Supplementary Material
Design and evaluation of a continuous annotation interface for real-time self-reporting of VR sickness

This PDF contains additional information on the following topics:

- (S.1) Training scene.
- (S.2) Speed modification of the VR sickness-inducing scene.
- (S.3) Individual time-series of VR sickness over time.
- (S.4) Qualitative feedback from semi-structured interviews.

Author's Contact Information:

S.1 Training scene

This section presents representative screenshots of the training scene. This scene is presented to the participants during the first five minutes of the second phase of the study, after they complete the initial questionnaire phase, sit in a non-rotating chair, and put on the HMD and controller.

We design the training scene to help participants familiarize themselves with VR and learn how to interact with the interfaces. As the study employs a within-subjects design with three sessions on separate days (one per interface, presented in a randomized order to mitigate potential memory issues), the training scene displays the specific interface that the participant will evaluate that day.

The training environment consists of a simple, static living-room scene. To maintain consistency with the interface designs, instructional elements are head-locked within the ergonomic field of view and positioned in a centered location for constant visibility; text instructions use the same typeface as the interfaces; and button interactions follow the same unified logic, including consistent highlighting colors on hover and identical trigger-based activation. After this scene, the actual test begins.

Fig. 1. Representative screenshots of the training scene.

S.2 Speed modification of the VR sickness-inducing scene

This section describes the speed modification applied to the VR sickness-inducing scene used in the study. During the study, participants were asked to annotate their symptoms using the provided interface while immersed in a VR sickness–inducing scene for ten minutes. It was important to select an appropriate scene for the experiment, one that elicits the target VR sickness symptoms without overwhelming the user. For this purpose, we used the Cybersicker scene [1]. This scene was developed to study induced VR sickness effects and consists of an amusement-ride scenario in which the user sits virtually in a cart that rotates around both the yaw and pitch axes.

By default, the rotation is modeled using sinusoidal functions that alternately accelerate motion along each axis, mimicking the behavior of a real amusement ride and producing a wave-like acceleration pattern. To elicit varying levels of discomfort and capture dynamic symptom fluctuations over time, we modified the scene's motion speed during the experience by increasing, decreasing, and maintaining it at different intervals as illustrated in Fig.2.

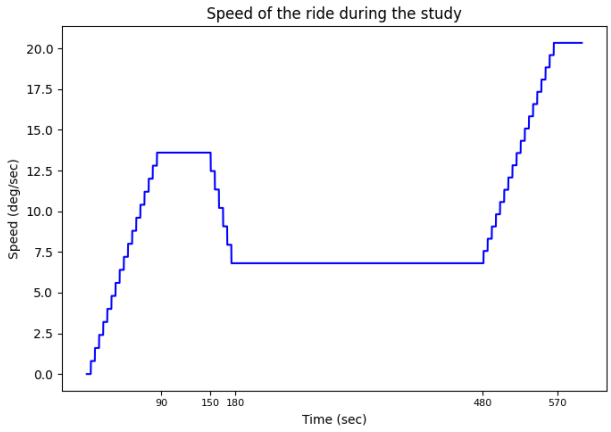

Fig. 2. Modified speed profile of the amusement ride during the Cybersicker scene. The speed increases during the first 90 seconds and is then maintained for 60 seconds. It is subsequently reduced to half for 30 seconds, followed by a constant speed phase lasting 300 seconds. Finally, the speed increases for 60 seconds and is maintained at this final level for an additional 30 seconds.

S.3 Individual time-series of VR sickness over time.

This section presents representative individual time-series plots of participants and sessions (Fig. 3 and 4). They show symptom intensity (Y-axis) over time (X-axis), recorded with the continuous interface (green) and the CSQ-VR questionnaire (red). To improve the readability of the plots, the Y-axis is scaled individually for each participant and session, ranging from zero to the maximum VR sickness level reported by that participant, rather than spanning the full theoretical range (0–1).

The plots illustrate the temporal information captured by the continuous interfaces compared to the discrete questionnaire. The continuous interfaces provide more temporal information throughout the experience, whereas the CSQ-VR only captures symptom intensity at three discrete time points. When the two measures overlap in time, they follow similar tendency, supporting the comparability of the interfaces with the standardized questionnaire.

Although the order in which interfaces were presented to participants was randomized to mitigate memory issues, the plots in this section are grouped by interface type and displayed according to overall participant preference (Dials, followed by Bars, and Halos).

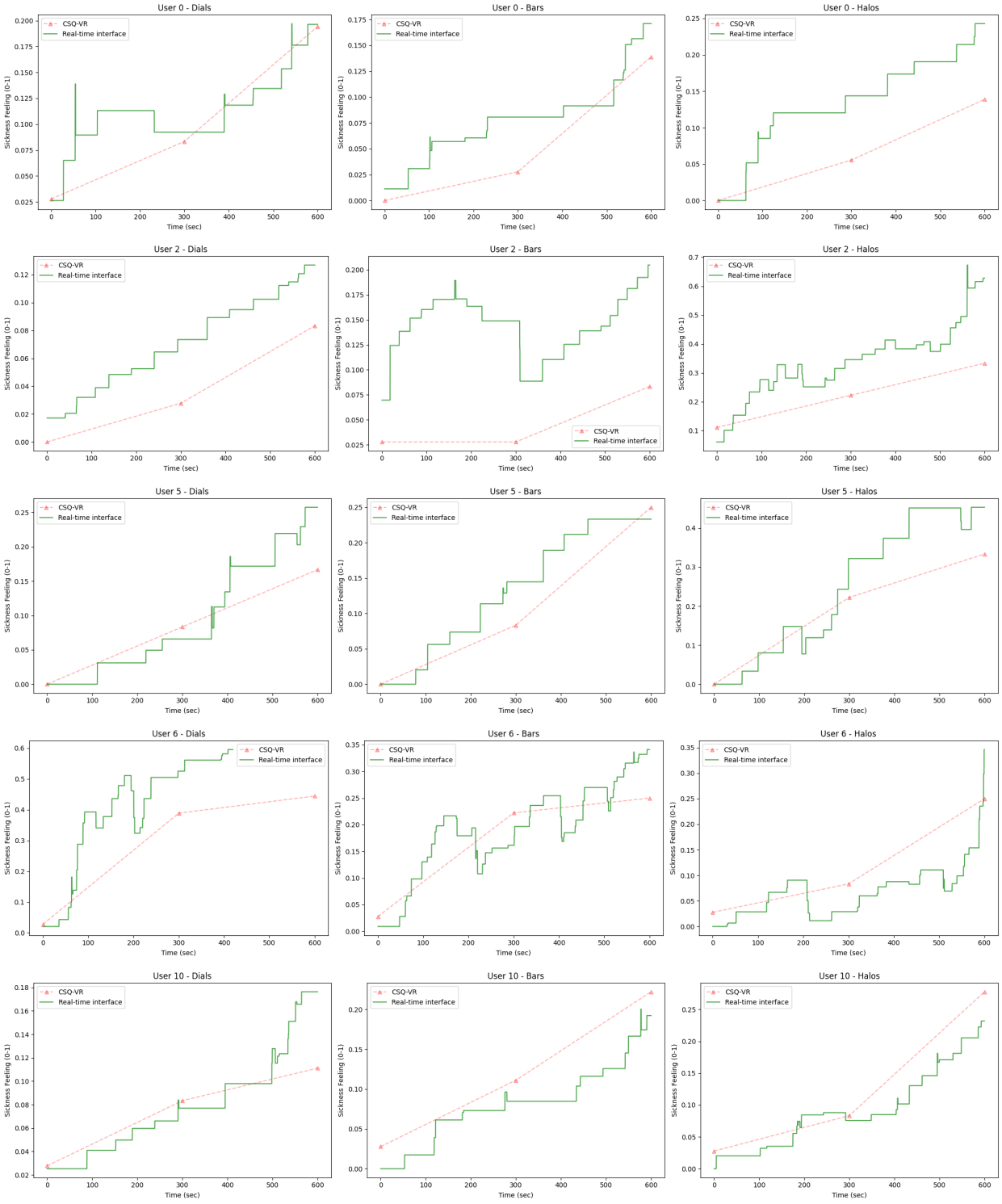

Fig. 3. Representative individual time-series showing VR sickness intensity over time, recorded with the continuous interface (green) and the CSQ-VR questionnaire (red).

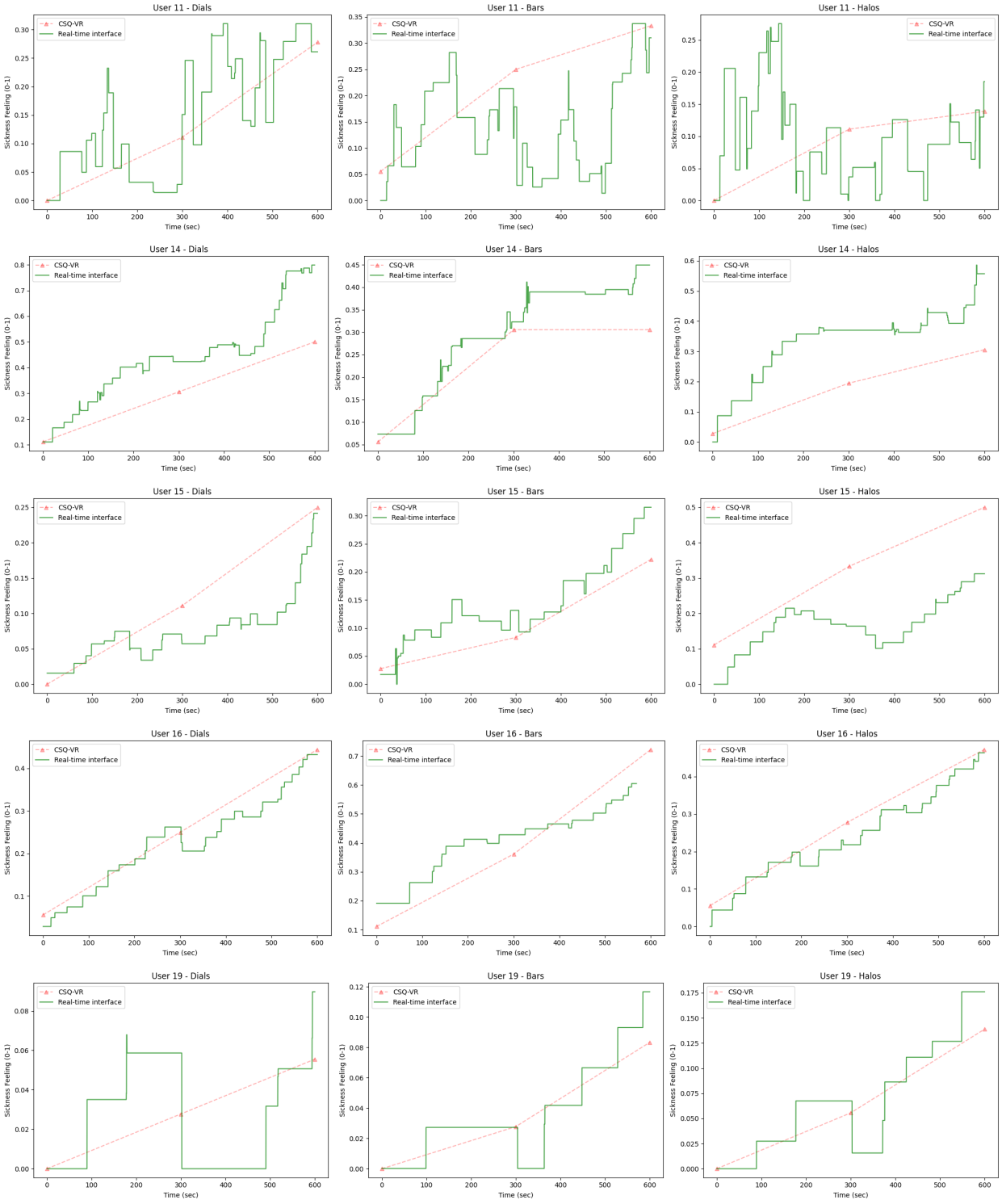

Fig. 4. Representative individual time-series showing VR sickness intensity over time, recorded with the continuous interface (green) and the CSQ-VR questionnaire (red).

S.4 Qualitative feedback from semi-structured interviews

To further evaluate the design of the interfaces, we conducted a series of semi-structured interviews with the participants. The purpose of these interviews is to collect qualitative feedback and identify potential interface improvements.

The semi-structured interviews were guided by a set of predefined themes and core questions that allowed flexibility for open-ended discussion. The interviewer took written notes of participants' responses.

Interviews were conducted at the end of each experimental session, consistent with the within-subjects design of the study, resulting in three interviews per participant (one per interface, across three separate days). The same interview structure and themes were discussed on each day. In the third session, once the participant had experienced all three interface prototypes, a ranking task was added to the interview. Participants rank the interfaces from most to least preferred (to support recall, images of each interface were shown to participants).

Overall Evaluation. Participants consistently expressed their preference for the proposed continuous interfaces over Likert Scale questionnaires. Some mentioned reasons for this preference are: continuous interfaces allowing participants to rely on relative perception when adjusting symptom intensity, since they are always visible, whereas questionnaires mostly require to remember the last score; continuous interfaces allowing to annotate intermediate values (e.g., between 2 and 3), reducing decision effort compared to selecting discrete values, where the majority of participants tend to choose the lowest available option when uncertain; additionally, answering questionnaires interrupted the immersive experience and tented to reduce symptoms during the response period, leading to uncertainty about whether to report how they felt before the interruption or their momentary intensity. This reinforced the opinion in continuous interfaces as more intuitive, faster to report, less cognitively demanding, and more faithful representation of real-time experience. Interestingly, participants frequently described the Halos interface as visually the most appealing and aesthetically pleasing. However, it was the lowest ranked (Dials=63; Bars=50; Halos=37) due to reported difficulties interpreting the radius of a 360° colored area as a precise indicator of symptom intensity. Besides, some participants found it harder to control using the joystick, as movement in any direction could be interpreted as increasing intensity. While similar representations have shown strong acceptance in other areas of HCI, they were perceived as less functional in this study for precise VR sickness reporting.

Possible sickness adaptation effect across sessions. Participants reported a possible adaptation to the VR sickness-inducing scene across sessions. Several participants experienced strong dizziness during the first session (three were unable to complete the full exposure on the first day), but completed subsequent sessions with reduced discomfort. This may suggest there is an adaptation effect that would be interesting to study in future research.

Potential interface improvements. Based on participants' feedback, we hypothesize that certain design parameters could benefit from personalization. In particular, an adjustable joystick speed may improve user experience, allowing users to choose between linear or non-linear speed control, where non-linear mapping would make slower intensity modulation for small adjustments and faster for larger changes. Additionally, adaptive visual settings, such as configurable highlighting background colors, could help maintain interface visibility across varying scene conditions.

Representative anonymized interview.

I - Interviewer

P01- Participant 01

Participant P01 completed three sessions on separate days:

S1 = Day 1 (Dials interface)

S2 = Day 2 (Bars interface)

S3 = Day 3 (Halos interface)

I: The purpose of this interview is to collect qualitative feedback on the interface you have just interact with and to identify potential areas for improvement. There are no correct or incorrect answers. Participants are encouraged to share their honest impressions and experiences. *Differentiation of symptom dimensions.* Did you find it easy to distinguish between the three symptom dimensions

represented in the interface? Did you use the three elements equally, or did some receive more attention than others? Did you notice whether some symptoms tended to appear before others during the experience?

P01–S1 (Dials): It is not clear where "dizziness" falls (element 1 or element 2 of the interface). I perceived it in the element 2, though during the CSQ questions I realized it fits better in the first category. I have used the three elements but more the element 2.

P01–S2 (Bars): The three controls were well-differentiated.

P01–S3 (Halos): [No specific comment provided for this dimension in S3].

I: *Influence of questionnaire administration (CSQ-VR).* When you answered the CSQ-VR questionnaire during the VR experience, did that moment influence your sickness symptoms making them worse or better? Do you think this influence was reflected in your questionnaire answers?

P01–S1 (Dials): When stopping for the questions, certain symptoms like disorientation become more noticeable because the brain had adapted to the previous rhythm. In the end, it seems to stabilize again. Since the questionnaires interrupt the experience, symptoms improve toward the end, which affects how I answer.

P01–S2 (Bars): Regarding the form canvas, I think it did influence me, but since I felt less sick this time, the impact was lower. Last time, when the canvas appeared, I was more aware of the symptoms.

P01–S3 (Halos): [No specific comment provided for this dimension in S3].

I: *Influence of the continuous interface on symptoms.* Did using the interface make your symptoms worse, better, or have no noticeable effect? Were you able to use the interface even when you felt uncomfortable or unwell?

P01–S1 (Dials): It may provide a stable fixed reference for some participants, which could help prevent dizziness. It would be interesting to examine where participants look using eye-tracking and whether there is a relationship between gaze behavior and higher levels of dizziness.

P01–S2 (Bars): [No specific comment provided for this dimension in S2].

P01–S3 (Halos): [No specific comment provided for this dimension in S3].

I: *Perceived trust and accuracy.* How easy did you identify your symptoms to transfer them to the interface? Do you feel that the interface recordings accurately reflected your internal state? How easy was it to distinguish different intensity levels when interacting with the interface (especially minimum and maximum values)?

P01–S1 (Dials): I am in favor of the dials; having the last mark constantly visible helps with measuring variation or relative change. This makes it easier and more reliable than a 1-to-7 scale (absolute measurement).

P01–S2 (Bars): The "pie chart" style is more visual than the bar for indicating symptom intensity (referring to the area that fills up).

P01–S3 (Halos): When symptoms increase, the circle fills up a lot and it feels like "too much." There might be some reluctance to fill it so much due to the non-linear scale.

I:For the following characteristics, participants are encouraged to consider extreme cases, i.e., whether a feature is so helpful that removing it would be detrimental, or whether its current implementation is confusing and might be better omitted or redesigned. *Icons (emojis).* Did the icons help you understand the meaning of each symptom dimension?

P01–S1 (Dials): They are understood very well and I liked them a lot. Perhaps the icon on the left influences the "urge to vomit."

P01–S2 (Bars): [No specific comment provided for this dimension in S2].

P01–S3 (Halos): [No specific comment provided for this dimension in S3].

I:*Wording support.* Did the wording reference help you understand the symptoms? Did you use the wording help during the experience? Was a single word sufficient, or would you have preferred more explanation?

P01–S1 (Dials): They look good; one word is enough. These are well-selected in my opinion since they summarize each symptom category.

P01–S2 (Bars): [No specific comment provided for this dimension in S2].

P01–S3 (Halos): [No specific comment provided for this dimension in S3].

I:*Text visualization.* Was the text easy to read in terms of typography and size?

P01–S1 (Dials):The triple wording signs extend too far to the right. The purple ones were not clearly visible because there was poor contrast with the background from my perspective. I had to slightly move my head until there was a lighter background behind the purple text.

P01–S2 (Bars): [No specific comment provided for this dimension in S2].

P01–S3 (Halos): [No specific comment provided for this dimension in S3].

I:*Graphic visualization.* Were the visual elements easy to perceive in terms of color, saturation, transparency, shadow contrast, and size?

P01–S1 (Dials): The colors help a lot and are well-selected. I actually think I would associate each color with its corresponding symptom.

P01–S2 (Bars): [No specific comment provided for this dimension in S2].

P01–S3 (Halos): [No specific comment provided for this dimension in S3].

I:*Positioning.* Did you find the position of the interface elements comfortable and appropriate within the VR environment?

P01–S1 (Dials): Comfortable to look at.

P01–S2 (Bars): The bars interfere more with the field of vision.

P01–S3 (Halos): [No specific comment provided for this dimension in S3].

I:*Interaction and functional design.* Was it easy or difficult to understand how to interact with the interface? Were interaction cues (e.g., color changes, trigger zones) clear and sufficiently large?

P01–S1 (Dials): Natural. The size/trigger is comfortable. Red lighting reminders are okay. It is easier to use the interface than to click the "continue" buttons in the tutorials.

P01–S2 (Bars): It is more comfortable to move the joystick from left to right than up and down (the moving joint is different). I prefer the dials because it is easier to point at them than at the bars with the trigger (bounding box).

P01–S3 (Halos): I liked the interaction and the reaction; aesthetically it feels more organic.

I:*Intrusiveness.* Did the interface interrupt your VR experience, or did it feel integrated?

P01–S1 (Dials): They are not intrusive.

P01–S2 (Bars): The bars interfere more with the field of vision.

P01–S3 (Halos): They don't seem intrusive. Maybe a bit more than the dials because they are slightly larger, but not by much.

I:*Speed and control.* Were you able to register your state quickly using the interface? Did the joystick sensitivity feel appropriate, or did you need to frequently correct intensity adjustments?

P01–S1 (Dials): Joystick speed: occasionally had to rectify, but very good overall. Interaction speed: okay.

P01–S2 (Bars): Joystick speed: okay. I had the feeling it grew faster with the dials than with the bars (perhaps due to the colored area).

P01–S3 (Halos): Joystick speed: okay, although perhaps it went up and down with more readjustments, not so much because of speed but because of the mental adjustment of the radio-to-filled-area relationship.

I:Ranking (Third Session Only). Looking at the image that shows all three interface prototypes you have seen during the three sessions, rank them from most to least preferred. Why?

P01-S3: 1: Dials, 2: Halos, 3: Bars. Bars are the worst because they are harder to aim at and more intrusive with the content. I had many doubts between Dials and Halos. However, the Halo has a very non-linear perception/scale: when I increase it slightly, the covered area is very small, but for the same increment at a higher level, the area covered is larger, giving the impression of higher intensity. Dials have a more linear scale/perception, and we are more accustomed to "pie charts." I like the interfaces more than the questionnaire. Since the interface is always there, you may feel inclined to give feedback more frequently than necessary. Although seeing it all the time is not intrusive, perhaps you could study whether it should be shown constantly. Maybe let the user decide when it appears or allow them to adjust the transparency.

References

[1] Daniel Zielasko and Yuen C. Law. "Cybersicker: An Open Source VR Sickness Testbed - Do you still have fun, or are you already sick?" In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*. Christchurch, New Zealand: Association for Computing Machinery, October 2024. ISBN: 9798400705359. DOI: 10.1145/3641825.3689503.