Semiotic Framework for Virtual Reality Usability and UX Evaluation: a Pilot Study

Barbara Rita Barricelli
Dept. of Computer Science
Università degli Studi di Milano
Milan, Italy
barricelli@di.unimi.it

Ambra De Bonis
Dept. of Computer Science
Università degli Studi di Milano
Milan, Italy
ambra.debonis@studenti.unimi.it

Serena Di Gaetano
Dept. of Computer Science
Università degli Studi di Milano
Milan, Italy
digaetano@di.unimi.it

Stefano Valtolina
Dept. of Computer Science
Università degli Studi di Milano
Milan, Italy
valtolin@di.unimi.it

ABSTRACT
This paper presents the results of a pilot study aimed at validating the Semiotic Framework for Virtual Reality (VR) usability and user experience evaluation (UX). The framework offers a theoretical model for VR applications classification and a combination of evaluation methods and a study protocol to be used for testing usability and UX in the VR field. The main goal of our approach is to provide a complete framework able at overcoming and correctly interpreting the discrepancies that may arise from the application of cognitive and semiotic methods of evaluation. The positive preliminary results of the pilot experiment led the authors to the design of a full-scale study that is already ongoing and that is focused on developing a complete tool of evaluation for VR.

Author Keywords
Virtual Reality, Semiotic Engineering, Usability, User eXperience.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; I.3.7 Three-Dimensional Graphics and Realism: Virtual Reality.

THE SEMIOTIC VR FRAMEWORK
The Semiotic Virtual Reality Framework [1][2][3] is the outcome of research in the field of semiotic analysis of Virtual Reality (VR) communication, focused on syntax, semantic, and pragmatics [4][5], that represent the three levels of the framework: the syntactic level as defined by the characteristics of the visual communication adopted in a VR application, the semantic level as related to the functional model chosen to design the application, and the pragmatic level as the one based on the human-computer interaction that changes the user's role. This approach stems on both a study of literature review on theoretical research by Eco [4] and Greimas [5], paired with a long-time experience in VR research and development and in Human-Computer Interaction and Design in general.

The Semiotic VR Framework can be used to classify and describe different kinds of virtual reality applications and to better understand communication in VR. It represents both a tool for evaluating existing VR applications and for supporting designers of VR systems in their decision-making processes. To exploit the framework potentials, designers and developers have to select the appropriate level of detail and likeliness in visualization, interaction, and modelling, choosing the appropriate sensory stimulation systems, determining the necessary languages for performing a successful human-VR communication. The framework can be depicted as a three-dimensional space (see Figure 1), where the three axes represent the range of variation of Structure (or syntax), Model (or semantic), and Interaction (or pragmatics representation) of the applications at hand. The Structure axis is relative to the syntactic level, which ranges from symbolic to highly realistic (better called likely). To identify a position in this axis, we need to consider the iconicity level and the likeliness level of Computer Graphics solution adopted as well: the iconicity level helps in locating the position, while the likeliness level suggests possible Computer Graphics solution to obtain the desired iconicity. The Model axis is relative to the semantic level; it ranges from mathematical to impressionistic. To identify a position in this axis, we will consider the detail level of the underlying mathematical, physical or chemical model that rule the evolution of the VR world, or the presence of symbolic or

GHItaly18: 2nd Workshop on Games-Human Interaction, May 29th, 2018, Castiglione della Pescaia, Grosseto (Italy)

logical model of the evolution. The Interaction axis is relative to the pragmatic level; it ranges from abstract to concrete, considering also the narrative aspect of communication of the system. To identify a position in this axis, we consider the interaction approach and the interaction devices that exercise different sensory systems. Using the framework, a VR application can be located in terms of parameters that allow to identify the expressive power of the communication solution provided by VR. The 3D-space can be described as organized into 8 octants, each one characterized by a triplet representing respectively the Interaction, Structure, and Model values (see Table 1).

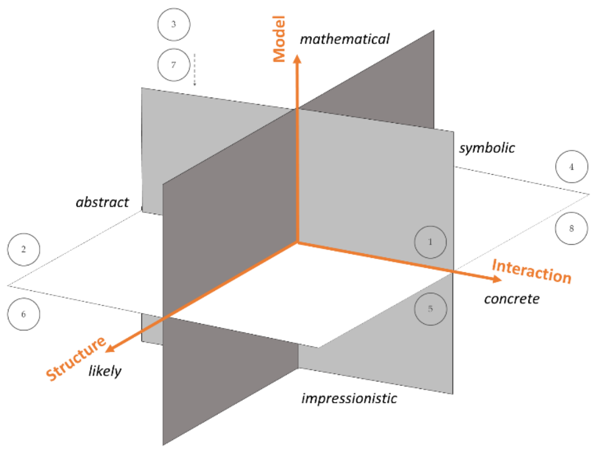

Figure 1. The Semiotic VR Framework in a 3D-space divided into octants.

Octant	Interaction	Structure	Model
1	Concrete	Likely	Mathematical
2	Abstract	Likely	Mathematical
3	Abstract	Symbolic	Mathematical
4	Concrete	Symbolic	Mathematical
5	Concrete	Likely	Impressionistic
6	Abstract	Likely	Impressionistic
7	Abstract	Symbolic	Impressionistic
8	Concrete	Symbolic	Impressionistic

Table 1. The VR Framework Octants

PILOT STUDY

To validate the framework, we designed a pilot study that applied a combination of semiotic and cognitive evaluation methods for measuring both usability and User eXperience (UX). We chose eight VR applications and designed a user test with a limited set of participants. In what follows, the test environment, participation selection, protocol, and results of the study are illustrated.

Test Environment

The test environment (depicted in Figure 2) was constituted by three main devices: a Samsung Gear VR headset equipped with a Samsung Galaxy S7, a monitor equipped with a Chromecast used for mirroring (real-time streaming) the Samsung Gear VR experience, and a laptop used as hotspot for Internet access and connected both with the Samsung Galaxy S7 and the monitor. This environment was set up in a silent laboratory where just the participants (one at a time) and one observer were present during the user test.

Figure 2. The environment setting for the user test.

We selected 8 VR applications, one for each of the framework's octant and asked the participants to complete a specific task with each of them. The observer, thanks to the mirroring of the interaction on the monitor, was able to taking notes about the participant behavior and to observe the way the user moved and acted in the virtual space.

In Car Racing VR [1] (Octant 1)
Car driving simulator that offers multiple camera modes. Easy to learn but difficult to master.

In Mind [2] (Octant 2)
Short arcade adventure that allows the user to explore a brain in search of neurological disorders.

Human Anatomy VR [3] (Octant 3)
Educational app that allows to explore the human body for learning general anatomy.

Star Tracker VR [4] (Octant 4)
3D Star Field depicted into a sphere surface that could be looked at from outside.

Bandit Six: Salvo [5] (Octant 5)
Shoot 'em up game in which the user needs to protect an island from the enemies.

A Night Sky [6] (Octant 6)
The user connects dot, i.e. stars, to build constellations with simple point and touch controls, and brings wonderful creatures to life.

[1] https://www.oculus.com/experiences/gear-vr/1409977735730829/

[2] https://www.oculus.com/experiences/gear-vr/742896805825051/

[3] https://www.oculus.com/experiences/gear-vr/1658650407494367/

[4] https://www.oculus.com/experiences/gear-vr/1438854922813902/

[5] https://www.oculus.com/experiences/gear-vr/1009334549088838/

[6] https://www.oculus.com/experiences/gear-vr/1613977911951627/

VISO Places [7] *(Octant 7)*
An exploration app that uses Google StreetView and Wikipedia and enables also group exploration.

PAINT VR [8] *(Octant 8)*
Through a wide set of brushes the user can paint in a 3D space and then export screenshots.

Participants
Being a pilot study, the number of participant was limited: we involved 10 users (5 female, 5 male), aged from 18 to 30 years, and 7 of them had experienced virtual reality with headset devices before (see Figure 3). According to the initial questionnaire they filled in and a non-structured interview, the 14 adjectives used to describe their previous experience with VR were *addictive/engaging* (5 times), *enjoyable* (4 times), *stimulating* (1 time), *tiring* (2 times), *annoying* (1 time), and *destabilizing* (1 time).

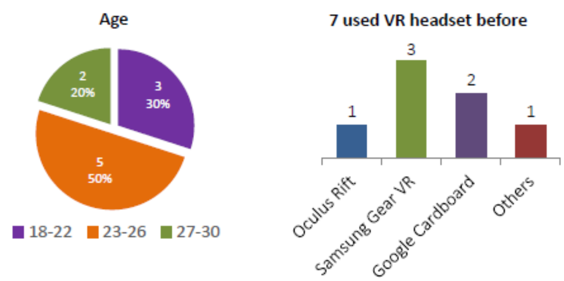

Figure 3. Age of the user test participants and their previous experience with VR headset.

User Test Protocol
The user test protocol of this pilot study was designed by pairing cognitive and semiotic methods of usability and UX evaluation, with the aim of collecting significant information for identifying not only eventual usability and UX problems, but also for highlighting the communication breakdowns that might take place during the interaction with the VR applications. After the submission to the participants of an initial demographic questionnaire, we organized the user tests into several steps. We divided the eight applications into two groups of four. The participants were asked to perform a task-based user test with the first set of four applications, adopting a think-aloud protocol. At the end of the first part of the test, four usability questionnaires (one per application) were filled in by the participants. The last two steps were repeated for the second set of four applications. To conclude the user test, a final UX evaluation questionnaire was submitted to the participants. The usability questionnaires were composed by 26 Likert scale questions – a combination of SUS (System Usability Scale) [6] and CSUQ (Computer System Usability Questionnaire) [7], while the UX evaluation questionnaire was created according to the UEQ (User Experience Questionnaire) [8] method. SUS is a questionnaire very broadly used, especially in industry, that can return reliable results even when administered to small sample of users. CSUQ is a questionnaire developed by IBM and is mostly focused on measuring the satisfaction in using the application or tool under evaluation. UEQ permits to assess feelings, impressions, and attitudes that arise when the users use the application of tool: it is aimed at measuring aspects that are typical of usability evaluation approach – efficiency, perspicuity, and dependability – and those that are typical of UX – originality and stimulation. All user test sessions were recorded, both the audio and the interaction with the apps. This allowed us to apply the Communicability Evaluation Method (CEM) [9], a Semiotic Engineering method for evaluating the communicability of an application, was applied.

App	SUS score
In Car Racing VR	76.25 (11.36)
In Mind	74.50 (9.34)
Human Anatomy VR	69.00 (13.84)
Star Tracker VR	73.00 (12.29)
Bandit Six: Salvo	74.75 (12.47)
A Night Sky	75.75 (8.88)
VISO Places	**61.25 (15.62)**
PAINT VR	**56.25 (18.68)**

Table 2. The SUS questionnaire results (with standard deviation)

App	CSUQ - SYSUSE
In Car Racing VR	3.77 (0.47)
In Mind	3.90 (0.29)
Human Anatomy VR	3.78 (0.82)
Star Tracker VR	3.94 (0.56)
Bandit Six: Salvo	3.62 (0.73)
A Night Sky	3.93 (0.57)
VISO Places	3.62 (0.72)
PAINT VR	3.14 (0.77)

Table 3. CSUQ results - SYSUSE (with standard deviation)

[7] https://www.oculus.com/experiences/gear-vr/943598329068595/

[8] https://www.oculus.com/experiences/gear-vr/1123989124339476/

App	CSUQ - INFOQUAL
In Car Racing VR	3.89 (0.40)
In Mind	3.51 (0.59)
Human Anatomy VR	3.39 (0.60)
Star Tracker VR	3.43 (0.68)
Bandit Six: Salvo	3.51 (0.82)
A Night Sky	3.57 (0.55)
VISO Places	**2.99 (0.85)**
PAINT VR	3.23 (0.92)

Table 4. CSUQ results – INFOQUAL (with standard deviation)

App	CSUQ - INTERQUAL
In Car Racing VR	**2.57 (0.82)**
In Mind	3.30 (0.80)
Human Anatomy VR	3.73 (0.71)
Star Tracker VR	3.83 (0.92)
Bandit Six: Salvo	3.63 (0.84)
A Night Sky	4.13 (0.60)
VISO Places	3.60 (1.04)
PAINT VR	**2.83 (1.20)**

Table 5. CSUQ results - INTERQUAL (with standard deviation)

App	CSUQ - OVERALL
In Car Racing VR	3.62 (0.40)
In Mind	3.66 (0.23)
Human Anatomy VR	3.63 (0.64)
Star Tracker VR	3.74 (0.59)
Bandit Six: Salvo	3.58 (0.69)
A Night Sky	3.83 (0.51)
VISO Places	3.83 (0.72)
PAINT VR	3.13 (0.78)

Table 6. CSUQ results - OVERALL (with standard deviation)

Usability and UX Evaluation Results

SUS

The app that reached the highest SUS score is In Car Racing. This app belongs to the octant (Concrete, Likely, Mathematical). The users defined it as simple to use and learn. On the other hand, the app with the lowest SUS score is PAINT VR that is characterized by Concrete interaction, like In Car Racing, but has Symbolic structure and Impressionistic model and together with VISO Places they did not reach the sufficient average score for SUS, which is 68. From the results reported in Table 2, it can be seen that the interaction, either concrete or abstract, did not contribute to the success or failure of SUS evaluation.

CSUQ

As to SYSUSE (system usefulness, Table 3) Star Tracker VR is the app with the highest SYSUSE result (3.94). It has been defined by the participants as very easy to use with an easy to learn navigation structure. On the other hand, PAINT VR received the lowest evaluation on SYSUSE, even if still sufficient (3.14). The participants judged the navigation very difficult and pointed out a vision fatigue experience. For INFOQUAL (information quality, Table 4), the best app is In Car Racing VR (3.89), while PAINT VR received a not-sufficient score (2.99) because of the lack of information and instructions of use. The app that scored the highest value for INTERQUAL (interface quality, Table 5) is A Night Sky (4.13), while the worst result is for In Car Racing VR (2.57), because of its too simple graphics and the vision fatigue reported by the participants. Finally, for OVERALL (overall satisfaction, see Table 6) the best app is A Night Sky (3.83), while the worst one is PAINT VR (3.13). A Night Sky reached very positive results thanks to its simplicity of use and the very good graphics. The participants reported that the fantasy genre strongly supports the user engagement. On the contrary, PAINT VR causes important vision fatigue episodes and also did not present all the features that the users would expect from such an application.

CEM

With the CEM analysis, we detected several communication breakdown, even if the cognitive usability analysis (SUS/CSUQ) results were quite positive for more than half the apps. Only two apps out of eight did not present communication breakdowns: In Car Racing VR and Bandit Six: Salvo. The CEM analysis results, in terms of tags and patterns, are reported in Table 7. The most used tag was “Where is it?” (3 times), followed by “I give up”, “What happened?”, and “Looks fine to me” (2 times). From the pattern, one can notice that for two apps, the participants gave up with trying to use the app after the tags “What happened?” and “I can’t do it this way”, both pointing out that the communication between the app and the user is not effective nor efficient.

UEQ

The UEQ questionnaire has been submitted once for each participant, meaning that the evaluation was performed on the entire experience of use with the Samsung Gear VR and not on the single apps. Figure 4 shows the results for each of the six aspects: attractiveness, perspicuity, efficiency, dependability, stimulation and novelty. Attractiveness (the capacity of attracting and stimulating the interest of the user) and stimulation (the capacity of a product to be motivating for the user and to raise curiosity) obtained an above-average evaluation (1.367 and 1.125). This is linked to the interest of the participants in using the Samsung Gear VR device, but still reveal some reservations. Perspicuity (easiness of getting familiar with the product), obtained a slightly below-average score (0.975); some participants in fact had problems in acquiring the use notions of the Gear VR. Efficiency and dependability, used to measure the quality of use of the device, obtained insufficient values (0.550 and 0.625) because the majority of the participants were affected by vision fatigue. A good result was scored by the Novelty aspect (1.150): the participants were surprised by the features of the device and appreciate its potentials.

App	Tags	Patterns
In Car Racing VR	-	-
In Mind	I give up. Where is it? What happened? Why doesn't it?	(I give up -> Why doesn't it?)
Human Anatomy VR	Looks fine to me. Thanks but no thanks. I can do otherwise.	-
Star Tracker VR	Where is it?	-
Bandit Six: Salvo	-	-
A Night Sky	What now? Where am I? Oops.	(Where am I? -> What now?)
VISO Places	I give up. What happened? I can't do it this way.	(What happened? -> I give up), (I can't do it this way -> I give up)
PAINT VR	Looks fine to me. Where is it?	-

Table 7. CEM analysis results – tags and patterns

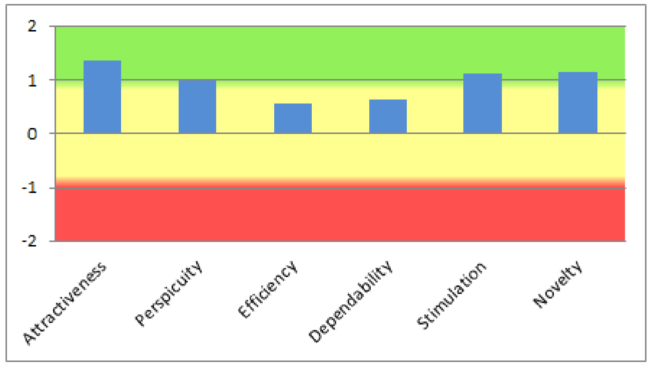

Figure 4. The UEQ results.

CONCLUSION

This pilot study was aimed at validating the Semiotic Framework for VR published in [1] and [2]. The results are presented for the first time, while the design of the pilot study was illustrated in a poster in [3]. This paper presents the results in detail, highlighting the discrepancies obtained by the application of different methods of usability and user experience evaluation, typical of the cognitive and the semiotic engineering approaches. These discrepancies are typical of the application of the so-called "classical" methods on what can be seen as "innovative" tools or applications: some well-known studies on the topic – e.g. [10][11][12] demonstrate that classical methods of usability evaluation applied to innovative tools and applications return negative results despite their popularity and success; this framework, its combination of methods, and the evaluation protocol are designed to be the answer to this problem with respect to the specific field of Virtual Reality. Given the positive results obtained in the pilot study, we are already performing a full-scale study involving 30 participants.

REFERENCES

1. D. Marini, R. Folgieri, D. Gadia, A. Rizzi. 2012. Virtual Reality as a communication process. *Virtual Reality* 16, 3, 233-241.
2. B. R. Barricelli, D. Gadia, A. Rizzi, and D. L. R. Marini. 2016. Semiotics of virtual reality as a communication process. *Behaviour & Information Technology* 35, 11, 879-896.
3. B. R. Barricelli, A. De Bonis, A. 2017. Semiotic Virtual Reality Framework Validation. *In Proceedings of the Doctoral Consortium, Posters and Demos at CHItaly 2017*. CEUR WS, Vol. 1910, 102-105.
4. U. Eco. 1976. *A Theory of Semiotics*. University Press.
5. A. J. Greimas. 1983. *Du Sens II: Essais Sémiotiques*.
6. J. Brooke. 1996. SUS: A Quick and Dirty Usability Scale. In: P.W. Jordan, B. Thomas, B.A. Weerdmeester & I.L. McClelland (Eds.), *Usability Evaluation in Industry*.
7. J. R. Lewis. 1995. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation

and Instructions for Use. *International Journal of Human-Computer Interaction* 7, 1, 57-78.

8. B. Laugwitz, T. Held, and M. Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *USAB 2008,* LNCS 5298, 63-76.
9. C. Sieckenius de Souza, and C. F. Leitao. 2009. *Semiotic Engineering Methods for Scientific Research in HCI*. Morgan and Claypool Publishers.
10. S. Greenberg, and B. Buxton. 2008. Usability evaluation considered harmful (some of the time*). In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, 111-120.
11. P. A. Silva, A. Dix. 2007. Usability: Not as we know it! *In Proceedings of the 21st British HCI Group Annual Conference (HCI 2007)*, 103-106.
12. J. Hart, C. Ridley, F. Taher, C. Sas, and A. Dix. 2008. Exploring the facebook experience: a new approach to usability. *In Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges (NordiCHI 2008)*, 471-474.