User exploration behavior for distorted Point Clouds

Author 1[1], Author 2[1], Author 3[2], Author 4[1]
[1]*Affiliation 1*
[2]*Affiliation 2*
Email: Authors's emails

***Abstract*—Nowadays Virtual Reality applications are adopted in a number of different scenarios, from gaming to training and teleconferencing. In this context, understanding the way the users navigate and interact in the virtual world is a key aspect of the development of effective and usable applications. In this paper, we study the movement of the user in the visualization of 3D content in a virtual environment. More specifically, we designed and run subjective experiments in which the users were asked to rate the quality of distorted point clouds that were visualized in a Head Mounted Display. One peculiarity of the experiments is the fact that users were allowed to move freely in the space around the object without any time constraints. F: qui poi va messa una riga sui nostri findings**

***Index Terms*—Virtual reality, Point Cloud, Quality Assessment, Subjective Experiment**

I. Introduction

In recent years we are witnessing a growing number of applications and services that take advantage of virtual reality (VR). In these systems, the user is immersed in a computer-generated environment in which he or she can use or interact with digital content. Currently, a VR immersive systems are mostly based on visual and audio content even if systems that can integrate olfactory and haptic dimensions are under study [1]–[3].

The evaluation of a user's quality of experience is a critical element in the design of VR systems and, to some extent, also in the development or failure of the system itself. The design of objective metrics for assessing the Quality of Experience (QoE) requires a deep knowledge of the human factors that directly affect quality perception. While this aspect has been studied for years for audio, 2D images, and videos, the research is still actively investigating these aspects in VR.

The purpose of this paper is to understand how people cope with an experience in VR. One of the main features of VR is that the user can move in the three-dimensional space. This means that, unlike traditional media enjoyment on 2D screens, in VR the user can move within a larger or smaller area depending on the application. That is, the user can act as if in a real situation. While this possibility greatly enhances the interactivity and the immersivity of the user, the freedom in the movement makes every user experience unique thus posing a great challenge in the definition of a general framework for the assessment of the quality of experience. In fact, one of the main requirements defined by the standards for quality evaluation [4], [5], is to guarantee common conditions that allow the reproduction of the tests and to compare the subjective scores provided by the users. When it comes to VR subjective tests, if users are allowed to freely move in the virtual space, it is not possible to define standard viewing conditions. For this reason, this works aims at investigating the impact of the user's movements on his/her quality evaluation.

In previous work, which is available for reviewers upon request (currently submitted for publication), a database of Point Cloud (PC) was created and a series of subjective experiments were performed to collect Mean Opinion Score (MOS). In this work, our focus is the in-depth analysis of people's behavior in virtual reality. In more detail, we assess whether there is a relationship among people's movement, artifact detection ability, and subjective judgment.

II. Dataset

The dataset used in this work has been designed for assessing the quality of PCs. It includes 4 pristine PCs: *Nefertiti*, selected from the PointXR dataset [6], *Longdress*, *Loot*, *Redandblack*, and *Soldier* from the JPEG Pleno dataset [7]. Twenty-four distorted PCs were generated by applying additive Gaussian white noise, random subsampling (i.e., subsampling of points performed randomly on the entire PC) and grid subsampling (i.e., subsampling of points performed on the grid superimposed on the point cloud) to the original PC. Distortions were applied to either (i) the entire PC (hereafter referred to as All (A)) or (ii) the edges of the pristine PC (hereafter referred to as Edge (E)). For each distortion, 4 levels have been selected. The parameters adopted to generate the distorted PC are reported in Tables I and II. As can be noticed, the parameters used are not exactly the same in the two applications area. They were selected empirically to return a PC in which distortions were not visible, barely noticeable, visible, and annoying. Expert evaluators selected parameter values.

Point cloud	Noise (dB)	Downsample Random (%)	Downsample Grid (10^{-4})
Longdress	[50, 60, 70, 80]	[20, 40, 60, 80]	[1, 2, 3, 4]
Loot	[50, 60, 70, 80]	[20, 40, 60, 80]	[1, 2, 3, 4]
Redandblack	[50, 60, 70, 80]	[20, 40, 60, 80]	[1, 2, 3, 4]
Soldier	[50, 60, 70, 80]	[20, 40, 60, 80]	[1, 2, 3, 4]
Nefertiti	[50, 60, 70, 80]	[20, 40, 60, 80]	[2, 3, 4, 5]

TABLE I
Selected distortion levels for the condition *All*.

III. Clustering approach

A. Motion histogram

Our goal is to cluster experiments that show a similar motion pattern to obtain a deeper understanding of user

Point cloud	Noise (dB)	Downsample Random (%)	Downsample Grid(10^{-4})
Longdress	$[55, 60, 65, 70]$	$[10, 30, 50, 70]$	$[1, 2, 3, 4]$
Loot	$[55, 60, 65, 70]$	$[10, 30, 50, 70]$	$[1, 2, 3, 4]$
Redandblack	$[55, 60, 65, 70]$	$[10, 30, 50, 70]$	$[1, 2, 3, 4]$
Soldier	$[55, 60, 65, 70]$	$[10, 30, 50, 70]$	$[1, 2, 3, 4]$
Nefertiti	$[50, 60, 70, 80]$	$[20, 40, 60, 80]$	$[2, 3, 4, 5]$

TABLE II
SELECTED DISTORTION LEVELS FOR THE CONDITION *Edge*.

behavior in VR. For each experiment, we build a *motion pattern histogram* that uniquely identifies the user motion during the experiment. The histogram, shown in the 2D example of Figure 1, carries in its values the time spent by a specific user in a specific position of the virtual space. The headset coordinates, quantized by a 0.25 bin step, are taken as reference. As the average time spent in a given bin of space increases, the color tends to be yellow. The label **O** indicates the original PC (position $x = -2$, $z = 0$), while **D** indicates the distorted PC (position $x = 2$, $z = 0$) the disturbance being a random noise (level 3) applied to the entire PC. The user vote is 4. From Fig. 1 we observe that the user, initially standing in position $(x, z) = (1, 2)$, watches the original PC **O** from a distance and more closely observes the distorted one **D** by walking around it. The evaluation takes 50 s overall.

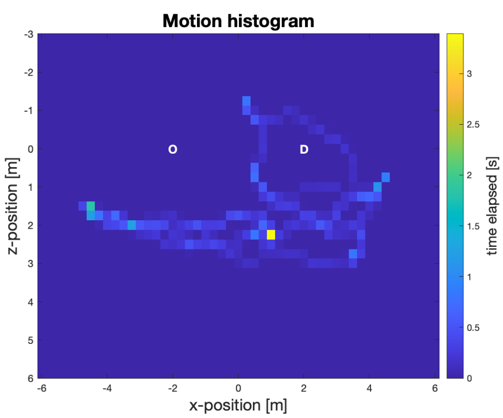

Fig. 1. Motion histogram example - user 1, experiment 67 - two point clouds

Although the example of Fig. 1 shows a 2D representation, where only the available horizontal coordinates are used, patterns are also created for 3D motion patterns by further exploiting the vertical direction y in a fully compliant manner.

B. Correlating patterns

Interdependencies among motion histograms can be evidenced by a simple (normalized) *correlation* measure that captures the *similarity* among histogram h_1 and histogram h_2 by the value $C_{1,2} = \langle h_1, h_2 \rangle / \|h_1\| \, \|h_2\|$, where $\langle \cdot, \cdot \rangle$ identifies an inner product and $\| \cdot \|$ a norm (in either the 2D or the 3D domain). For the sake of robustness against small position offsets, a Gaussian filter ($\sigma = 1.85$) is applied to histograms prior to correlating them.

The resulting correlation matrix C can then be exploited as a reliable tool to investigate interdependencies in motion patterns. Specifically, C can be used as an adjacency matrix from which to build a graphical representation by means of network science tools. This is graphically illustrated in the example Fig. 2 for the experiments of user 6 (green nodes) and user 12 (yellow nodes). The graphical layout exploits the correlation measure $C_{i,j}$ as the strength (weight) of the edge linking the motion pattern h_i to the motion pattern h_j. The specific layout was obtained by a force-directed algorithm using the Fruchterman-Reingold approach [8], which is an efficient form of dimensionality reduction displaying nodes in the 2D space in such a way that strongly correlated patterns lie close to one another, while loosely correlated patterns tend to be placed far away. As one can infer from the figure, this approach is able to visualize the presence of clusters of data, e.g., the different behavior in motion of users 6 and 12 due to the fact that the two classes of nodes clearly separate into two disjoint groups.

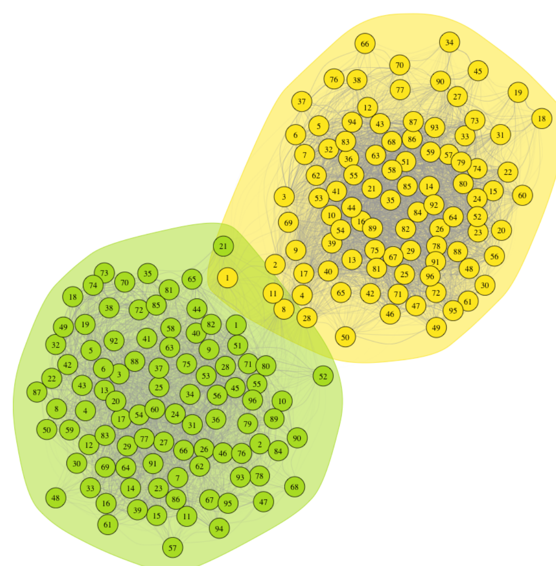

Fig. 2. Graph layout (Fruchterman-Reingold force-directed approach) of the interdependence between the 96 experiments of user 6 (green) and user 12 (yellow).

C. Clustering approach

Clustering is performed by applying the Louvain algorithm [9] to the connectivity matrix C, an *unsupervised* algorithm having a close relationship with the visual structures that emerge from the force-directed layout of the Fruchterman-Reingold approach [10]. It is an extremely fast and easy-to-implement community detection algorithm that can produce high-quality and accurate results. In order to mitigate the greedy nature of the Louvain algorithm, and strengthen its output, the *consensus clustering* technique of [11] was applied to 100 independent community detection outcomes, in order to examine the consistency of each node's module affiliation across partitions and to statistically extract an *average* community assignment that takes into consideration of how many times, in the 100 runs, two nodes appear in the same community. The selected optimal number of runs was 100, and the corresponding threshold was set to 0.3.

IV. Results

The classification is available in Figure 3 both for motion data exploiting 2D coordinates x (horizontal) and z (depth), as well as for a full 3D mapping taking into account the vertical coordinate y. Average motion histograms per cluster are shown.

In both the 2D and 3D visualization cases we observe three main clusters, namely:

1) Cluster 1 refers to subjects observing the two PCs from a distance ($\sim 4\,\mathrm{m}$ in average) and moving from left to right in their inspection, and occasionally moving closer and behind;
2) Cluster 2 refers to subjects watching more closely the two PCs ($\sim 2\,\mathrm{m}$ average distance), and much more frequently moving closer and behind;
3) Cluster 3 refers to subjects mainly standing still in their observations, at an average distance of $\sim 3\,\mathrm{m}$.
4) Clusters 4 and 5 in the 3D map, instead, correspond to a subject watching the PCs from a lower vertical position (Cluster 4, subject 3, possibly kneeling) or standing still but from a higher distance of $\sim 5\,\mathrm{m}$ (Cluster 5, subject 10).

As one can infer from the confusion matrix of Fig. 4 (below), there is a neat correspondence between clusters in the 2D and 3D domains. Fig. 3 also reports the cluster distribution among the 22 users, clearly revealing that most users adopt a single motion strategy in the virtual world.

Fig. 4 further shows the voting behavior under different noise/distortion effects, for the main 3D clusters. Values in the plots represent the *average value* and the *slope* of a linear fitting linking the noise/distortion level to the users' vote, where the slope is expected to be negative (~ -1) under an ideal user's evaluation. We observe that:

DR *Random subsampling*: we observe a different behavior when the distortion is applied to the whole PC (A) and when it is applied to PC edges (E), namely:

- **E** the presence of positive/zero slopes reveals that the level of distortion is, independently of the specific motion pattern and selected PC, hardly recognizable by the user, with high distortion levels seldom perceived as more acceptable than lower ones; **T: explain the reason**
- **A** the negative slopes ensure that there is a neat correlation between the level of distortion and the user experience, particularly enhanced under a detailed inspection of the virtual environment (cluster 2, walking around and closely) where slopes assume a lower value. Overall the user experience is less positive than in DR-A, as we can infer from the average values.

DG *Grid subsampling*: the comparable average vote and slope levels ensure a similar user experience for the whole (A) and edge (E) distortion setups. With respect to the motion patterns, instead, some distinctions arise, namely:

- **A** in this case the (negative) linear relation between distortion level and vote shows a more pronounced effect in both clusters 2 and 3 (walking around and closely, and standing still, respectively);
- **E** we observe an enhanced linear behavior (much lower slopes) only in cluster 2 (walking around and closely).

NS *White Gaussian noise*: white Gaussian noise stands out as the most acceptable disturbance, with higher average vote values in both whole (A) and edge (E) distortion setups; we also have:

- **A** particularly low slopes identify a strong linear effect clearly linking the level of distortion to the vote, a behavior that is particularly well captured in cluster 2 (walking around and closely). In this cluster, the user walked around the object and at close range, showing behavior similar to what one would have in a real situation.
- **E** average vote values tend to saturate to the maximum (hence slopes tend to be much less pronounced), especially in cluster 2 (walking around and closely), with a slight degradation of the user experience generally associated with cluster 1 (watching from a distance). As can be noticed, in this case, there is little or no movement of the subject. This could be due to either unfamiliarity with the virtual world or poor involvement in the experiment.

V. Conclusions

In this contribution, we analyze the user behavior while using an application in VR. To this end, we apply a clustering approach to data collected through a subjective quality assessment experiment, in which two PCs (original and its distorted version) were shown in a virtual environment. The results

Acknowledgment

This work was partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE0000001 - program "RESTART").

References

[1] J. Mendes Miranda Martins and M. de Paiva Guimarães, "Using olfactory stimuli in virtual reality applications," in *2018 20th Symposium on Virtual and Augmented Reality (SVR)*, 2018, pp. 57–64.

[2] A. Sheremetieva, I. Romanv, S. Frish, M. Maksymenko, and O. Georgiou, "Touch the story: An immersive mid-air haptic experience," in *2022 International Conference on Interactive Media, Smart Systems and Emerging Technologies (IMET)*, 2022, pp. 1–3.

[3] A. Zenner, K. Ullmann, and A. Krüger, "Combining dynamic passive haptics and haptic retargeting for enhanced haptic feedback in virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2627–2637, 2021.

[4] ITU-R BT.500-13, “Methodology for the subjective assessment of the quality of television pictures,” *International Telecommunications Union*, 2012.

[5] ITU P.913, “Methods for the subjective assessment of video quality, audio quality and audiovisual quality of internet video and distribution quality television in any environment,” *International Telecommunications Union*, 2021.

[6] S. Subramanyam, I. Viola, J. Jansen, E. Alexiou, A. Hanjalic, and P. Cesar, “Subjective QoE Evaluation of User-Centered Adaptive Streaming of Dynamic Point Clouds,” in *14th International Conference on Quality of Multimedia Experience (QoMEX)*, 2022.

[7] E. d’Eon, B. Harrison, T. Myers, and P. Chou, “8i Voxelized Full Bodies - A Voxelized Point Cloud Dataset,” *ISO/IEC JTC1/SC29 JointWG11/WG1 (MPEG/JPEG)*, 2017.

[8] T. M. Fruchterman and E. M. Reingold, “Graph drawing by force-directed placement,” *Software: Practice and experience*, vol. 21, no. 11, pp. 1129–1164, 1991.

[9] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[10] A. Noack, “Modularity clustering is force-directed layout,” *Physical review E*, vol. 79, no. 2, p. 026102, 2009.

[11] A. Lancichinetti and S. Fortunato, “Consensus clustering in complex networks,” *Scientific reports*, vol. 2, no. 1, pp. 1–7, 2012.

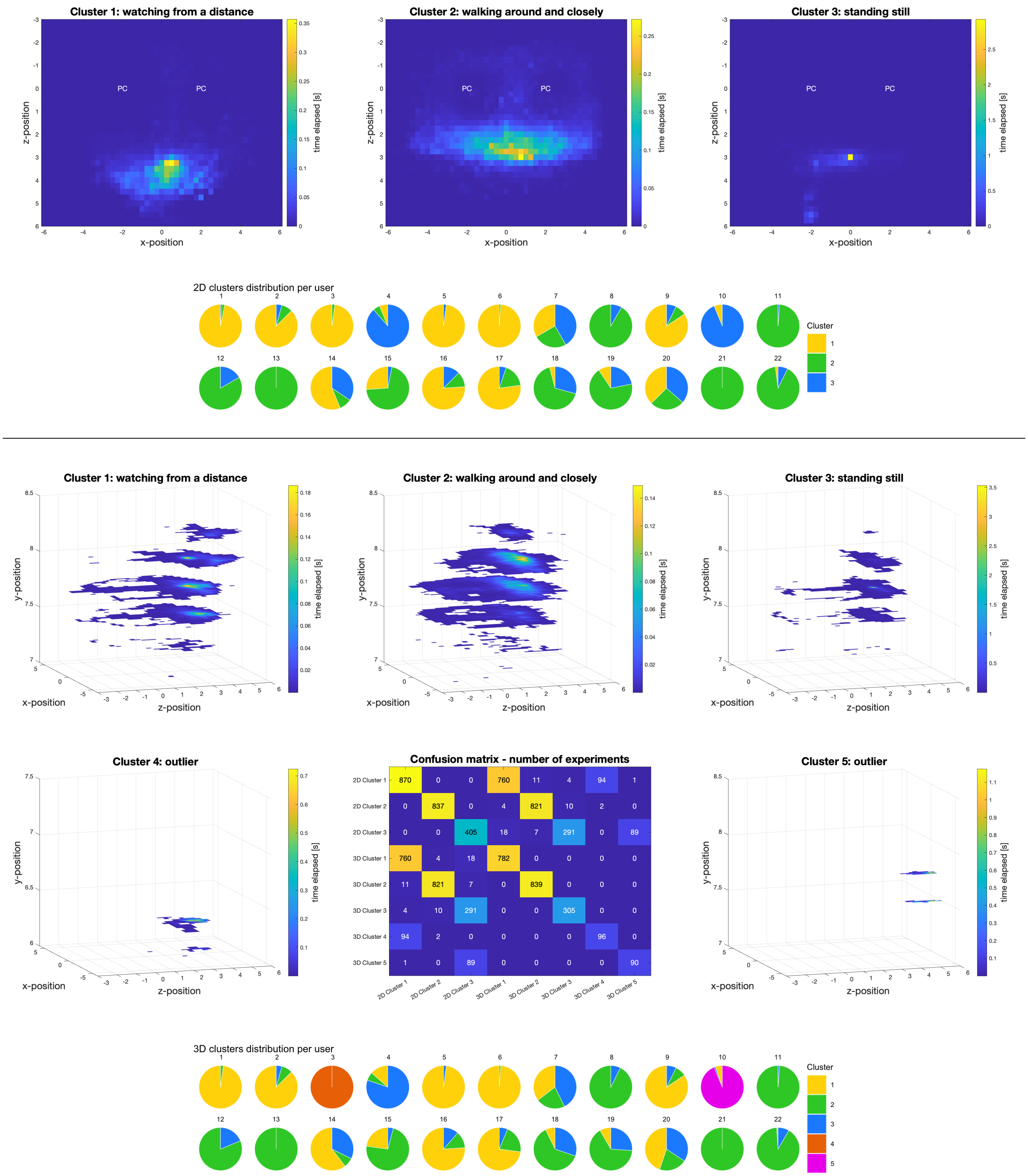

Fig. 3. Average motion histograms and cluster distribution among users

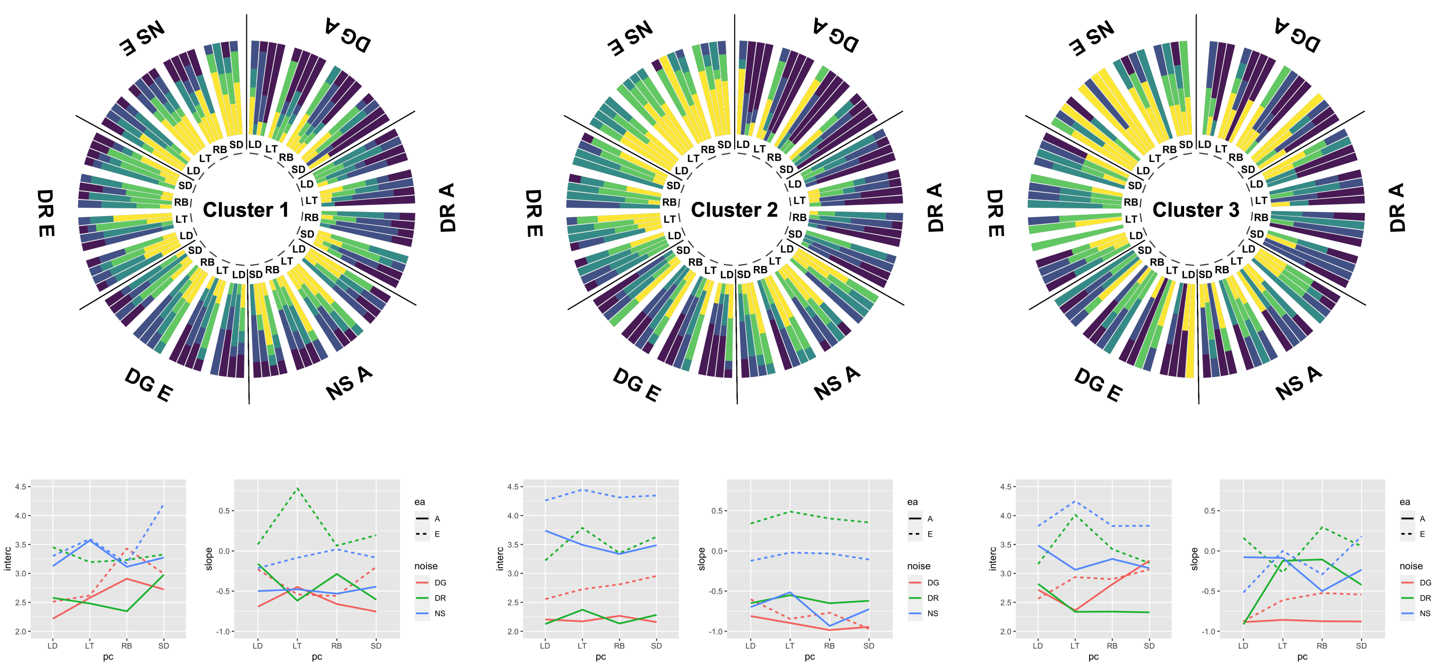

Fig. 4. 3D clusters - Voting behavior under different noise/distortion effects - two PCs