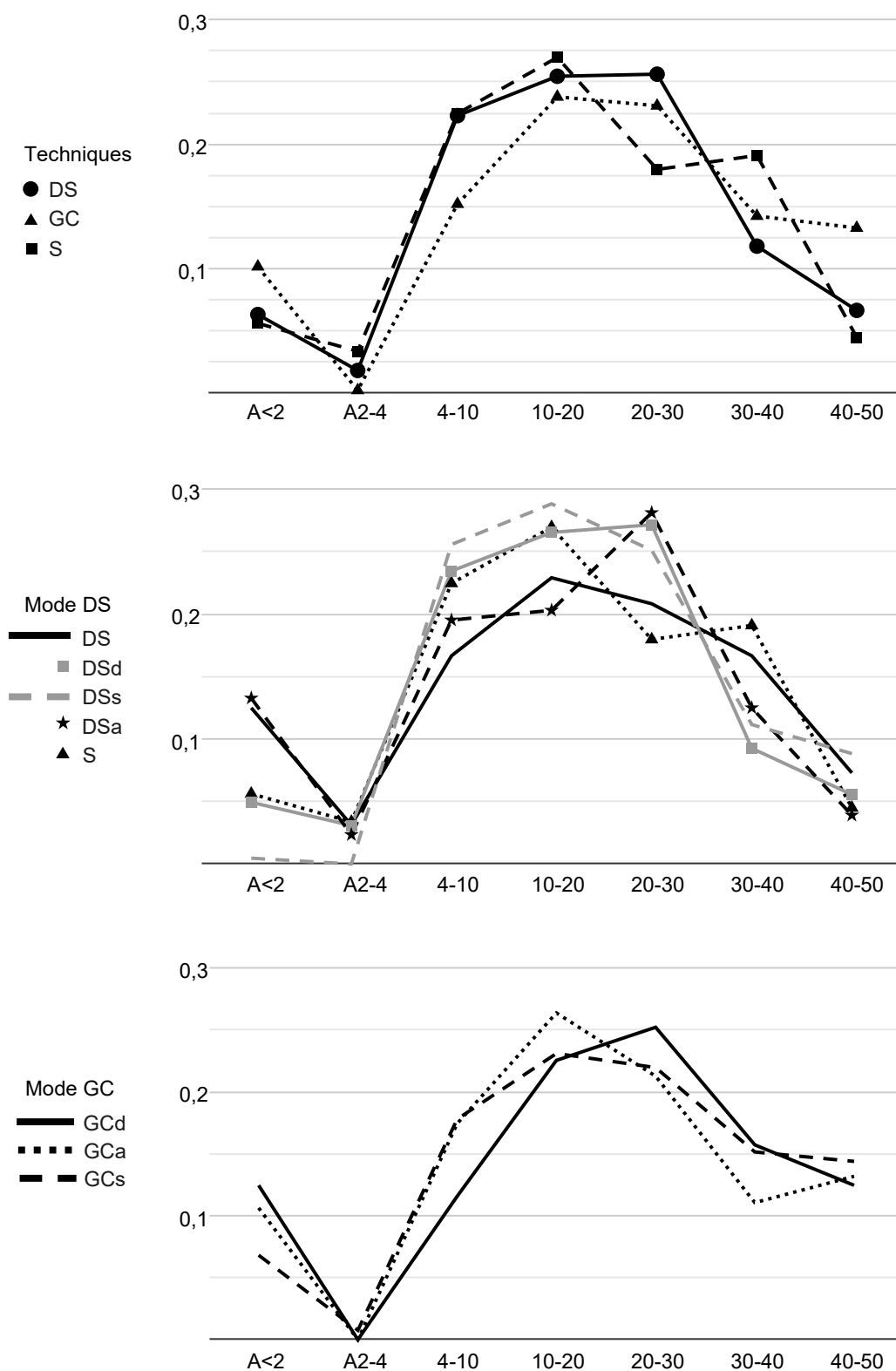


Figure 85 – Score A counts by stratum. Top: techniques. Center: DS pairs. Bottom: GC pairs.

Table 63 – HRb: correlations of other variables with score A

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	-	-	-	-	0,1	-
DS	-	-	-	-	-	0,15
GC	-	-	-	-	0,1	-
S	-	-	-	-	-	-

#### A.3.1.4.1 Score A summary

Considering all the evidence gathered about score A, the following order (descending performance) is suggested: DS and DSa  $\geq$  DSd and S  $>$  GC  $>$  DSs

### A.3.2 Score B

#### A.3.2.1 Distributions and outliers

The sample size is 517 choices for DS (17 per participant) and 678 for GC (21 per participant), with no outliers found. Concerning modes, sample sizes are almost equal per technique (Table 64), differing about 23%, which is not a concern for statistical analysis.

Data distributions clearly deviate from normality and appear to follow no particular distribution (Figure 86, top). Therefore this score was analyzed using an Aligned Rank Transform ANOVA.

Table 64 – HRb: group size and outliers on score B

Technique	N	N (per mode)	Outliers
DS	517	103,4	0
GC	678	135,6	0

#### A.3.2.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results was significant for techniques ( $p < 0.01$ ) and pairs ( $p < 0.01$ ). Boxplots (Figure 86, middle) show minor differences between Techniques (Table 65) and only GCa and GCs pairs stand out from the rest, with between 14 and 25 % lower scores (Table 67). Pairwise comparisons cannot be made among either DS or GC pairs. Differences in Color modes were not significant and the difference is negligible (Table 66; figure 86, bottom).

Therefore we can conclude that GCa and GCs pairs are responsible for the lower score of GC technique (7% lower than DS). Color played no role.

#### A.3.2.3 Strata analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing the dynamics of choices with maximum score B, as in this special case, B =



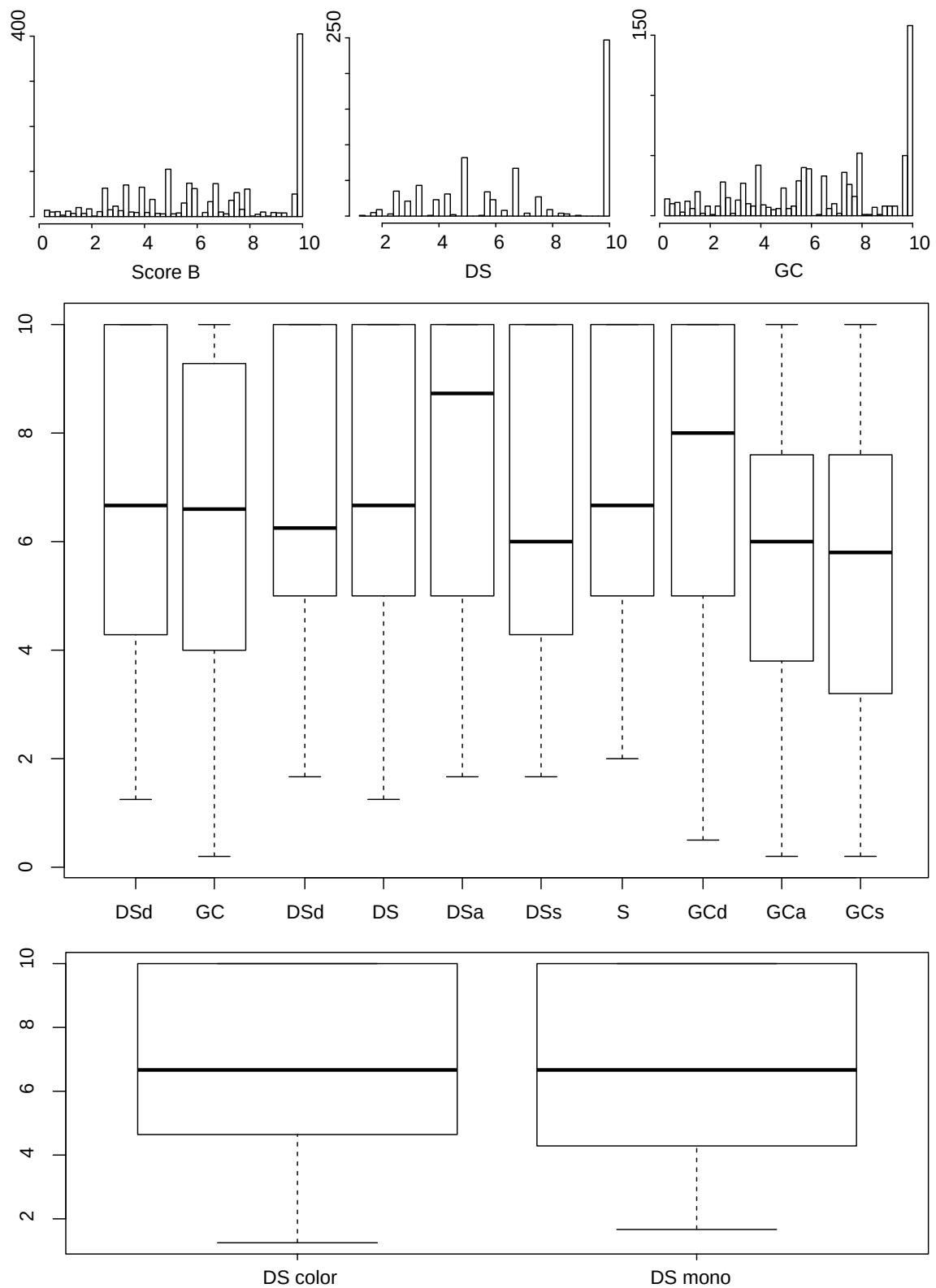


Figure 86 – Score B. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

Table 65 – HRb: significant differences (%) per technique, score B

Technique	Mean	s.d.	Difference
DS	6,74	2,8	7,23
GC	6,28	2,94	-6,75

Table 66 – HRb: significant differences (%) per Color mode, score B

Color mode	Mean	s.d.	Difference
DS Color	6,67	2,75	-
DS Mono	6,83	2,87	-

Table 67 – HRb: significant differences (%) between pairs, score B

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	6,52	2,7	0	-	-	-	-	-	-	-
DS	6,9	2,8	-	0	-	-	-	-	-17,2	-19,7
DSd	6,64	2,77	-	-	0	-	-	-	-14	-16,6
DSa	7,38	2,89	-	-	-	0	-	-	-22,7	-25
DSs	6,46	2,77	-	-	-	-	0	12,7	-	-14,3
GCd	7,28	2,76	-	-	-	-	-11,3	0	-21,6	-24
GCa	5,7	2,86	-	20,8	16,3	29,3	-	27,5	0	-
GCs	5,53	2,89	-	24,6	20	33,4	16,7	31,6	-	0

10 means that the best visible grain in the moment was chosen. This was done by counting the proportion of  $B = 10$  per technique (Figure 87, top) and pairs (Figure 87, middle). By examining techniques, DS is slightly better than S (23 %), and both are way superior to GC (65-105 %). Concerning visualization modes, one can see that 3Da stands out on DS, followed by 2D, while on GC, 2.5D is way superior to 3D modes.

#### A.3.2.4 Correlations

Table 68 shows that only a few weak correlations ( $r < 0.2$ ) were found between score B and selected variables (Movement, Oscillations), particularly on GC technique. Thus, the magnitude found here is too small to draw any conclusions.

Table 68 – HRb: correlations of other variables with score B

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	-	0,1	0,06	-	-	-
DS	-	-	-	-	-	-
GC	-	0,15	0,14	-	-	-
S	-	-	-	-	-	-

##### A.3.2.4.1 Score B summary

Considering all the evidence gathered about score B (ANOVA, counts of  $B = 10$ , and correlations), the following order (descending performance) is suggested: DSa > DS and GCd and DSd > S and DSs > GCa and GCs.

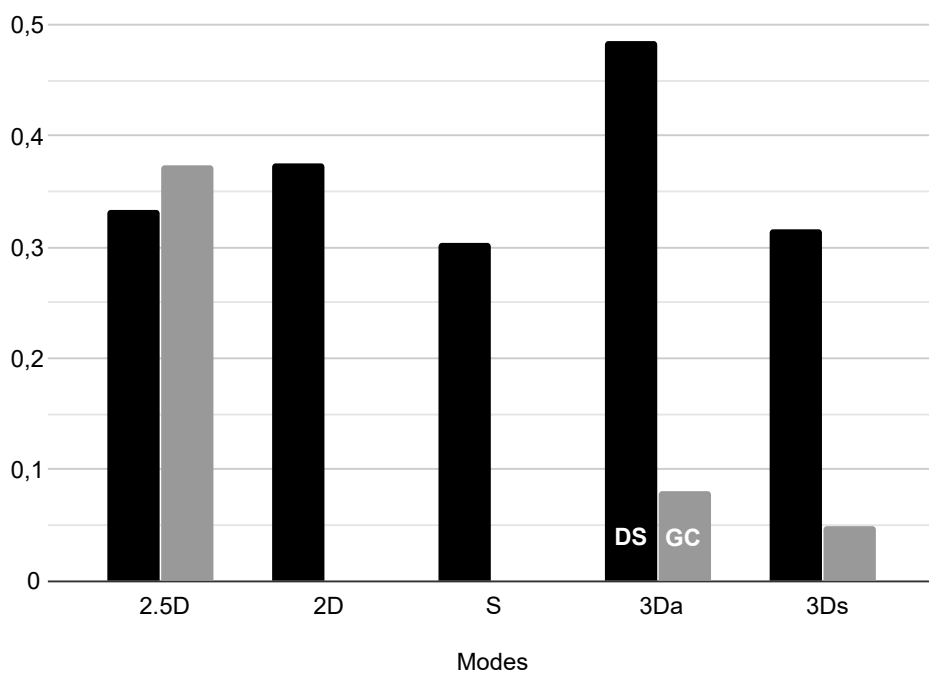
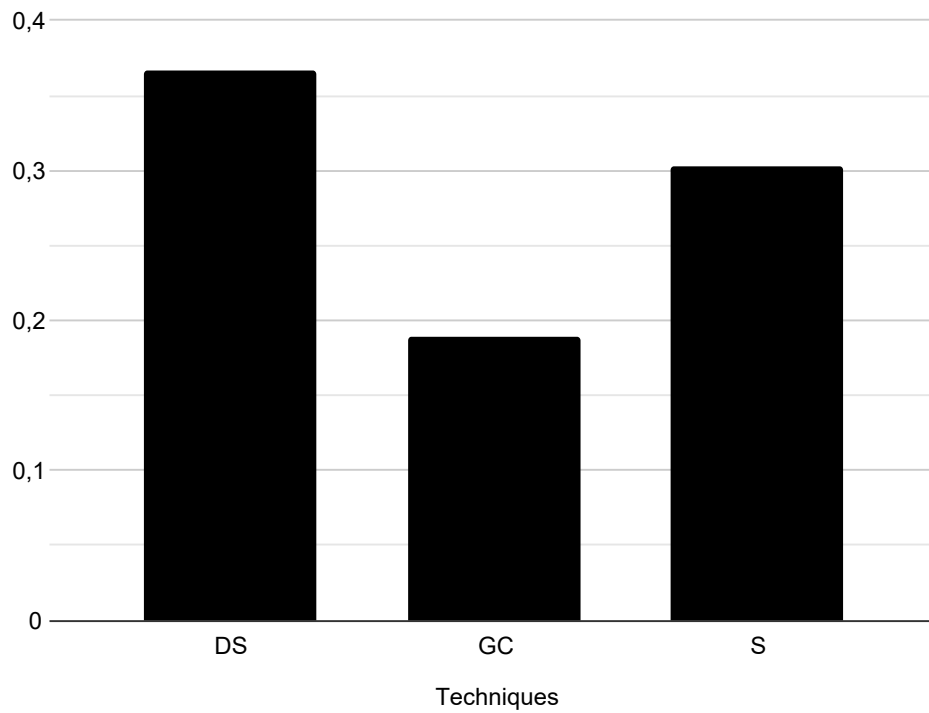


Figure 87 – Score B = 10 counts (%) by score A stratum. Top: techniques. Center: DS and GC pairs. Bottom: color modes.

### A.3.3 Score C

#### A.3.3.1 Distributions and outliers

Being a secondary score, this analysis was only performed for good choices or top 10 choices (Score A  $\leq 10$ ). The sample size is 211 choices for DS (7 per participant) and 138 for GC (4 per participant), with 30 outliers found (8.6 %). Concerning modes, sample sizes are almost equal per technique (Table 69).

Data distributions clearly deviate from normality, however, as DS and GC distributions are widely distinct (Figure 88, top), which was confirmed by a positive Kolmogorov-Smirnov test ( $p < 0.01$ ), techniques were directly compared. Each technique was then analyzed with a GLMM ANOVA for visualization mode differences, adopting Gamma distributions.

Table 69 – HRb: Group size and outliers on score C

Technique	N	N (per mode)	Outliers
DS	211	42,2	15
GC	138	46	15

#### A.3.3.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results was significant for visualization modes on both DS and GC ( $p < 0.01$ ). Technique differed substantially, where DS had a 95 % lower score than GC (Table 70). Boxplots (Figure 88, middle) show moderate differences within DS modes, where DSs stands out with a higher score (66-92 %) while within GC, GCd emerge with a score about 70 % lower (Table 72). Differences in Color modes on DS were not significant ( $p < 0.01$ ) and are negligible (Table 71 and figure 88, bottom ).

Therefore we can conclude that DS is better in general and GCd have the worst score, while DSs have the best.

Table 70 – HRb: significant differences (%) per technique, score C

Technique	Mean	s.d.	Difference
DS	82,38	46,04	-94,93
GC	1628,12	1262,44	1876,17

Table 71 – HRb: significant differences (%) per Color mode, score C

Color mode	Mean	s.d.	Difference
DS Color	81,92	42,82	-
DS Mono	82,96	50,02	-

#### A.3.3.3 Strata analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing the dynamics of score C along the nine strata of score A, which may show if

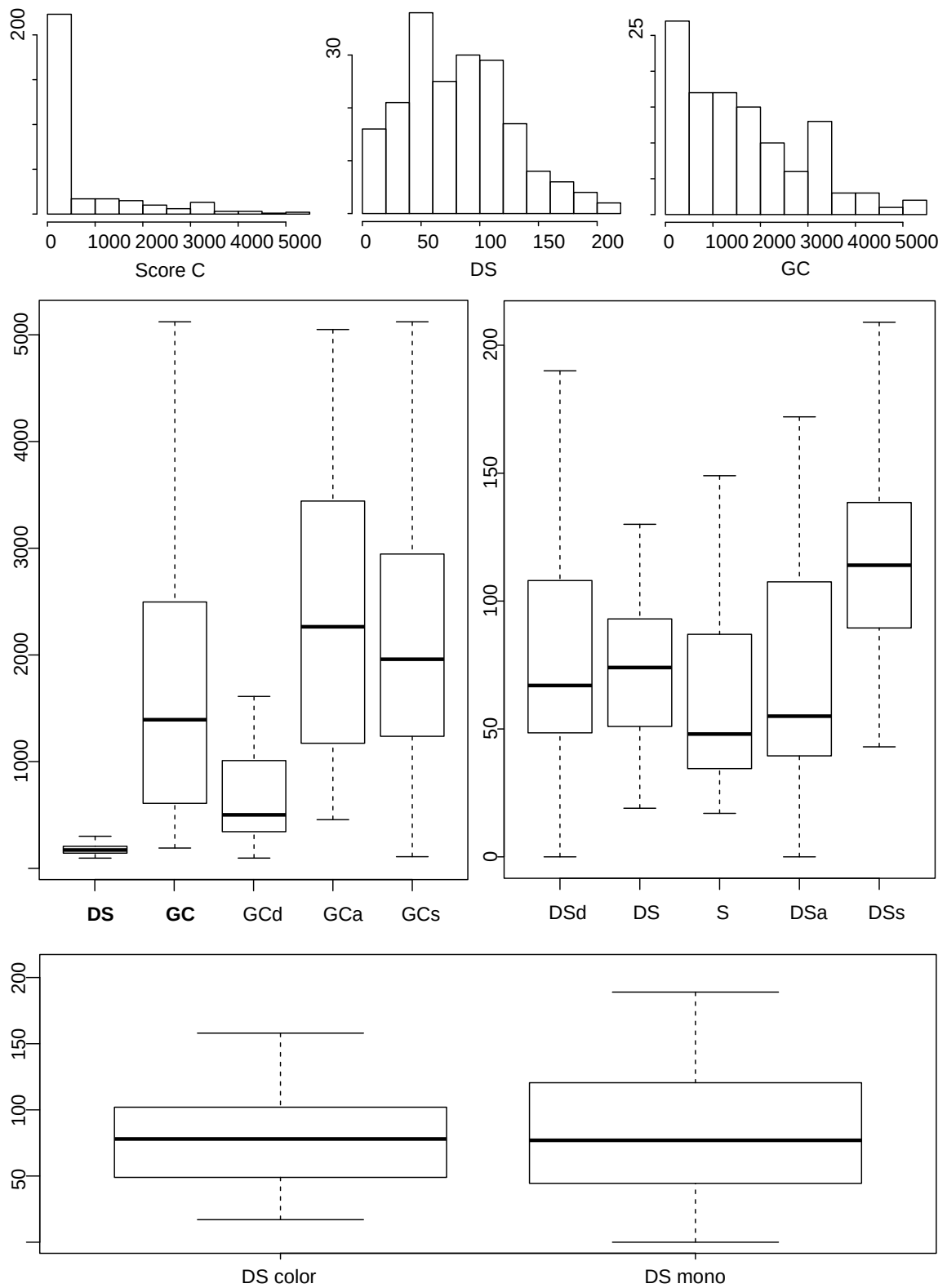


Figure 88 – Score C. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

Table 72 – HRb: significant differences (%) between pairs, score C

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	59,85	32,7	0	-	-	-	92,2	1019	4000,5	3283,3
DS	74,13	36,97	-	0	-	-	55,1	803,5	3210,6	2631,6
DSd	77,68	47,84	-	-	0	-	48,1	762,3	3059,6	2507
DSa	69	45,52	-	-	-	0	66,7	870,8	3457,1	2835
DSs	115,05	39,79	-47,9	-35,5	-32,4	-40	0	482,1	2033,2	1660,1
GCd	669,85	452,88	-91	-88,9	-88,4	-89,6	-82,8	0	266,4	202,3
GCa	2454	1389	-97,5	-96,9	-96,8	-97,1	-95,3	-72,7	0	-
GCs	2025	1170	-97	-96,3	-96,1	-96,5	-94,3	-66,9	-	0

any condition provides faster feature detection on the first three strata, i.e., of the top ten bigger grains in the sample. Overall, each condition was deemed very consistent across those strata.

Analysis on techniques (Figure 89, top) reinforces the substantial differences found between GC and the others, and how S and DS performance are pretty close. Within DS (Figure 89, middle), DSd is clearly worse than the others, followed by DSs. S seems a little better, particularly at the 2-4 stratum. Within GC (Figure 89, bottom), GCd is better at the 4-10 stratum, although no grain was detected on 2-4. GCs is clearly better than GCa on the first two strata (Score A < 4).

#### A.3.3.4 Correlations

Table 73 shows mostly weak correlations ( $r < 0.2$ ) between score C and selected variables (Time, Movement, Oscillations) or pre/post tests scores (MRT, SBSOD and TLX). Score C's Relationships with Time, Movement, and Oscillations are expected as all variables are related to time spent, and score C is essentially a time-related variable. However, the magnitude found here is too small to draw any conclusions.

Table 73 – HRb: correlations of other variables with score C

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	-	0,16	0,23	-0,14	0,06	-
DS	0,11	0,1	0,1	-	-	-
GC	-	-	-	-0,15	-	-
S	0,1	-	-	-	-	-

#### A.3.3.4.1 Score C summary

Considering all the evidence gathered about score C (ANOVA, strata analysis, and correlations), the following order (descending performance) is suggested: S > DS and DSa > DSs > DSd » GCd > GCs > GCa. Given the small difference when score C is converted to time in seconds, DSs surpassed DSd and changed positions in the ranking after the strata analysis.

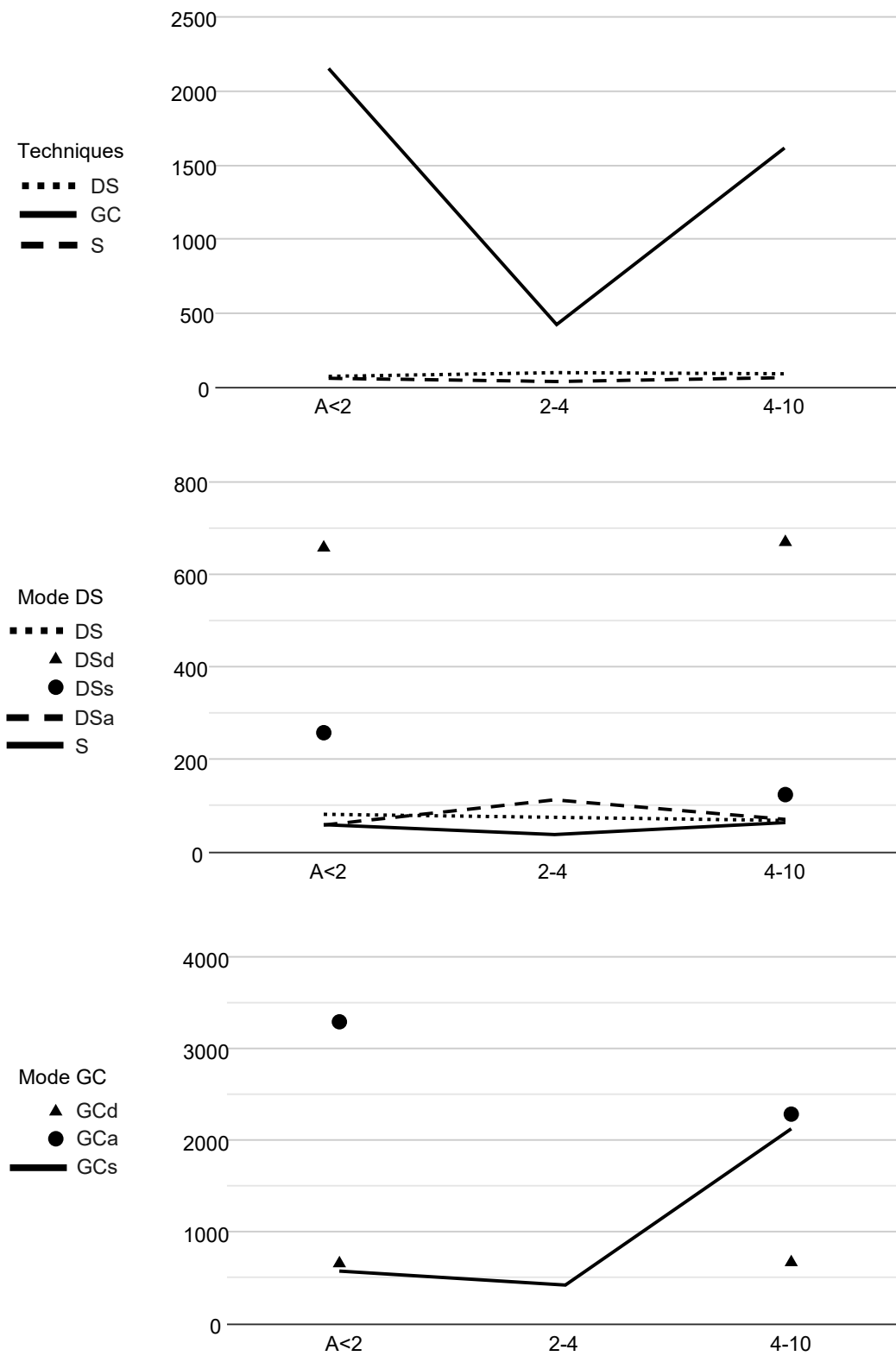


Figure 89 – Score C means by score A stratum. Top: techniques. Center: DS pairs. Bottom: GC pairs.

### A.3.4 Score D

#### A.3.4.1 Distributions and outliers

Being a secondary score, this analysis was only performed for good choices or top 10 choices (Score A  $\leq$  10). As D = 1 has a distinct meaning, those scores were removed from the data sample. The remaining sample size has 175 choices for DS (6 per participant) and 112 for GC (4 per participant), without outliers. Concerning modes, sample sizes are almost equal per technique (Table 74).

Data distributions deviate from normality, and due to the small data set (Figure 90, top), the Aligned Rank Transform ANOVA was used in this case.

Table 74 – HRb: group size and outliers on score D

Technique	N	N (per mode)	Outliers
DS	175	35	0
GC	112	37	0

#### A.3.4.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA was not significant for techniques, pairs, nor color mode. Indeed, boxplots (Figure 90, middle) shows no clear differences between Techniques and any other pair and means are very close (Tables 75, 76 and 77). GCa and GCs are not displayed as all choices yielded a score D of 1.

Table 75 – HRb: significant differences (%) per technique, score D

Technique	Mean	s.d.
DS	0,43	0,23
GC	0,4	0,2

Table 76 – HRb: significant differences (%) per Color mode, score D

Color mode	Mean	s.d.
DS Color	0,43	0,22
DS Mono	0,43	0,19

Table 77 – HRb: significant differences (%) between pairs, score D

	Mean	s.d.
S	0,38	0,22
DS	0,41	0,25
DSd	0,4	0,24
DSa	0,46	0,26
DSs	0,47	0,21
GCd	0,4	0,2
GCa	-	-
GCs	-	-



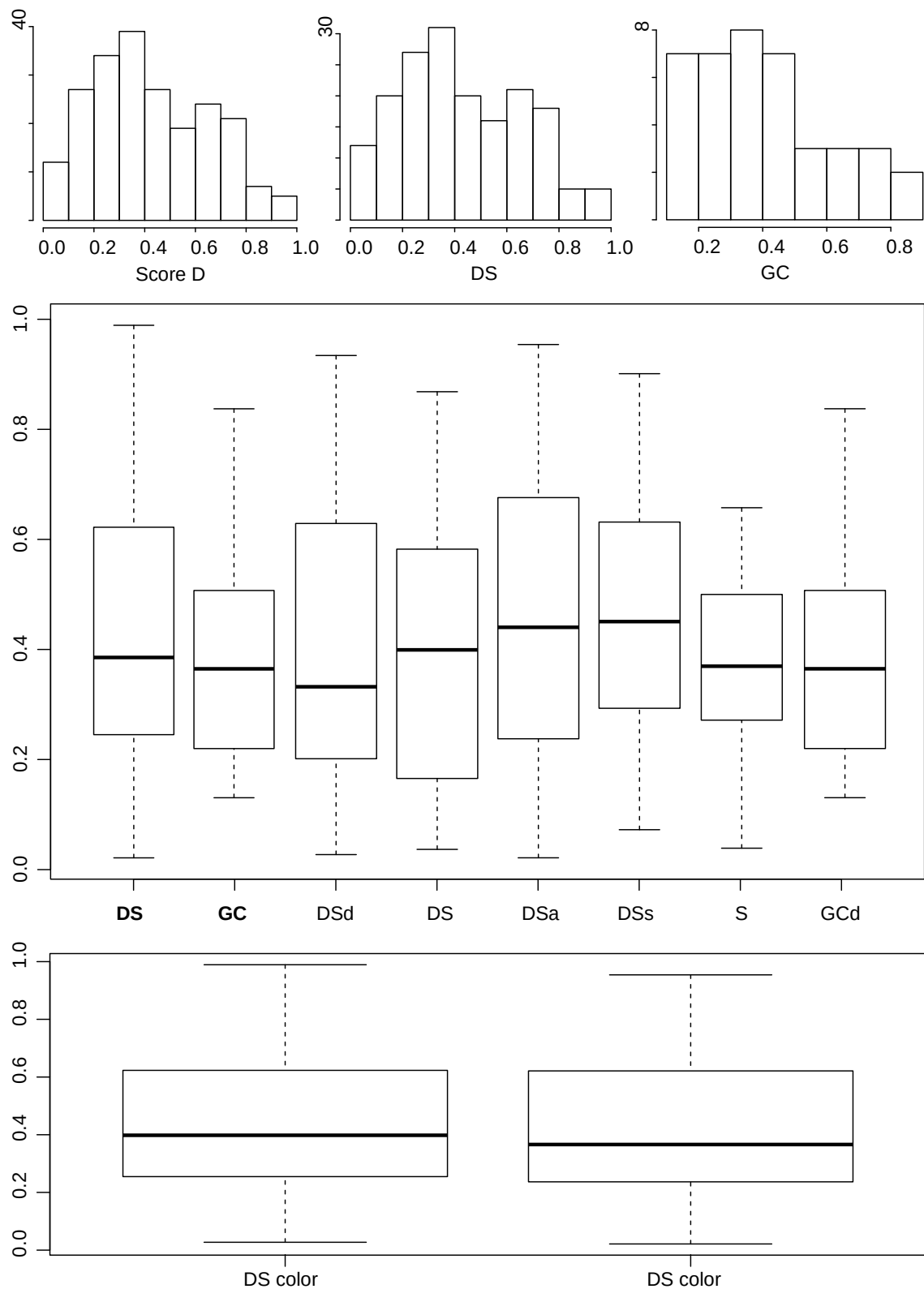


Figure 90 – Score D. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

### A.3.4.3 Strata analysis

The purpose of this analysis, in this case, is to find any distinction by detailing the dynamics of choices with maximum score D, as in this special case,  $D = 1$  means that the selected grain was chosen in its first appearance. This was done by counting the proportion of  $D = 1$  per technique (Figure 91, top), pairs (Figure 91, middle) and per score A strata (Figure 91, bottom).

By examining techniques, one can see that GC clearly stands out by six times. Concerning visualization modes, 3D modes clearly emerge both within GC and DS. Strata analysis helps to set DS apart from S, with better performance on 2-4 stratum.

### A.3.4.4 Correlations

Table 78 shows weak (around 0.2) and moderate correlations (around 0.5) between score D and selected variables (Time, Movement, Oscillations) or pre/post tests scores (MRT, SBSOD and TLX). It is worth noting that the stronger correlations are all negative and related to the first three variables. This was expected as score D is time-sensitive, thus, as the participant moves or oscillates more, the score is supposed to get smaller. Notably, the table shows that S is more sensitive to oscillations, with a moderate correlation of 0.5. Score D gets smaller every time a given grain is missed, which may happen more often due to the larger jumps in S.

Table 78 – HRb: correlations of other variables with score D

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	-0,16	-0,14	-	-	0,07	0,1
DS	-0,27	-0,31	-0,36	-	0,21	0,15
GC	-0,1	-0,31	-0,26	0,1	-	-
S	-0,4	-0,27	-0,5	-	-	-

#### A.3.4.4.1 Score D summary

Considering all the evidence gathered about score D, particularly counts of score  $D = 1$ , the following order (descending performance) is suggested: GCs and GCa > DSs and DSa > GCd and DSd > DS and 4S.

## A.4 Heterogeneous regions task: smaller grains clustering

### A.4.1 Score A

#### A.4.1.1 Distributions and outliers

The sample size is 749 choices for DS (24 per participant) and 604 for GC (19 per participant), without outliers. Concerning modes, sample sizes are almost equal per

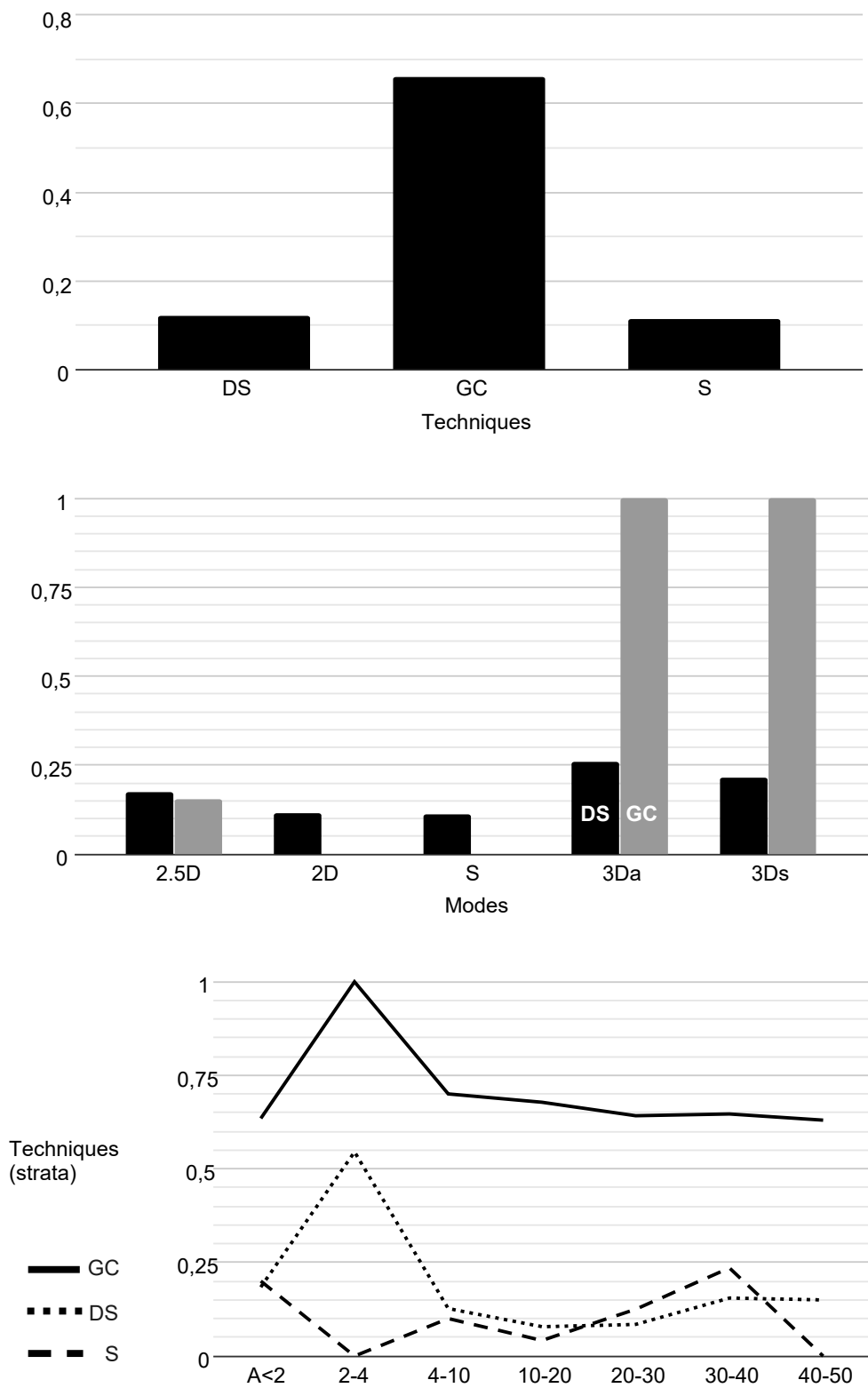


Figure 91 – Score D. Top: counts D = 1 per technique (%). Center: per DS and GC pairs. Bottom: per score A stratum.

technique (Table 79), differing about 25%, which is not a concern for statistical analysis.

Data distributions clearly deviate from normality and appear to follow the Weibull distribution (Figure 92, top). Therefore this score was analyzed using a GLMM, adopting the Weibull distribution family.

Table 79 – HRs: group size and outliers on score A

Technique	N	N (per mode)	Outliers
DS	749	150	0
GC	604	201	0

#### A.4.1.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results were significant for techniques ( $p < 0.01$ ) and pairs ( $p < 0.01$ ), however, modes cannot be compared within each technique, i.e., DS modes can only be compared to GC modes. Indeed, boxplots (Figure 92, middle) shows only mild differences between Techniques and between GC and DS Pairs. Inspecting pairwise comparisons (Table 82) DS and S differences to GC pairs range from 23 to 29 % less. Therefore, the main difference in this score is due to techniques, where DS has a score about 26 % lower (Table 80). Differences in Color modes were not significant and are negligible (Table 81).

Therefore we can conclude that DS and S have a slightly smaller score A than GC, and color played no role in this score.

Table 80 – HRs: significant differences (%) per technique, score A

Technique	Mean	s.d.	Difference
DS	11,45	12,42	-25,68
GC	15,41	14,18	34,56

Table 81 – HRs: significant differences (%) per Color mode, score A

Color mode	Mean	s.d.	Difference
DS Color	11,66	12,75	-
DS Mono	11,2	12,04	-

Table 82 – HRs: significant differences (%) between pairs, score A

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	11,63	12,01	0	-	-	-	-	35,1	-	34,1
DS	12,08	13,06	-	0	-	-	-	30	-	29,1
DSd	11,59	12,77	-	-	0	-	-	35,5	27,5	34,5
DSa	11,07	12	-	-	-	0	-	41,9	33,5	40,9
DSs	11,18	12,48	-	-	-	-	0	40,5	-	39,4
GCd	15,72	14,17	-26	-23,1	-26,2	-29,5	-28,8	0	-	-
GCa	14,78	14,33	-	-	-21,5	-25,1	-	-	0	-
GCs	15,6	14,12	-25,4	-22,5	-25,6	-29	-28,3	-	-	0

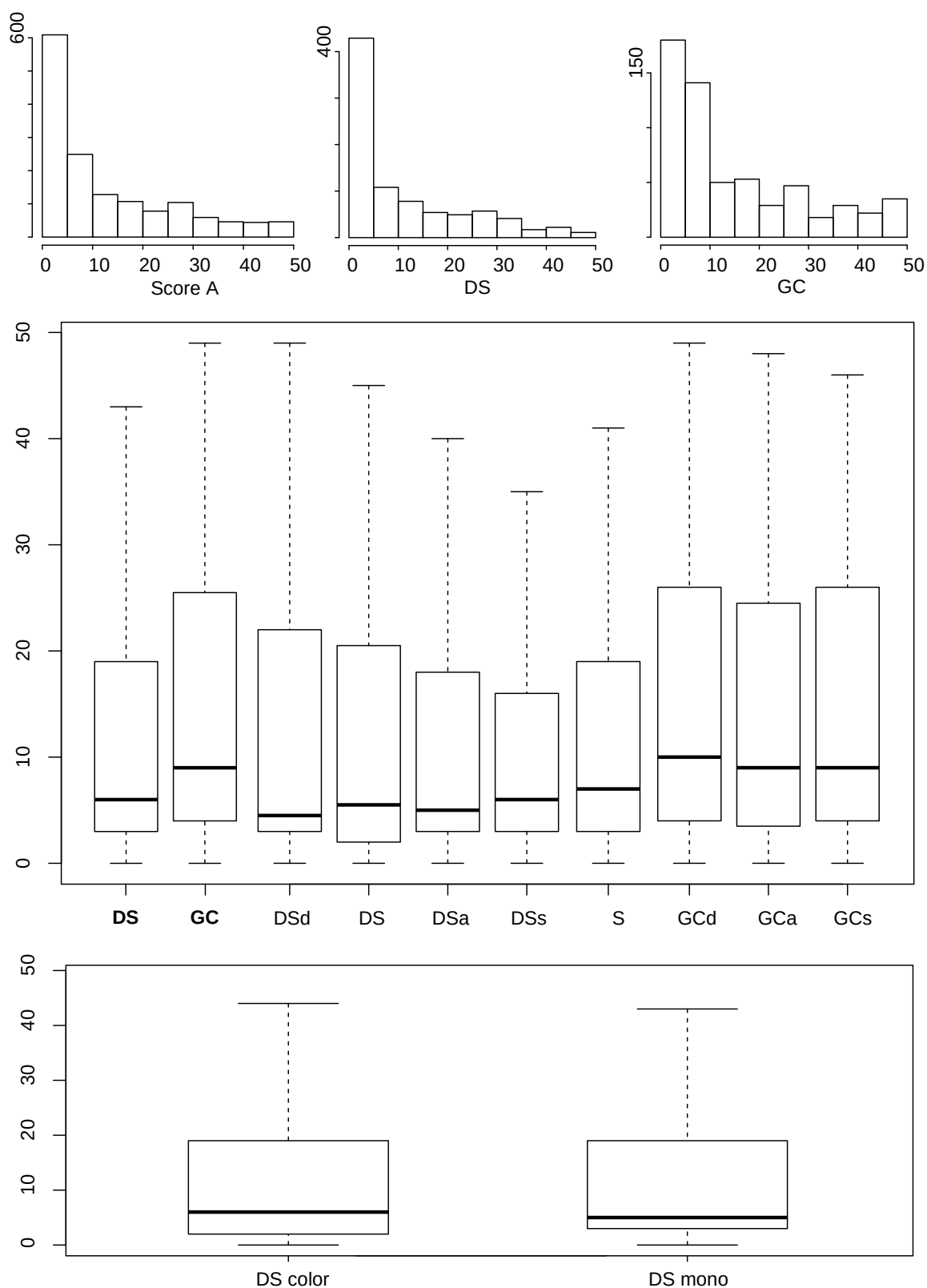


Figure 92 – Score A. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

### A.4.1.3 Strata analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing how choices are distributed between nine strata (Score  $A < 2$ , between 2 and 4, 4-10, 10-20, 20-30, 30-40, 40-50, 50-60, higher than 60). In this case, most choices should be within the first strata ( $A < 10$ ).

By examining techniques (Figure 93, top), DS, and S are better than GC in the first two strata (Score  $A < 4$ ). Visualization modes performed quite similarly on the DS technique (Figure 93, middle) with a small advantage of DSd, followed by DS. On GC, GCa has a small lead over the others on the first two strata.

### A.4.1.4 Correlations

Table 83 shows only a few weak (around 0.2) between score A and selected variables (Time, Movement, Oscillations) or pre/post tests scores (MRT, SBSOD, and TLX). However, the magnitude found here is too small to draw any conclusions.

Table 83 – HRs: correlations of other variables with score A.

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	-	-	-	-	-	0,06
DS	-	-	0,08	-	-	-
GC	-	-	-	-	-	0,13
S	-	-	-	-	-	-

#### A.4.1.4.1 Score A summary

Considering all the evidence gathered about score A, especially strata analysis, the following order is suggested: DSd and DS  $>$  DSa and DSs  $>$  S  $>$  GCa  $>$  GCd and GCs.

## A.4.2 Score B

### A.4.2.1 Distributions and outliers

The sample size is 749 choices for DS (24 per participant) and 604 for GC (19 per participant), with no outliers found. Concerning modes, sample sizes are almost equal per technique (Table 84), differing about 25%, which is not a concern for statistical analysis.

Data distributions clearly deviate from normality and appear to follow a Beta distribution (Figure 94, top). For the sake of simplicity, as GLMMs for such distributions are not well documented, the distribution was inverted to a Weibull distribution, and this score was analyzed adopting the Weibull distribution family.

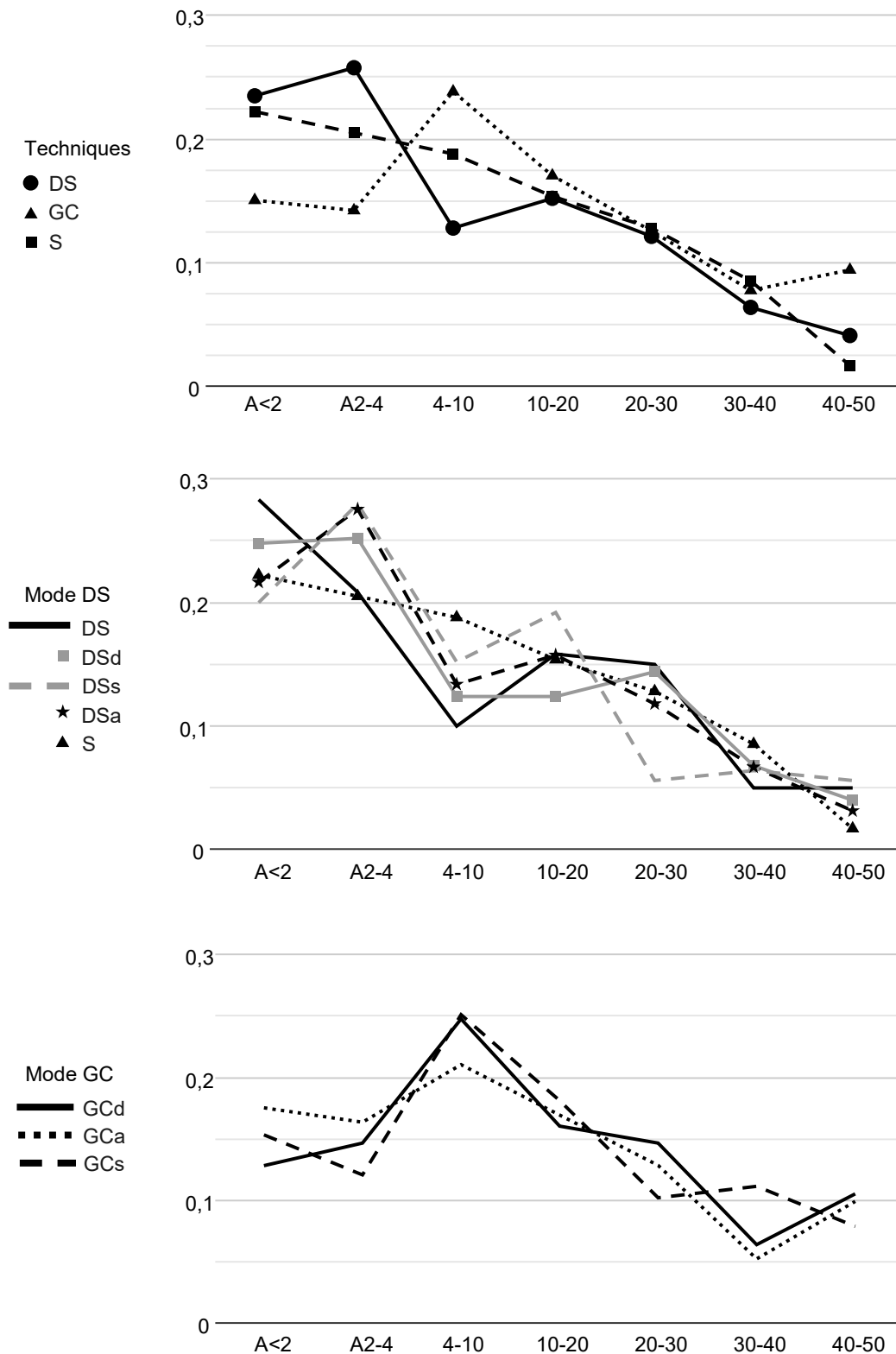


Figure 93 – Score A counts by stratum. Top: techniques. Center: DS pairs. Bottom: GC pairs.

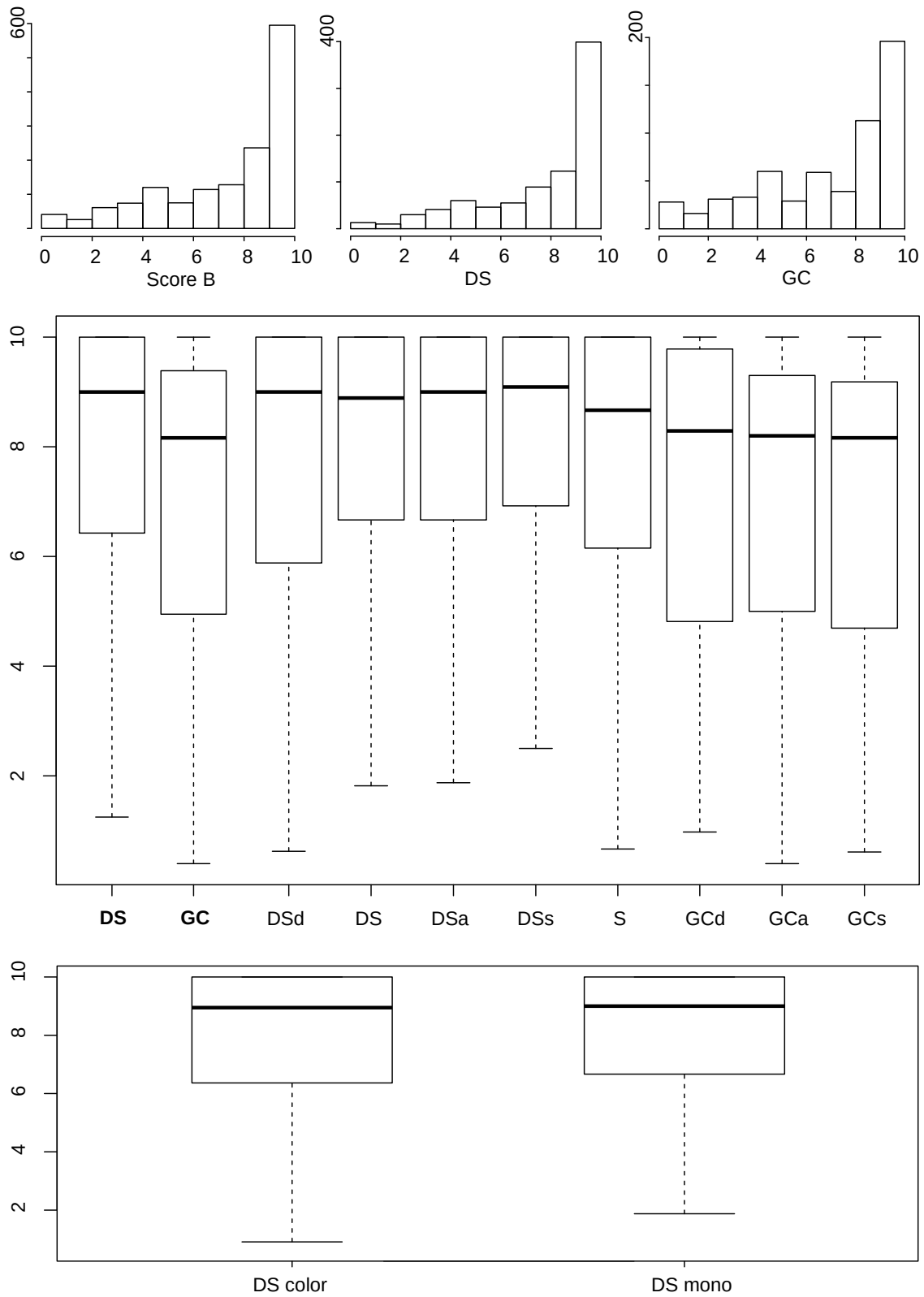


Figure 94 – Score B. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.



Table 84 – HRs: group size and outliers on score B

Technique	N	N (per mode)	Outliers
DS	749	150	0
GC	604	201	0

#### A.4.2.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results were significant for techniques ( $p < 0.01$ ) and pairs ( $p < 0.01$ ), however, modes cannot be compared within each technique, i.e., DS modes can only be compared to GC modes. Boxplots (Figure 94, middle) show mild differences between Techniques (Table 85) and only GCa and GCs pairs stand out from the rest, with between 11 and 15 % lower scores (Table 87). Techniques differ by 12 % in favor of DS (Table 85). Differences in Color modes were not significant, and the difference is negligible (Table 86).

Therefore we can conclude that GCa and GCs pairs are responsible for the lower score of GC technique. Color played no role.

Table 85 – HRs: significant differences (%) per technique, score B

Technique	Mean	s.d.	Difference
DS	7,84	2,42	12,25
GC	6,98	2,81	-10,92

Table 86 – HRs: significant differences (%) per Color mode on score B

Color mode	Mean	s.d.	Difference
DS Color	7,79	2,49	-
DS Mono	7,9	2,34	-

Table 87 – HRs: significant differences (%) between pairs, score B

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	7,82	2,23	0	-	-	-	-	-	-	-12,9
DS	7,73	2,63	-	0	-	-	-	-	-	-11,8
DSd	7,71	2,51	-	-	0	-	-	-	-	-11,6
DSa	7,92	2,33	-	-	-	0	-	-	-11,2	-14
DSs	8,06	2,38	-	-	-	-	0	-11,6	-12,7	-15,4
GCd	7,12	2,7	-	-	-	-	13,2	0	-	-
GCa	7,03	2,86	-	-	-	12,6	14,6	-	0	-
GCs	6,81	2,88	14,8	13,4	13,2	16,3	18,3	-	-	0

#### A.4.2.3 Strata analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing the dynamics of choices with maximum score B, as in this special case,  $B = 10$  means that the best visible grain at the moment was chosen. This was done by counting the proportion of  $B = 10$  per technique (Figure 95, top), pairs (Figure 95, middle) and color modes (Figure 95, bottom).

By examining techniques, S and DS are more than twofold better than GC. Concerning visualization modes, one can see balance within DS modes with a small advantage of DS and DSs over DSa and DSd, while in GC, GCd is way superior to 3D modes.

#### A.4.2.4 Correlations

Table 88 shows that only a few weak correlations ( $r < 0.2$ ) were found between score B and selected variables (Time, Movement, Oscillations) or pre/post tests scores (MRT, SBSOD, and TLX). However, the magnitude found here is too small to draw any conclusions.

Table 88 – HRs: correlations of other variables with score B.

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	-	-0,05	-0,06	-	-	-
DS	-	-	-0,07	-	-	-
GC	-	-	-	-	-	-0,11
S	-	-	-	-	-	-

##### A.4.2.4.1 Score B summary

Considering all the evidence gathered about score B, particularly counts of  $B = 10$ , the following order (descending performance) is suggested: DSs and DS > DSd and DSa and S > GCd > GCa > GCs.

### A.4.3 Score C

#### A.4.3.1 Distributions and outliers

Being a secondary score, this analysis was only performed for good choices or top 10 choices (Score A  $\leq 10$ ). The sample size is 497 choices for DS (16 per participant) and 257 for GC (8 per participant), without outliers. Concerning modes, sample sizes are almost equal per technique (Table 89), differing about 13%, which is not a concern for statistical analysis.

Data distributions clearly deviate from normality, however, as DS and GC distributions are widely distinct (Figure 96, top), which was confirmed by a positive Kolmogorov-Smirnov test ( $p < 0.01$ ), techniques were directly compared. Each technique was then analyzed with a GLMM ANOVA for visualization mode differences, adopting Gamma distributions.

Table 89 – HRs: group size and outliers on score C

Technique	N	N (per mode)	Outliers
DS	497	99,4	0
GC	257	86	0

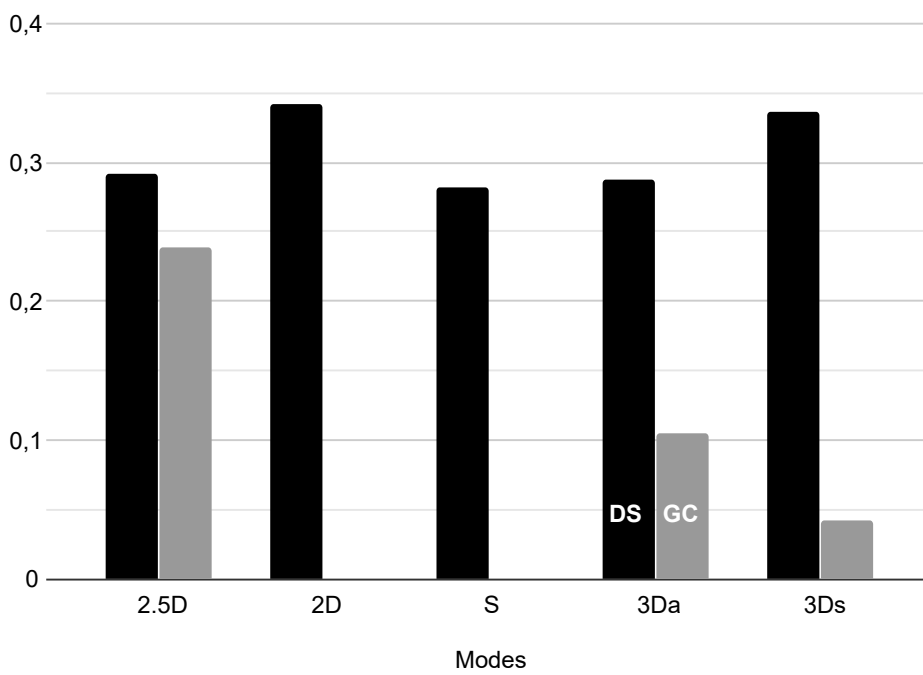
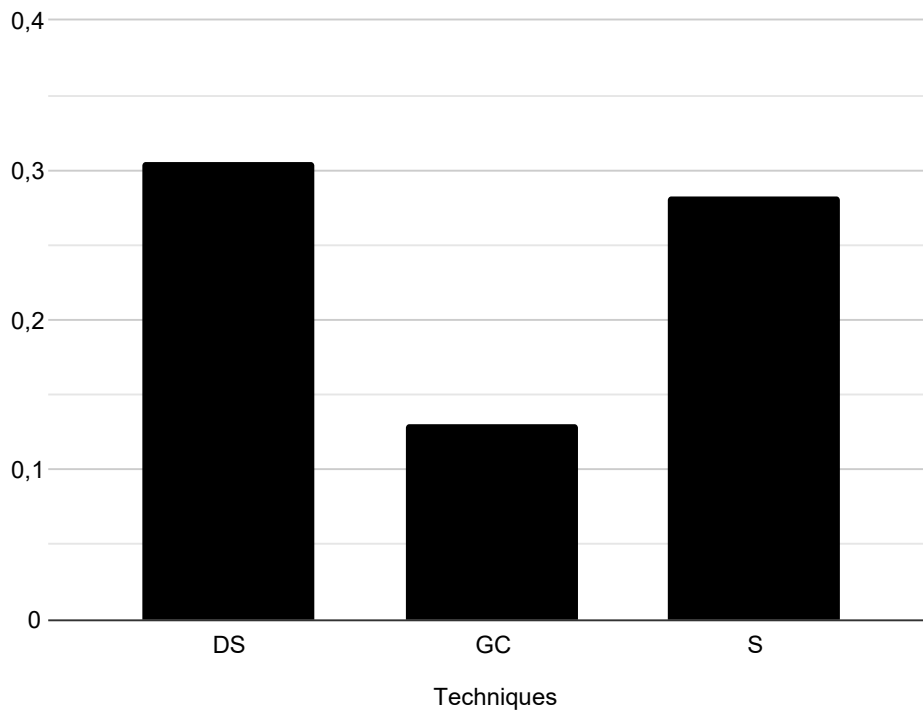


Figure 95 – Score B = 10 counts (%) by score A stratum. Top: techniques. Center: DS and GC pairs. Bottom: color modes.

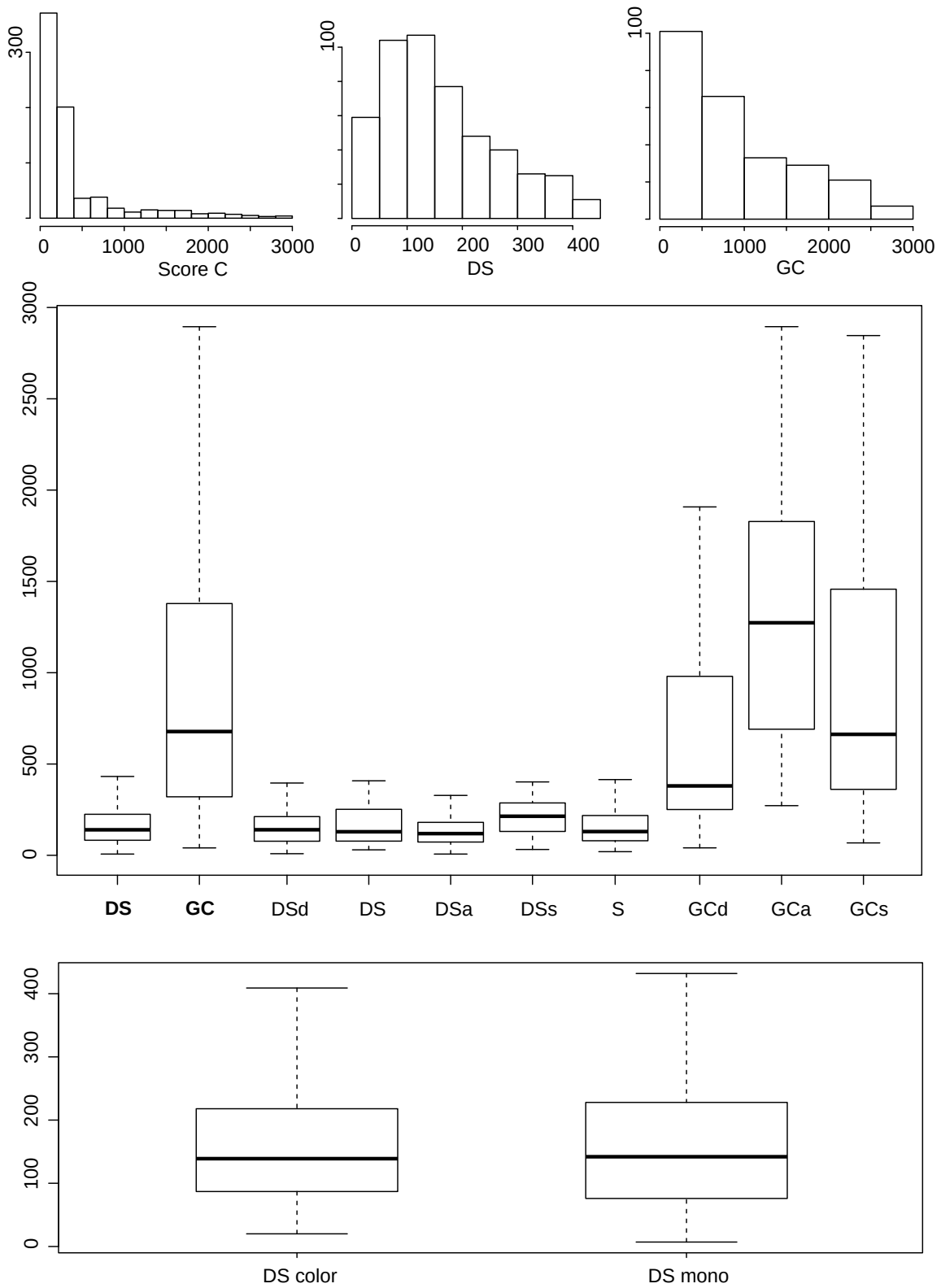


Figure 96 – Score C. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.

### A.4.3.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results was significant for visualization modes on both DS and GC ( $p < 0.01$ ). Technique differed substantially, where DS had a 82 % lower score than GC (Table 90). Boxplots (Figure 96, middle) show moderate differences within DS modes, where DSs stands out with a higher score (33-52 %) while larger differences are visible within GC, where GCd scores are smaller than GCa and GCs, by 50 and 28 %, respectively (Table 92). Differences in Color modes on DS were not significant ( $p < 0.01$ ) and are negligible (Table 91 and Figure 96, bottom ).

Therefore we can conclude that DS and S are better in general, and GCa has the worst score, while S and DS, except for DSs, have the best.

Table 90 – HRs: significant differences (%) per technique, score C

Technique	Mean	s.d.	Difference
DS	161,71	101,97	-82,23
GC	910,32	719,67	462,9

Table 91 – HRs: significant differences (%) per Color mode on score C

Color mode	Mean	s.d.	Difference
DS Color	164,16	101,23	-
DS Mono	159,01	102,93	-

Table 92 – HRs: significant differences (%) between pairs, score C

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	160,66	104,93	0	-	-	-	33,2	318,1	741,6	486,1
DS	169,64	112,26	-	0	-	-	-	295,9	697	455,1
DSd	157,04	102,3	-	-	0	-	36,2	327,7	761	499,6
DSa	140,41	91,82	-	-	-	0	52,4	378,4	863	570,6
DSs	214,02	94,76	-24,9	-	-26,6	-34,3	0	213,8	531,8	340
GCd	671	586,71	-76	-74,7	-76,6	-79	-68,1	0	101,2	40,1
GCa	1352	755,85	-88,1	-87,4	-88,3	-89,6	-84,1	-50,3	0	-
GCs	941,77	724,19	-82,9	-81,9	-83,3	-85	-77,2	-28,6	-	0

### A.4.3.3 Strata analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing the dynamics of score C along the nine strata of score A, which may show if any condition provides faster feature detection on the first three strata, i.e., of top 10 bigger grains. Overall, each condition was deemed very consistent across those strata.

Analysis of techniques (Figure 97, top) reinforces the substantial differences found between GC and the others, and how S and DS performance are pretty close overall. Within DS (Figure 97, middle), 2D is slightly worse than S, followed by DSa, while within GC (Figure 97, bottom), the analysis shows GCd as the best.

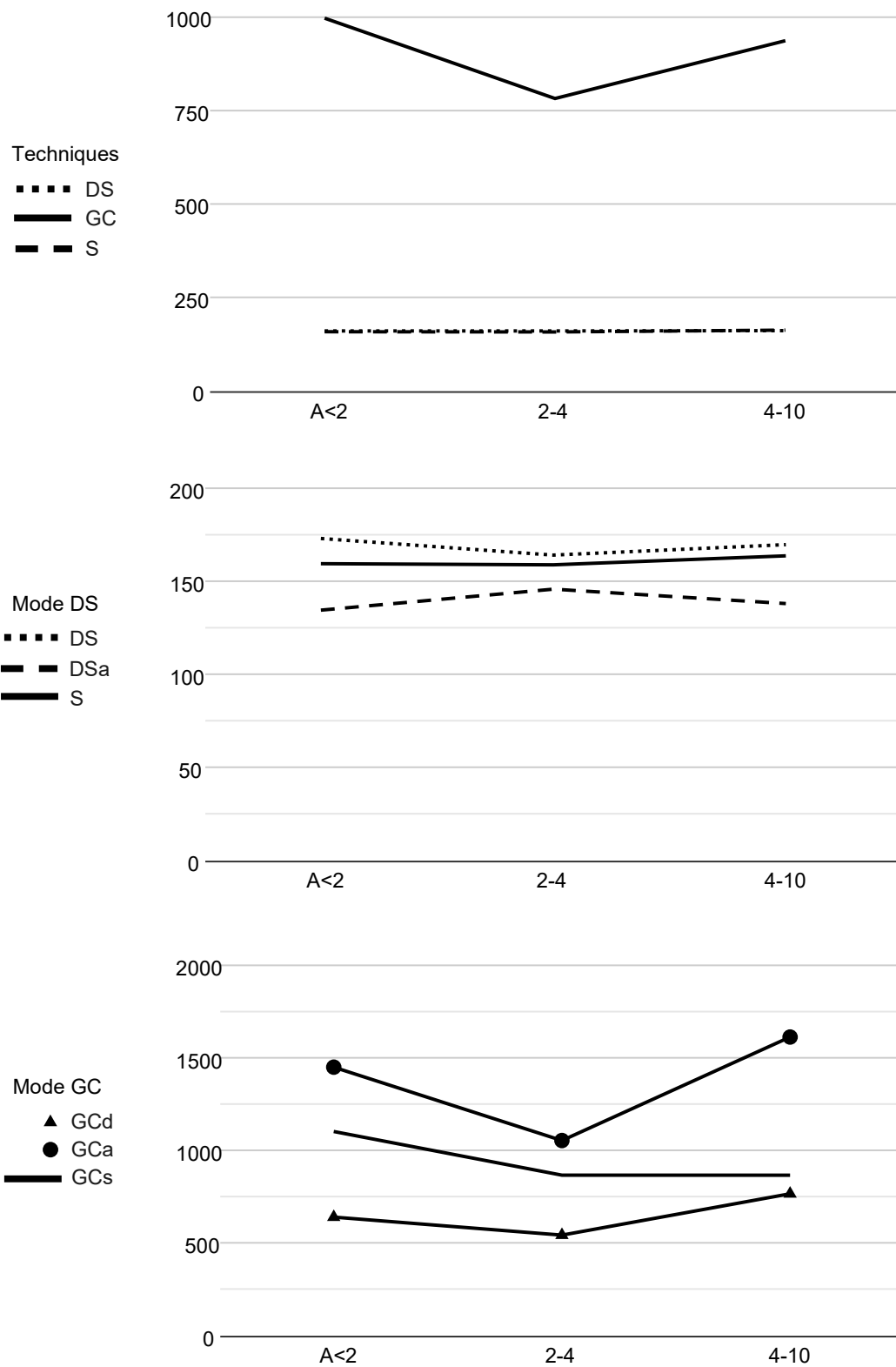


Figure 97 – Score C means by score A stratum. Top: techniques. Center: DS pairs. Bottom: GC pairs.

#### A.4.3.4 Correlations

Table 93 shows mostly weak correlations ( $r < 0.2$ ) between score C and selected variables (Time, Movement, Oscillations) or pre/post tests scores (MRT, SBSOD and TLX). Relationships of score C with Time, Movement, and Oscillations are expected as all variables are related to time spent, and score C is essentially a time-related variable. However, the magnitude found here is too small to draw any conclusions.

Table 93 – HRs: correlations of other variables with score C.

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	0,08	0,22	0,29	-0,1	-	-
DS	0,1	-	0,11	0,1	0,1	-
GC	0,19	0,12	0,11	-0,17	-	-
S	-	-	-	-	0,25	-

##### A.4.3.4.1 Score C summary

Considering all the evidence gathered about score C, particularly the strata analysis, the following order (descending performance) is suggested: DSa > S and DSd > DS > DSs » GCd > GCs > GCa.

#### A.4.4 Score D

##### A.4.4.1 Distributions and outliers

Being a secondary score, this analysis was only performed for good choices or top 10 choices (Score A  $\leq$  10). As D = 1 has a distinct meaning, those scores were also removed from the data sample. The remaining sample size has 749 choices for DS (24 per participant) and 604 for GC (19 per participant), without outliers. Concerning modes, sample sizes are almost equal per technique (Table 94), differing about 25%, which is not a concern for statistical analysis.

Data distributions deviate from normality, and due to the small data set (Figure 98, top), the Aligned Rank Transform ANOVA was used in this case.

Table 94 – HRs: group size and outliers on score D

Technique	N	N (per mode)	Outliers
DS	749	150	0
GC	604	201	0

##### A.4.4.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results were significant for techniques ( $p < 0.01$ ), pairs ( $p < 0.01$ ), and Color in DS ( $p < 0.01$ ).

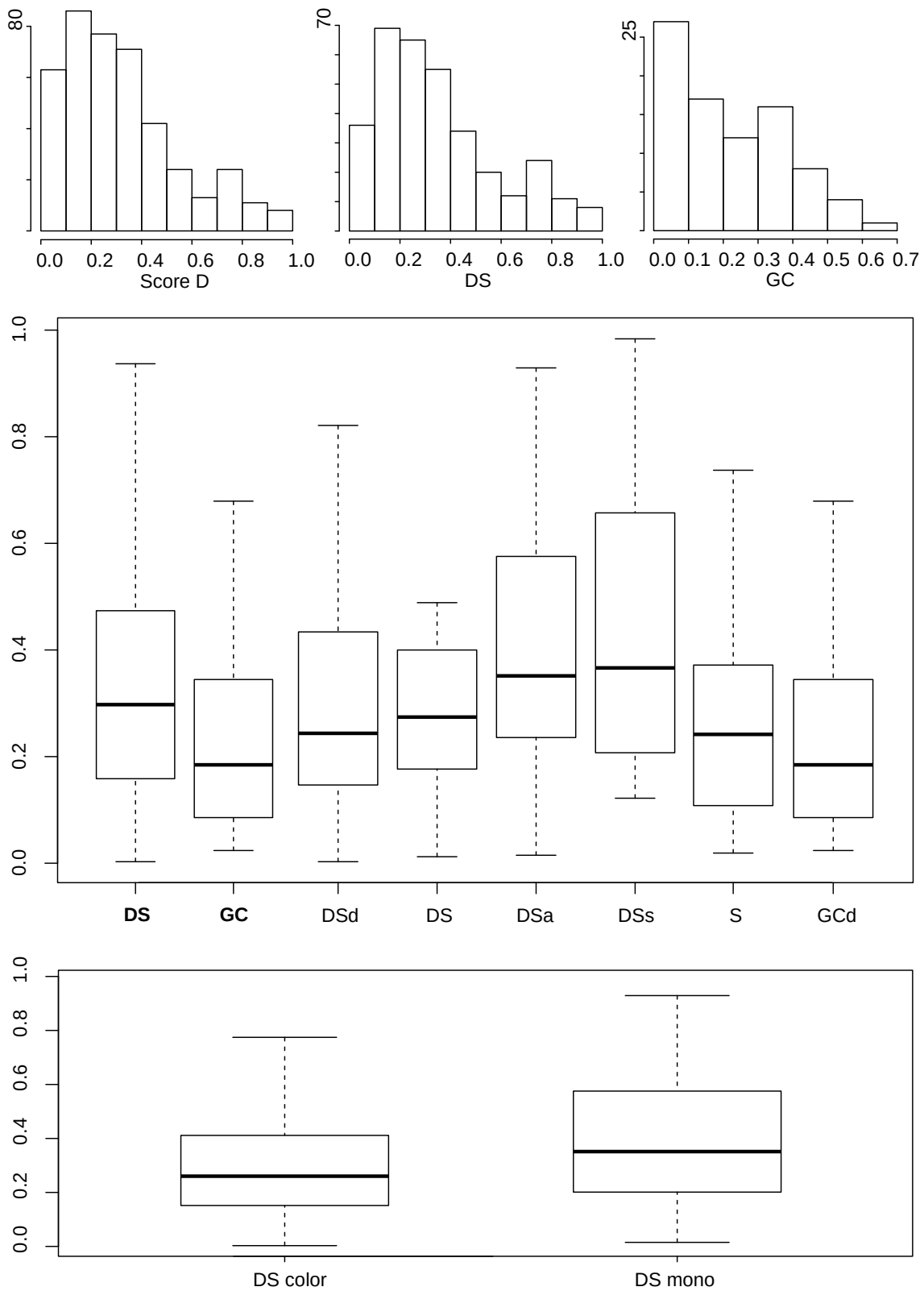


Figure 98 – Score D. Top: histograms are scaled for best view (global, DS group and GC group). Middle: Boxplots group techniques and pairs. Bottom: Color mode.



Boxplots (Figure 98, middle) shows clear overall differences between Techniques and milder differences between pairs. Indeed, pairwise comparisons (Table 97) are only allowed between the lowest scores (S and GCd) and the highest (DSa or DSs), where the difference lies between 40 and 95 %. Techniques differ by 54 % in favor of DS (Table 95). Differences in Color modes in DS is 26 % in favor of Mono (Table 96).

Table 95 – HRs: significant differences (%) per technique on score D

Technique	Mean	s.d.	Difference
DS	0,34	0,23	54,31
GC	0,22	0,15	-35,19

Table 96 – HRs: significant differences (%) per Color mode on score D

Color mode	Mean	s.d.	Difference
DS Color	0,31	0,22	-20,68
DS Mono	0,39	0,24	26,07

Table 97 – HRs: significant differences (%) between pairs, score D

	Mean	s.d.	S	DS	DSd	DSa	DSs	GCd	GCa	GCs
S	0,29	0,24	0	-	-	38,9	52,4	-	NA	NA
DS	0,28	0,16	-	0	-	-	-	-	NA	NA
DSd	0,3	0,2	-	-	0	-	-	-	NA	NA
DSa	0,4	0,23	-28	-	-	0	-	-43,8	NA	NA
DSs	0,44	0,26	-34,4	-	-	-	0	-48,8	NA	NA
GCd	0,22	0,15	-	-	-	78,2	95,5	0	NA	NA
GCa	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
GCs	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

#### A.4.4.3 Strata analysis

The purpose of this analysis is to deepen the distinctions found by statistical analysis by detailing the dynamics of choices with maximum score D, as in this special case,  $D = 1$  means that the selected grain was chosen in its first appearance. This was done by counting the proportion of  $D = 1$  per technique (Figure 99, top), pairs (Figure 99, middle) and per score A strata (Figure 99, bottom).

By examining techniques, GC clearly stands out by more than two times over DS and three over S. Concerning visualization modes, 3D modes clearly emerge within GC. On DS, DSs show a small disadvantage over the others. Strata analysis helps to set DS apart from S, with better performance on the first two strata (Score  $A < 4$ ).

#### A.4.4.4 Correlations

Table 98 shows weak (around 0.2) and moderate correlations (around 0.5) between score D and selected variables (Time, Movement, Oscillations) or pre/post tests scores

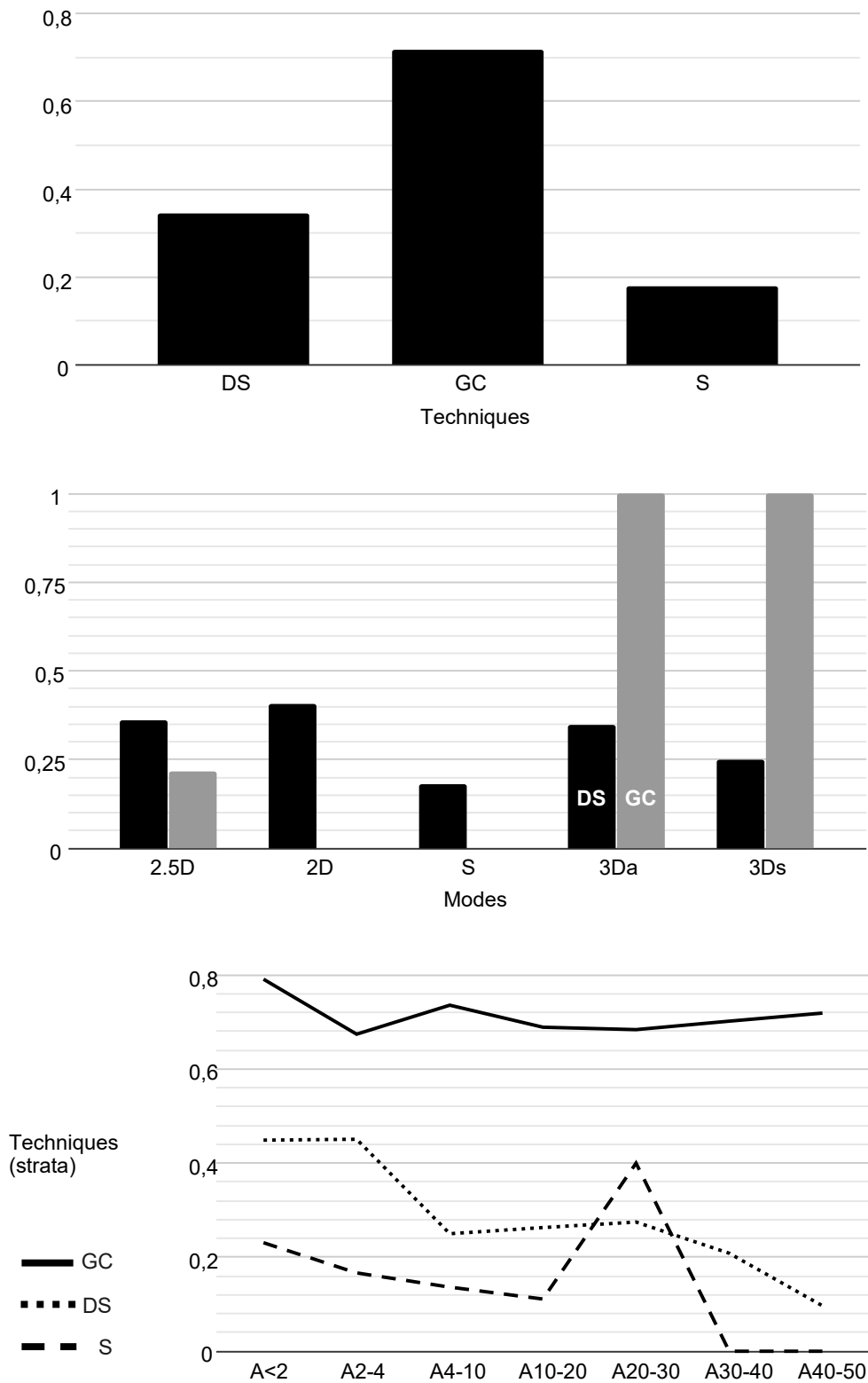


Figure 99 – Score D. Top: counts D = 1 per technique (%). Center: per DS and GC pairs. Bottom: per score A stratum.

(MRT, SBSOD and TLX). It is worth noting that the stronger correlations are all negative and related to the first three variables. This was expected as score D is time-sensitive, thus, as the participant moves or oscillates more, the score is supposed to get smaller. Notably, the table shows that S is more sensitive to movement and oscillations, with a moderate correlation on both. As movement in S occurs in larger jumps, it is easier to escape a given grain, and score D gets smaller every time this happens.

Table 98 – HRs: correlations of other variables with score D.

	Time	Movement	Oscillations	MRT	SBSOD	TLX
Global	-0,14	-0,19	-	-	0,11	-
DS	-0,21	-0,38	-0,36	-	0,18	-
GC	-	-0,32	-0,28	-	-	-
S	-0,36	-0,46	-0,5	-	-	-0,26

#### A.4.4.4.1 Score D summary

Considering all the evidence about score D, particularly strata analysis, the following order (descending performance) is suggested: GCs and GCa > GCd > DSs and DSa > DSd and DS > S. The order DS-GC was inverted after strata analysis as GCs and GCa had only scores D = 1, which fundamentally contributed to GC supremacy over the other techniques in this score.

## A.5 Shape task analysis

Shape analysis was performed with a distinct method than the other two tasks. The main score is a boolean variable, showing if the choice was correct or not. Then, considering correct choices, the time response was further analyzed.

### A.5.1 Main score: correct hits

The GC technique was 103 % better than DS (Figure 100, middle), and both visualization and color modes had a negligible influence on performance, with a maximum of 8 % variation within GC technique. The other responses may help to distinguish the variations further.

### A.5.2 Time response

#### A.5.2.1 Distributions and outliers

The sample size is 798 choices for DS (26 per participant) and 503 for GC (16 per participant), with 178 outliers found (13.6 %). Concerning modes, sample sizes are almost equal per technique (Table 99). Data distributions clearly deviate from normality and

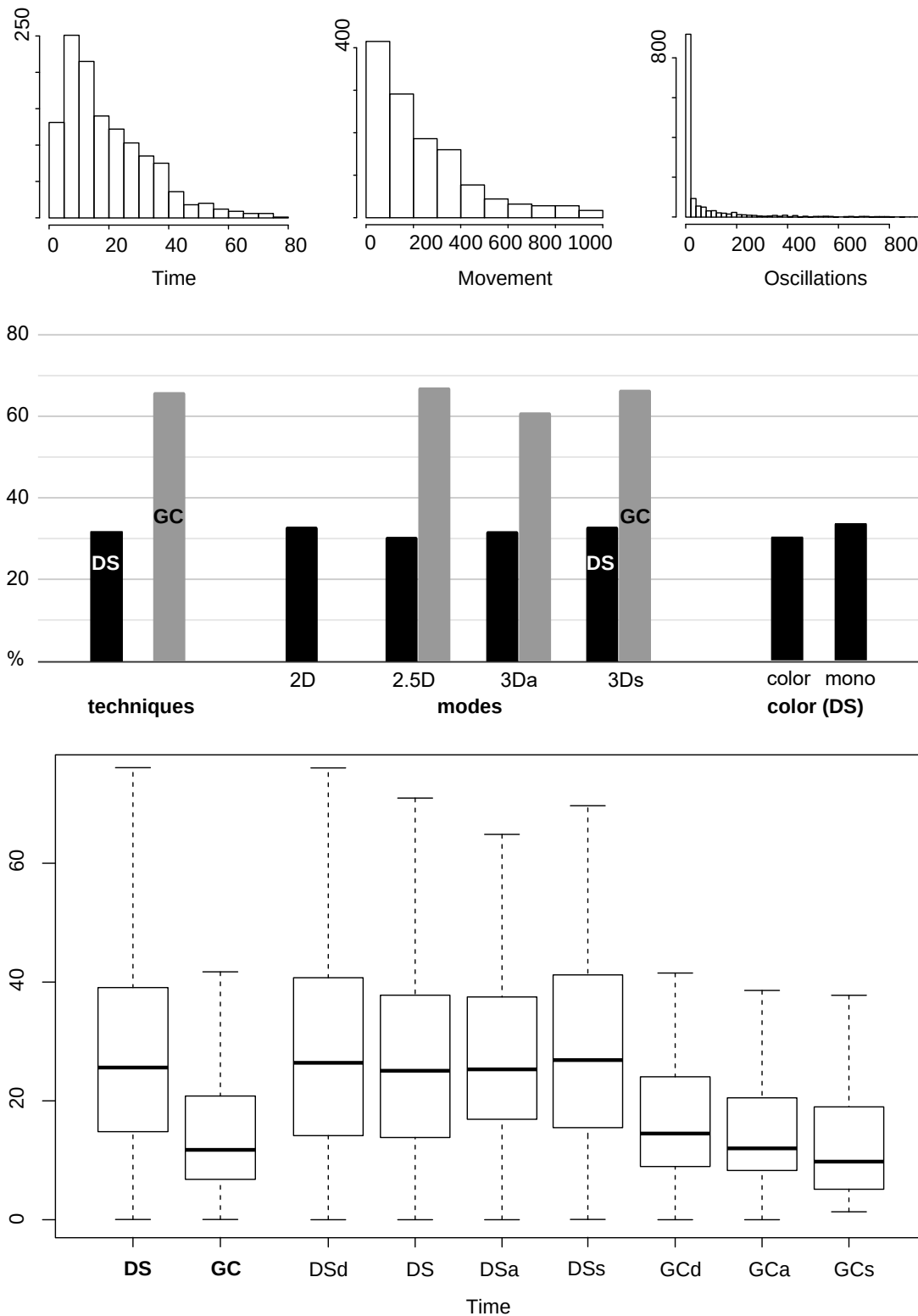


Figure 100 – Shape task. Top: histograms of Time, Movement and Oscillations responses.

appear to follow a Gamma distribution (Figure 100, top left). Therefore this score was analyzed using a GLMM, adopting the Gamma distribution family.

Table 99 – Shape analysis: group size and outliers on time response

Technique	N	N (per mode)	Outliers
DS	798	159,6	107
GC	503	167,6	71

### A.5.2.2 Boxplots, ANOVAs, and comparisons

GLMM ANOVA results were significant for techniques ( $p < 0.01$ ) and pairs ( $p < 0.01$ ). Boxplots (Figure 100, bottom) shows clear differences between Techniques (Table 100) and minor differences within GC modes. Pairwise comparisons (Table 102) show that DS modes cannot be compared, and only GCs and GCd can be compared within GC. The difference between DS and GC pairs lies between 28 and 53 % in favor of GC. Differences in Color modes was not significant and is negligible (Table 101). This analysis reinforced the current findings with GC showing times 44 % shorter than DS, however, GCs can now be distinguished from GCd by 35 %.

Table 100 – Shape analysis: significant differences per technique on time response

Technique	Mean	s.d.	Difference (%)
DS	26,95	17,93	78,63
GC	15,08	11,74	-44,02

Table 101 – Shape analysis: significant differences per Color mode on time response

Color mode	Mean	s.d.	Difference (%)
DS Color	26,85	16,54	-
DS Mono	27,06	19,38	-

Table 102 – Shape analysis: significant differences between pairs on time response

	Mean	s.d.	DS	DSd	DSa	DSs	GCd	GCa	GCs
DS	27,18	16,84	0	-	-	-	-28,1	-42,3	-53,8
DSd	26,76	16,53	-	0	-	-	-27	-41,4	-53,1
DSa	26,88	20,32	-	-	0	-	-27,3	-41,7	-53,3
DSs	27,25	18,35	-	-	-	0	-28,3	-42,5	-53,9
GCd	19,52	13,86	39,2	37	37,7	39,5	0	-	-35,7
GCa	15,66	10,53	73,5	70,8	71,6	73,9	-	0	-
GCs	12,54	10,18	116,7	113,4	114,4	117,2	55,7	-	0

### A.5.2.3 Correlations

Table 103 shows weak (around 0.2) to moderate (0.5) correlations between Time and selected variables (Movement, Oscillations) or pre/post tests scores (MRT, SBSOD and TLX). It is worth noting that the strongest correlations are related to movement and

oscillations variables. This was expected as such variables are time-sensitive, thus, as the participant moves or oscillates more, the Time response gets higher. Considering that GC's average time was considerably shorter than DS, the higher Movement-Time correlation on GC (0.5 x 0.26 in DS) probably means that GC induced more movement on average, which also reflected in the 0.2 correlation to oscillations, as more oscillations are expected with increased movement.

Table 103 – Shape task: correlations of other variables with time response

	Movement	Oscillations	MRT	SBSOD	TLX
DS	0,26	-	-	-	0,15
GC	0,5	0,2	-	0,1	-0,1

#### A.5.2.3.1 Shape task summary

Considering all the evidence gathered about score and Time responses on Shape task, the following order (descending performance) is suggested: GCs > GCd > GCa > DS.

# Annex





## ANNEX A – Ethics committee approval

Ethics approval was requested in order to perform user studies. A detailed project was submitted in the Plataforma Brasil online platform, and it was approved by the EACH (University of Sao Paulo) committee under code CAAE-04436918.8.0000.5390, as attested by the following document (in Portuguese).

The document contains detailed process data (title, main researcher, code, institution, and sponsors), contact data of the ethics committee (complete address, zip code, telephone number, e-mail), a summary of the project (abstract, objectives, assessment of risks and benefits), a checklist of delivered and accepted documents (project abstract, complete project, user consent form, signed statements from researcher and host institution), as well as reviewer's comments, considerations, recommendations, conclusions, and final approval statement.

USP - ESCOLA DE ARTES,  
CIÊNCIAS E HUMANIDADES  
DA UNIVERSIDADE DE SÃO



### PARECER CONSUBSTANCIADO DO CEP

#### DADOS DO PROJETO DE PESQUISA

**Título da Pesquisa:** Identificação e cognição de objetos em visualização imersiva

**Pesquisador:** ANDRE MONTES RODRIGUES

**Área Temática:**

**Versão:** 2

**CAAE:** 04436918.8.0000.5390

**Instituição Proponente:** UNIVERSIDADE DE SAO PAULO

**Patrocinador Principal:** UNIVERSIDADE DE SAO PAULO

#### DADOS DO PARECER

**Número do Parecer:** 3.217.152

#### Apresentação do Projeto:

Pesquisa experimental sobre avaliação de desempenho na identificação de objetos virtuais e na percepção de suas propriedades visuais e geométricas em sistemas interativos imersivos. Será realizada na Escola Politécnica da Universidade de São Paulo (USP), com intenção de avaliar como algumas variáveis da configuração de um sistema interativo imersivo afeta a identificação de objetos virtuais (gerados por computador) e a percepção de tamanho e forma de tais objetos. As principais variáveis avaliadas serão métodos de renderização, visualização e interação. Dois sistemas computacionais serão testados e o desenho experimental será fatorial misto (ou híbrido).

Participarão do estudo 260 alunos graduandos e pós-graduandos com idade entre 18 e 65 anos; profissionais e estudantes de áreas relacionadas ao desenvolvimento do sistema que aceitarem participar na pesquisa, incluindo autorização de divulgação dos dados em publicações científicas, com preservação do anonimato.

#### Objetivo da Pesquisa:

Objetivo Primário:

- Revelar as configurações de visualização e interação que permitem melhor detecção de objetos, seja por seu tamanho, aspecto de forma, volumetria ou outras características sensoriais, em duas condições, uma onde há muito espaço livre entre objetos e outra onde todo o espaço é preenchido

**Endereço:** Av. Arlindo Béttio, nº 1000

**Bairro:** Ermelino Matarazzo

**UF:** SP

**Município:** SAO PAULO

**CEP:** 03.828-000

**Telefone:** (11)3091-1046

**E-mail:** cep-each@usp.br

USP - ESCOLA DE ARTES,  
CIÊNCIAS E HUMANIDADES  
DA UNIVERSIDADE DE SÃO



Continuação do Parecer: 3.217.152

por eles.

**Objetivo Secundário:**

- Avaliar diferentes tecnologias de visualização, tais como monitor 2D, monitor 3D, dispositivos de realidade virtual (HMD - head-mounted display) e outros dispositivos imersivos;
- Avaliar diferentes tecnologias de interação, tais como teclado e mouse, controladores de jogo, interfaces de toque, sensores de rotação e posição, objetos reais rastreáveis no espaço e similares;
- Avaliar a diferença entre visualização mono e estereoscópica;
- Avaliar o impacto de efeitos e recursos sonoros, odores e características táteis, em condição de total congruência ou de incongruência sensorial, além de situações intermediárias;
- Avaliar diversos modos de renderização resultantes da combinação de configurações de cor (natural ou monocromático) e de sombreado (simplificado ou realista), além de outras variações similares envolvendo cores e modelos computacionais de iluminação;
- Avaliar técnicas alternativas para selecionar objetos em ambientes virtuais, tais como cursor virtual no visor, rastreamento das mãos, ponteiros rastreados, seleção tátil ou direta;
- Avaliar diversos modos de renderização (ver exemplos na figura 3), tais como corte sólido (fig. 3, esquerda), corte “vazado” monocromático (fig. 3, centro), corte vazado colorido, corte vazado monocromático com sombreado realista (fig. 3, direita), corte vazado colorido com sombreado realista e outras variações similares envolvendo cores e modelos computacionais;
- Avaliar técnicas de navegação em ambientes virtuais densos e herméticos, tais como visualização por plano de corte dinâmico, visualização por técnicas multiescalares e outras similares que se valham de filtros visuais (ocultamento e agrupamento seletivo de objetos) e redutores de oclusão visual (transparência e uso de objetos alternativos).
- Avaliar qualitativamente fatores de usabilidade e interação humano-computador.

**Avaliação dos Riscos e Benefícios:**

**Riscos:**

Os possíveis riscos que o experimento pode proporcionar são eventuais desconfortos ou mesmo possíveis sensações de náusea ao usar o dispositivo móvel de realidade virtual ou outros tipos de óculos 3D. Nesse caso os participantes serão orientados a interromper imediatamente o experimento. No caso de odores, se utilizados, terão caráter puramente sensorial, sem nenhuma atividade química ou biológica. No caso de objetos físicos o principal aspecto a ser estudado é a textura superficial, o que não apresenta risco

**Endereço:** Av. Arlindo Béttio, nº 1000

**Bairro:** Ermelino Matarazzo

**UF:** SP

**Município:** SAO PAULO

**CEP:** 03.828-000

**Telefone:** (11)3091-1046

**E-mail:** cep-each@usp.br

USP - ESCOLA DE ARTES,  
CIÊNCIAS E HUMANIDADES  
DA UNIVERSIDADE DE SÃO



Continuação do Parecer: 3.217.152

de ferimento.

**Benefícios:**

Apoiar ou refutar as hipóteses da pesquisa; Demonstrar a viabilidade de se utilizar tecnologias de realidade virtual de baixo custo ou mesmo tecnologias convencionais para aplicações educacionais ou comerciais cujo sucesso depende em grande parte da compreensão espacial e identificação de objetos e de suas características geométricas, visuais, táteis e olfativas; Contribuir com a literatura de estudos perceptivos, principalmente nas áreas de computação gráfica e realidade virtual; Aprofundar os conhecimentos gerais sobre todos os processos desenvolvidos, contribuindo para a evolução técnica na área.

**Comentários e Considerações sobre a Pesquisa:**

Estudo com metodologia pertinente aos objetivos proposto e com potencial para contribuir para o avanço do conhecimento sobre processos perceptivos em ambientes virtuais, em especial naqueles caracterizados por alto preenchimento do espaço útil e alta densidade visual.

**Considerações sobre os Termos de apresentação obrigatória:**

Foram apresentados os termos conforme resolução CONEP 466/12: declaração do pesquisador; carta contendo ciência do responsável pela instituição que se dará a coleta dos dados; folha de rosto devidamente preenchida e assinada; PB informações básicas sobre o projeto; projeto na íntegra de maneira adequada.

**Recomendações:**

Sem recomendações.

**Conclusões ou Pendências e Lista de Inadequações:**

Projeto aprovado.

**Considerações Finais a critério do CEP:**

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1253127.pdf	13/02/2019 18:56:34		Aceito
Projeto Detalhado / Brochura Investigador	ProjetoCompleto_final.pdf	13/02/2019 18:46:28	ANDRE MONTES RODRIGUES	Aceito
TCLE / Termos de Assentimento /	TcleFinal.pdf	13/02/2019 13:41:37	ANDRE MONTES RODRIGUES	Aceito

**Endereço:** Av. Arlindo Béttio, nº 1000

**Bairro:** Ermelino Matarazzo

**UF:** SP

**Município:** SAO PAULO

**CEP:** 03.828-000

**Telefone:** (11)3091-1046

**E-mail:** cep-each@usp.br

USP - ESCOLA DE ARTES,  
CIÊNCIAS E HUMANIDADES  
DA UNIVERSIDADE DE SÃO



Continuação do Parecer: 3.217.152

Justificativa de Ausência	TcleFinal.pdf	13/02/2019 13:41:37	ANDRE MONTES RODRIGUES	Aceito
Declaração de Pesquisadores	dec.pdf	05/12/2018 12:33:39	ANDRE MONTES RODRIGUES	Aceito
Declaração de Instituição e Infraestrutura	Declaracao_Infraestrutura.pdf	04/12/2018 19:00:31	ANDRE MONTES RODRIGUES	Aceito
Folha de Rosto	Folha_Rosto_Assinada.pdf	04/12/2018 18:51:26	ANDRE MONTES RODRIGUES	Aceito

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

SAO PAULO, 22 de Março de 2019

---

**Assinado por:**  
**Rosa Yuka Sato Chubaci**  
**(Coordenador(a))**

**Endereço:** Av. Arlindo Béttio, nº 1000  
**Bairro:** Ermelino Matarazzo **CEP:** 03.828-000  
**UF:** SP **Município:** SAO PAULO  
**Telefone:** (11)3091-1046 **E-mail:** cep-each@usp.br



## ANNEX B – Participant consent form

As requested by the ethics committee, a participant consent form was prepared, approved, and filled by all participants before experimental procedures. One copy was given to the participant and another retained by the experimenter. The original document is presented below (in Portuguese).

Following international standards, this document discloses to the participant all necessary information about the experiment and the scientific project. Details are given about the involved researchers, the ethics committee approval, and the research project, focusing on the scientific objectives and expected outcomes.

In the first place, following Brazilian law, the participant is invited as a volunteer. This aspect is further reinforced, informing that the participant is allowed to stop and quit the experiment. The document also states and guarantees the confidentiality and anonymity of the collected data.

Most of the document is dedicated to detailing the experimental procedure, describing the technological setup, the character of the content and pre/post questionnaires, and providing an estimate of total time and experiment location. The document also informs about possible risks, comfort issues, available countermeasures, and how the participant may contact the ethics committee if necessary.

Finally, the participant is asked to provide a name, RG (citizen register number), date, and signature. The experimenter also signs the document. Complete contact information is given on the last page.

## TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Prezado(a) Participante,

Você está sendo convidado(a) para participar voluntariamente de um projeto de pesquisa intitulado **Identificação e cognição de objetos em visualização imersiva**. A pesquisa é realizada pela Escola Politécnica da Universidade de São Paulo (USP), orientada pelo Prof. Dr Romero Tori e coordenada pelo pesquisador doutorando André Montes Rodrigues. O projeto foi aprovado pelo Comitê de Ética em Pesquisa em Seres Humanos da CEPH (CAAE - 04436918.8.0000.5390). **Esta pesquisa tem como objetivo avaliar como algumas variáveis da configuração de um sistema interativo imersivo afeta a identificação de objetos virtuais (gerados por computador) e a percepção de tamanho e forma de tais objetos. As principais variáveis avaliadas serão métodos de renderização, visualização e interação.**

**A pesquisa contribuirá para demonstrar a viabilidade de se utilizar dispositivos e tecnologias de realidade virtual de baixo custo ou mesmo tecnologias convencionais, mais acessíveis ao público geral, em aplicações educacionais ou comerciais cuja eficácia dependa de adequada compreensão espacial e boa identificação de objetos e de suas características.** Para isso, a sua participação é muito importante, e ela se dará da seguinte forma: Durante um experimento, você utilizará um computador convencional, dispositivo móvel (celular ou tablet) ou um dispositivo imersivo (realidade virtual) e lhe será apresentado um ambiente virtual, que pode ser tanto uma sala populada com objetos diversos ou um ambiente caracterizado pelo preenchimento total do espaço por objetos de forma irregular. Nestes ambientes sua tarefa consiste em procurar e identificar determinados objetos com características específicas. É possível que experimente um ambiente virtual puro ou incrementado com alguns objetos reais, em sintonia visual com suas representações virtuais, sendo possível neste caso tocar e manipular tais objetos. Finalmente, também é possível que experimente odores neste ambiente. É importante frisar que tais objetos e odores não são nocivos à saúde nem apresentam riscos de ferimentos.

Antes do experimento, você preencherá um questionário para obtenção de dados de identificação e para sondagem de experiências anteriores relevantes, e outro após o mesmo, sobre sua experiência. O tempo estimado para o preenchimento de todos os formulários é de 25 min e da execução do experimento 35 min, totalizando 60 min. Informamos ainda que há a possibilidade de inclusão em grupo de controle ou experimental.

Informamos que poderá ocorrer algum tipo de desconforto, náusea ou tontura pelo uso das tecnologias de realidade virtual. Você deverá avisar imediatamente à equipe de pesquisa caso note algum desses sintomas, e o experimento será interrompido imediatamente.

Gostaríamos de esclarecer que sua participação é totalmente voluntária, podendo você recusar-se a participar ou desistir a qualquer momento, sem que isto acarrete qualquer ônus ou prejuízo à sua pessoa. Não há custos de nenhuma espécie na participação da pesquisa e nenhum dano está associado à participação.

Informamos que os dados coletados serão utilizados somente para os fins desta pesquisa, e serão tratados com o mais absoluto sigilo e confidencialidade, de modo a preservar a sua identidade. Informamos, ainda, que a sua participação no procedimento será filmada e que as imagens poderão ser divulgadas, se autorizado.

O principal local da pesquisa será no Laboratório de Tecnologias Interativas (Interlab), localizado em Av. Prof. Luciano Gualberto travessa 3, no. 158, sala C2-45, São Paulo, no prédio de Engenharia Elétrica da Escola Politécnica da Universidade de São Paulo (USP). Outras salas do mesmo edifício também poderão ser utilizadas.

Caso você tenha mais dúvidas ou necessite de maiores esclarecimentos, entre em contato nos endereços listados no final desse documento ou procure o Comitê de Ética em Pesquisa da CEPH. Este termo



deverá ser preenchido em duas vias de igual teor, sendo uma delas, devidamente preenchida e assinada, entregue a você.

Além da assinatura nos campos específicos pelo pesquisador e por você, solicitamos que sejam rubricadas todas as folhas deste documento. Isto deve ser feito por ambos (pelo pesquisador e por você, como sujeito ou responsável pelo sujeito de pesquisa) de tal forma a garantir o acesso ao documento completo.

**Agradecemos, desde já, a sua valiosa colaboração.**

Eu,....., portador do documento de identidade,....., declaro que fui devidamente esclarecido e concordo em participar **VOLUNTARIAMENTE** da pesquisa intitulada "Identificação e cognição de objetos em visualização imersiva".

São Paulo, \_\_\_\_\_ de \_\_\_\_\_ de 20\_\_

\_\_\_\_\_  
Assinatura do participante

\_\_\_\_\_  
Assinatura do(a) Pesquisador(a)

Qualquer dúvida com relação à pesquisa poderá ser esclarecida com o pesquisador, conforme o endereço abaixo:

**André Montes Rodrigues**  
Av. Prof. Luciano Gualberto travessa 3, no. 158  
Prédio da Engenharia Elétrica - sala C2-45  
Cidade Universitária  
CEP 05508-010  
São Paulo – SP  
TEL (11) 3091-5282  
E-mail: andre.montes.rodrigues@usp.br

Qualquer dúvida com relação aos aspectos éticos da pesquisa poderá ser esclarecida com o Comitê de Ética em Pesquisa com Seres Humanos do CEPH, no endereço abaixo:

CEP da Escola de Artes, Ciências e Humanidades (EACH) da Universidade de São Paulo (USP)  
Endereço: Av. Arlindo Béttio, nº 1000, bairro Ermelino Matarazzo, CEP 03.828-000, São Paulo/SP  
Tel: (11) 3091-1046  
E-mail: cep-each@usp.br



## ANNEX C – MRT form

The MRT (Mental Rotation Test) form was prepared according to [Vandenberg e Kuse \(1978\)](#) in order to assess participant spatial ability with small objects. The form was approved by the ethics committee and filled one time by all participants before the main experiment. This document is presented below (in Portuguese).

The participant starts by providing name, university ID number, and genre. The first section (four initial pages) presents a tutorial, showing how the test works and what the participant should accomplish. For each line of drawings (five per line), the first drawing indicates the object of interest, and the participant must pick two out of the other four drawings that depict the same object from a different point of view. The remaining two images show a different object. This tutorial provides the opportunity to practice four times, revealing the answers after the first and fourth challenges.

The next two sections are the test itself, presenting 12 objects each, for a total of 24 objects. Each section should be finished in 3 minutes, and a five-minute break is recommended after the first section.

## Teste de Rotações Mentais

\*Obrigatório

Nome \*

Sua resposta

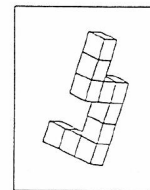
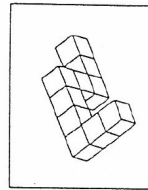
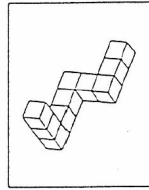
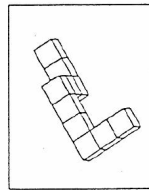
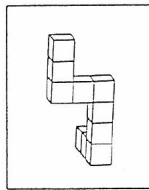
Número USP \*

Sua resposta

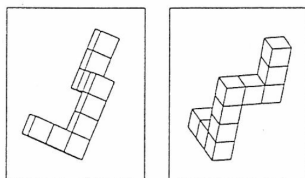
Gênero \*

- Feminino  
 Masculino

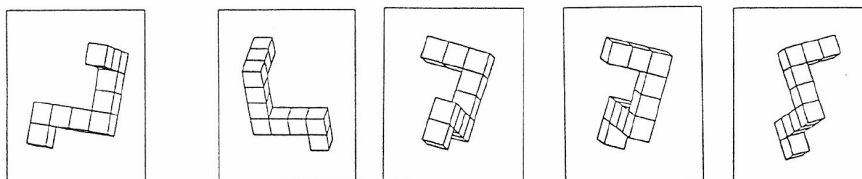
Observe as seguintes figuras:



Note que todas representam o mesmo objeto visto de diferentes ângulos. Tente imaginar o objeto de movimentando (ou você em relação ao objeto) conforme olha de uma imagem à outra.



Acima estão duas imagens de um novo objeto, diferente daquele representado nas primeiras 5 imagens. Perceba que essas duas imagens mostram um objeto diferente e que não pode ser rotacionado de forma a ser idêntico ao objeto mostrado nas 5 primeiras imagens. Agora observe o objeto abaixo na primeira imagem à esquerda. Duas das quatro imagens à direita mostram o mesmo objeto.

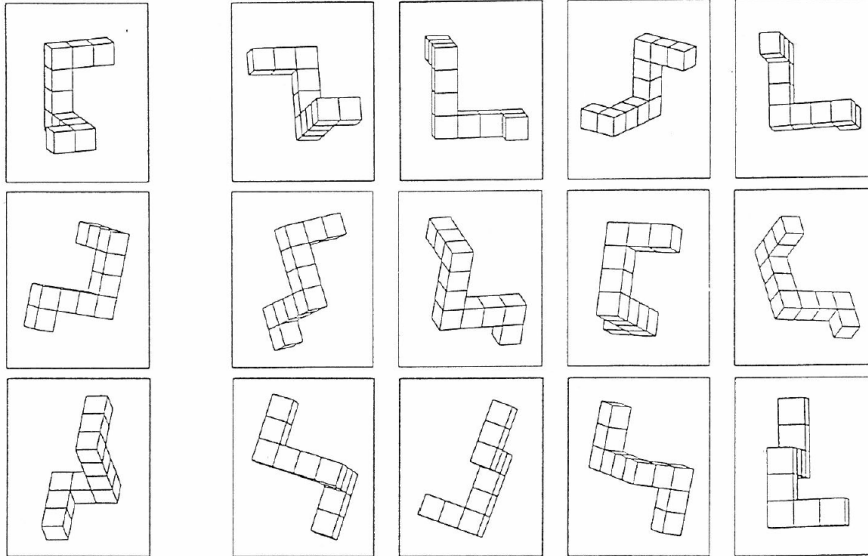


Você consegue identificar essas imagens ? Marque abaixo quais acha que são.

- 1
- 2
- 3
- 4



Se você marcou a primeira e a terceira, fez escolhas corretas.  
Agora, abaixo temos mais 3 problemas similares. Novamente o objeto à esquerda é mostrado em duas das 4 imagens.



Quais são, na linha 1 ?

- 1  
 2  
 3  
 4

E na linha 2 ?

- 1  
 2  
 3  
 4



E na 3 ?

1

2

3

4

As escolhas corretas são segundo e terceiro (linha 1), primeiro e quarto (linha 2) e primeiro e terceiro (linha 3).

Quando fizer o teste, lembre-se que para cada problema existem APENAS duas imagens que representam corretamente o objeto. Qual é sua melhor estratégia para resolver os problemas ? Tome cuidado, pois cada escolha errada subtrai uma correta, portanto escolha apenas as que tiver certeza, mesmo se for apenas uma, mas espera-se que você tentará encontrar as duas.

Na próxima sessão você fará os primeiros 12 problemas, procurando não ultrapassar 3 minutos no total (15 segundos por problema)

Página 1 de 3

**PRÓXIMA**

Nunca envie senhas pelo Formulários Google.

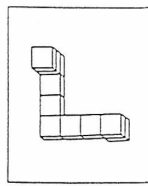
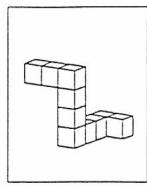
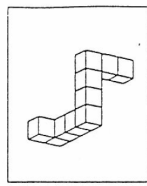
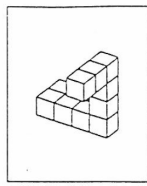
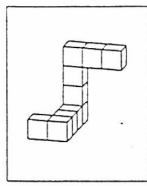
Este formulário foi criado em Universidade de São Paulo. Denunciar abuso - Termos de Serviço - Termos Adicionais

Google Formulários



## Teste de Rotações Mentais

### Seção 1

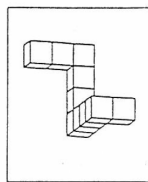
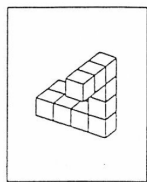
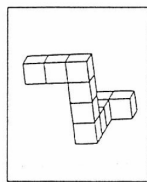
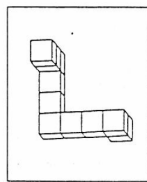
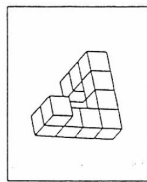


1

2

3

4



1

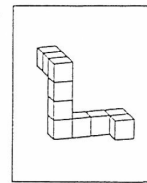
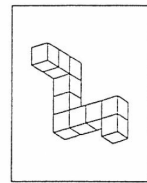
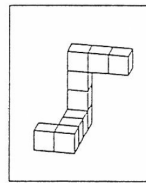
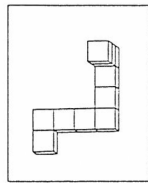
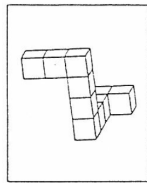
2

3

4





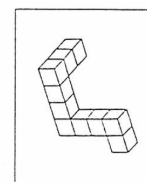
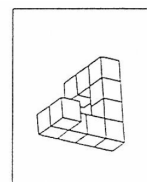
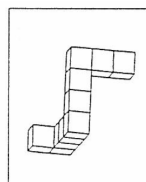
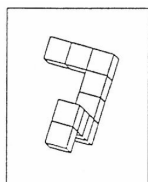
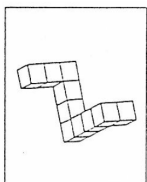


1

2

3

4

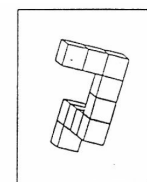
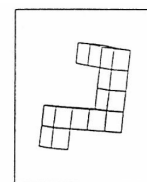
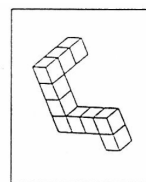
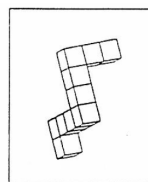
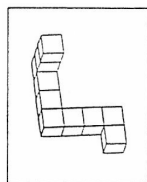


1

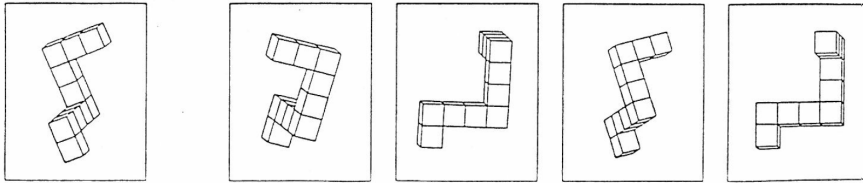
2

3

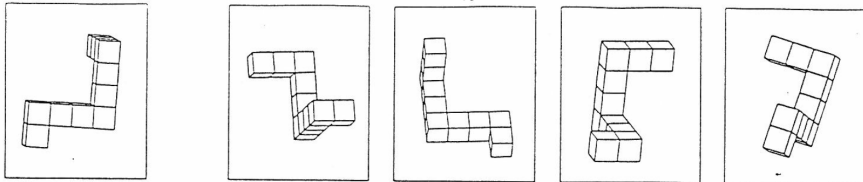
4



- 1
- 2
- 3
- 4

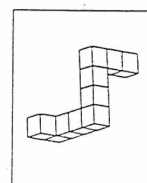
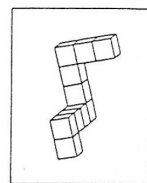
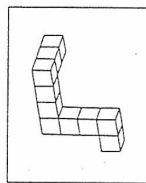
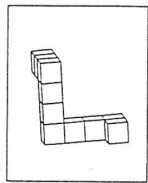
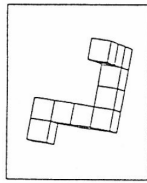


- 1
- 2
- 3
- 4



- 1
- 2
- 3
- 4



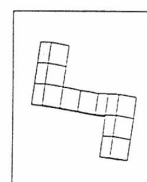
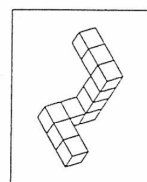
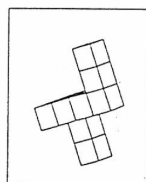
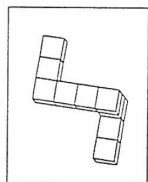
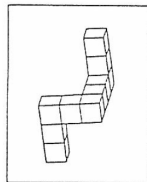


1

2

3

4

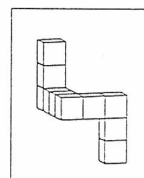
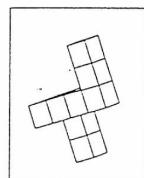
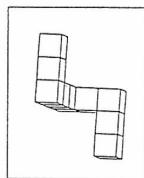
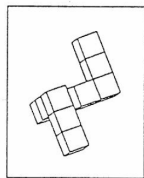
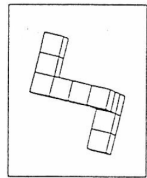


1

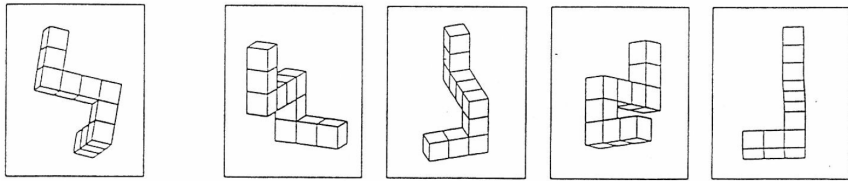
2

3

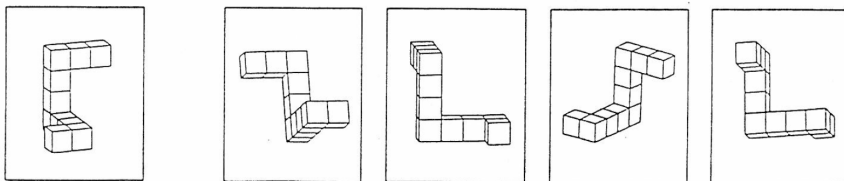
4



- 1
- 2
- 3
- 4



- 1
- 2
- 3
- 4



- 1
- 2
- 3
- 4



Agora faça uma pausa de 4 a 5 minutos antes de prosseguir para os próximos 12 problemas.  
Novamente procure fazer em até 3 minutos.

Página 2 de 3

**VOLTAR**

**PRÓXIMA**

Nunca envie senhas pelo Formulários Google.

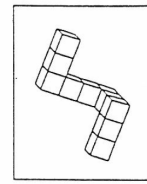
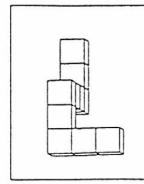
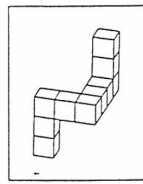
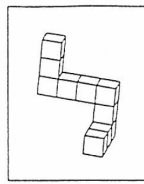
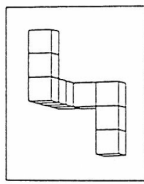
Este formulário foi criado em Universidade de São Paulo. [Denunciar abuso](#) - [Termos de Serviço](#) - [Termos Adicionais](#)

Google Formulários

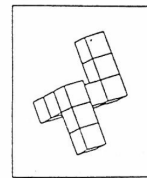
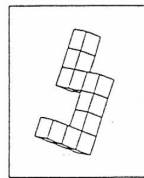
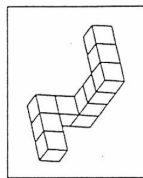
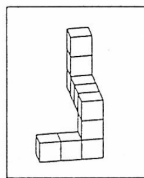
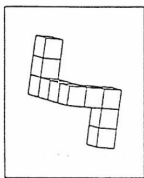


## Teste de Rotações Mentais

### Seção 2

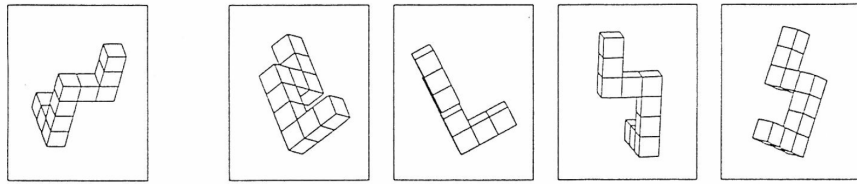


- 1
- 2
- 3
- 4

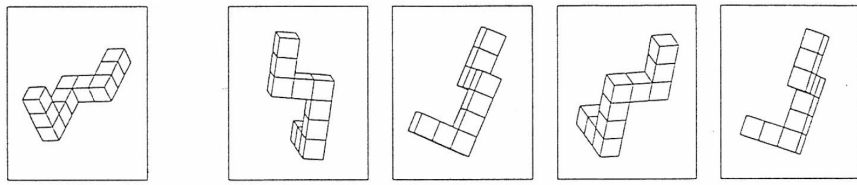


- 1
- 2
- 3
- 4

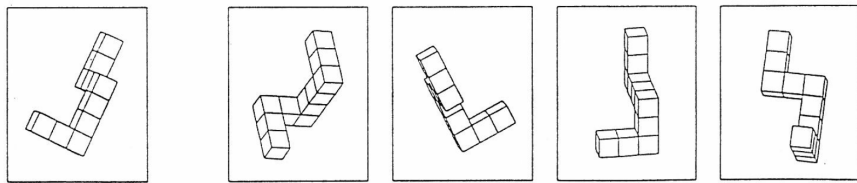




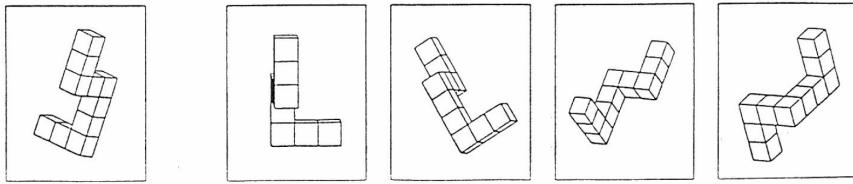
- 1
- 2
- 3
- 4



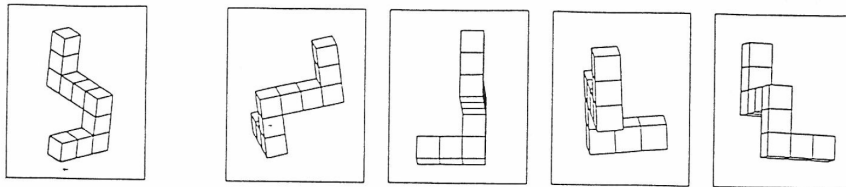
- 1
- 2
- 3
- 4



- 1
- 2
- 3
- 4



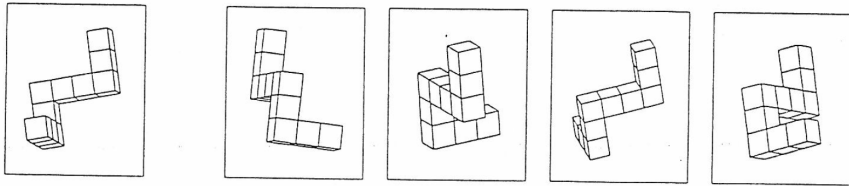
- 1
- 2
- 3
- 4



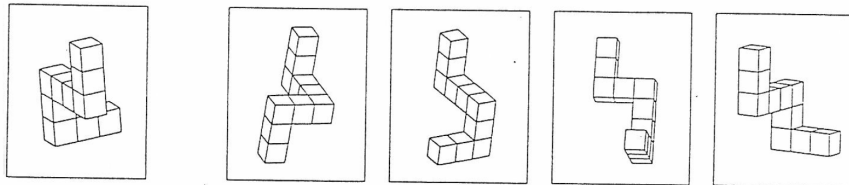
- 1
- 2
- 3
- 4



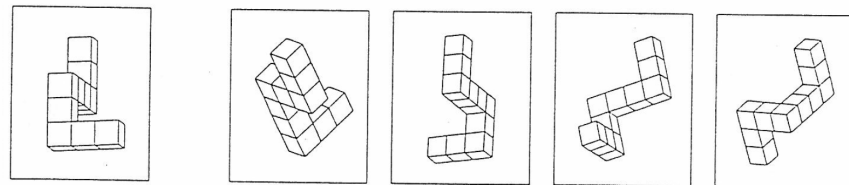




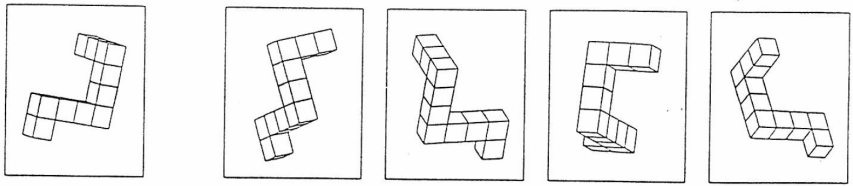
- 1
- 2
- 3
- 4



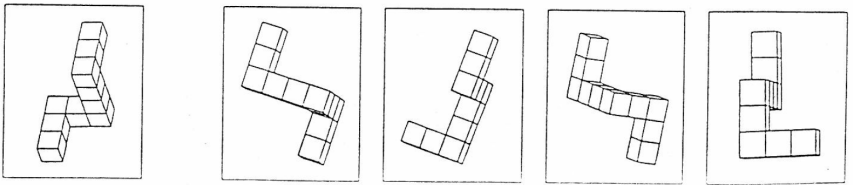
- 1
- 2
- 3
- 4



- 1
- 2
- 3
- 4



- 1
- 2
- 3
- 4



- 1
- 2
- 3
- 4



---

Página 3 de 3

**VOLTAR**

**ENVIAR**

Nunca envie senhas pelo Formulários Google.

Este formulário foi criado em Universidade de São Paulo. [Denunciar abuso](#) - [Termos de Serviço](#) - [Termos Adicionais](#)

Google Formulários





## ANNEX D – SBSOD form

The SBSOD (Santa Barbara Sense of Direction Scale) form was prepared according to [Hegarty et al. \(2002\)](#) to assess participant environmental spatial abilities. The form was approved by the ethics committee and filled one time by all participants before the main experiment. The translated form (Portuguese) is presented on the next pages.

The SBSOD is a simple questionnaire comprising 15 statements about spatial navigation abilities, preferences, and experiences. The participant chooses a number on a 7-level Likert scale, stating how much agrees or disagrees with that statement. Themes are grouped into three main classes: giving and understanding directions, navigating urban environments, and locating oneself and other objects within a larger environment. The main goal is to measure the performance and effort level in key tasks. The higher the ranking, the better the score. The original questions are (R means that score is reversed):

- I am very good at giving directions. (R)
- I have a poor memory for where I left things.
- I am very good at judging distances. (R)
- My ‘sense of direction’ is very good.(R)
- I tend to think of my environment in terms of cardinal directions (N, S, E, W). (R)
- I very easily get lost in a new city.
- I enjoy reading maps. (R)
- I have trouble understanding directions.
- I am very good at reading maps. (R)
- I don’t remember routes very well while riding as a passenger in a car.
- I don’t enjoy giving directions.
- It’s not important to me to know where I am.
- I usually let someone else do the navigational planning for long trips.
- I can usually remember a new route after I have traveled it only once. (R)
- I don’t have a very good ‘mental map’ of my environment.

## Questionário de Senso de Direção

Este questionário consiste de diversas afirmações sobre suas habilidades, preferências e experiências de navegação espacial. Após cada afirmação você deve escolher um número para indicar quanto concorda com ela. Escolha 1 se concorda completamente que a afirmação se aplica a você; 7 se discorda completamente; 4 se não concordar nem discordar e qualquer outro número se sua concordância for intermediária.

Sou muito bom/boa em dar direções.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente

Tenho uma péssima memória para lembrar onde deixei as coisas.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente

Sou muito bom/boa para julgar distâncias.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente

Meu "senso de direção" é muito bom.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente



Tendo a pensar sobre o ambiente que estou em termos de pontos cardeais (Norte, Sul, Leste, Oeste).

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente

Me perco facilmente em uma nova cidade.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente

Gosto de ler mapas.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente

Tenho dificuldade em entender direções.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente

Sou muito bom/boa em ler mapas.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente

Não me lembro muito bem de rotas quando estou de passageiro em um veículo.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente



Não gosto de dar direções.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente

Não é muito importante para mim saber onde estou.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente

Normalmente deixo outra pessoa fazer o planejamento de navegação para longas viagens.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente

Normalmente consigo me lembrar de uma nova rota após passar por ela apenas uma vez.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente

Não tenho um bom "mapa mental" de meu ambiente.

	1	2	3	4	5	6	7	
concordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	discordo totalmente

VOLTAR

ENVIAR

Nunca envie senhas pelo Formulários Google.





## ANNEX E – TLX form

The Nasa TLX (Task Load Index) form was prepared according to [Hart e Staveland \(1988\)](#) in order to estimate participant workload. The form was approved by the ethics committee and filled by all participants after each task of the main experiment. The original document is presented on the next pages (in Portuguese).

The original TLX has two sections. The first section comprises six questions related to a subjective subscale: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. The second part of the TLX allows the participant to weight the importance of each subscale to that task, however, it is usually skipped in most HCI studies. The full version of each question is presented below:

- Mental Demand: How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?
- Physical Demand: How much physical activity was required? Was the task easy or demanding, slack or strenuous?
- Temporal Demand: How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?
- Overall Performance: How successful were you in performing the task? How satisfied were you with your performance?
- Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?
- Frustration Level: How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?

The adopted questionnaire is a Portuguese translation of the paper-and-pencil version available at NASA's website, featuring shorter questions. For the sake of simplicity, the original Likert scale was reduced from 21 to 11 levels.

## Questionário de carga cognitiva (NASA TLX)

Esse questionário visa verificar o nível de esforço exigido pelo experimento e o nível de estresse decorrente. Leia atentamente a descrição de cada pergunta e responda francamente.

\*Obrigatório

Nome \*

Sua resposta

Data \*

Data

dd/mm/aaaa

Demanda mental - qual foi o nível de exigência mental da tarefa?

\*

0 1 2 3 4 5 6 7 8 9 10

Muito baixo             Muito alto

Demanda física - qual foi o nível de exigência física da tarefa ? \*

0 1 2 3 4 5 6 7 8 9 10

Muito baixo            Muito alto



**Demanda de tempo - quão apressado foi o ritmo da tarefa ? \***

0 1 2 3 4 5 6 7 8 9 10

Muito baixo             Muito alto

**Desempenho - quão bem sucedido você foi em cumprir o que foi solicitado ? \***

0 1 2 3 4 5 6 7 8 9 10

Perfeito             Falhei

**Esforço - quão pesado foi o trabalho para atingir o seu nível de desempenho ? \***

0 1 2 3 4 5 6 7 8 9 10

Muito baixo             Muito alto

**Frustração - quão inseguro, desanimado, irritado, estressado ou incomodado você se sentiu ? \***

0 1 2 3 4 5 6 7 8 9 10

Nada             Muito

ENVIAR

Nunca envie senhas pelo Formulários Google.

Este formulário foi criado em Universidade de São Paulo. [Denunciar abuso](#) - [Termos de Serviço](#) - [Termos Adicionais](#)

Google Formulários

