Experiences with Repertory Grid Analysis for Investigating Effectiveness of Virtual Environments

Anthony Steed and Janet McDonnell
Department of Computer Science, University College London
Gower St, London, WC1E 6BT, United Kingdom
{A.Steed | J.McDonnell}@cs.ucl.ac.uk

Abstract

We investigate the application of the repertory grid analysis (RGA) in the investigation of effectiveness or virtual environments. RGA is a structured interview technique based on the Kelly's Personal Construct Theory (Kelly, 1955). Through two pilot experiments, we show that RGA is a useful companion to other forms of evaluation such as presence questionnaires. It avoids problems of imposition of terminology when using questionnaires and allows the participant to develop their own terminology for discussing experiences of virtual environments. The technique allows us to explore the aspects of the virtual environment that may enhance or detract from the sense of being there.

1. Introduction

One definition of an effective virtual environment might be an environment where the person experiences a sense of presence within that environment and acts according to the stimuli received, not the situation of presentation. There are many facets to research on presence (Draper et al. 1998), but one key one that has preoccupied the community is measurement of presence through questionnaires (e.g. Witmer and Singer, 1998, Schubert et al., 1999). Questionnaires are often unsatisfying because they are post-experience and there may be a number of problems with the language that they use. However, they are still commonly used, since alternatives such as the breaks in presence method (Slater and Steed, 1999) and physiological measures (Meehan et al., 2002) are still being developed.

There is also a gulf between the nature of experiments on presence, which determine the effect of a small number of independent variables on an overall rating of presence, and the need to evolve usability methods that are adapted to the situation of virtual environments where there may be no well-defined task to support (Bowman et al. 2002, Tromp et al. 2003). In this abstract we present repertory grid analysis (RGA) as a potential way of bridging this gulf.

2. Repertory Grid Analysis

2.1. Background to Method

According to the theory of personal constructs (Kelly, 1955), constructs are ways of construing the world, enabling people to respond to what they experience in ways which are 'explicitly formulated or implicitly acted out' (op.cit. p.6). Kelly's repertory grid technique was originally developed for use in clinical psychology as a technique for exploring individuals' personal constructs about interpersonal relationships in the context of psychotherapy. However the technique has been applied extensively to the elicitation of personal constructs for many other purposes. The technique has a sound theoretical basis as a conversational tool without the necessity to 'buy into' Kelly's original claims that psychological events are real phenomena. The work we report here uses repertory grids as a means of arriving at an understanding of how participants make sense of experiences of (construe) virtual environments through a 'grid conversation'.

The repertory grid technique is a content-free procedure for exploring and for forming personal constructs. Underlying the technique is the notion that people can represent their experiences – the situations with which they are faced – by placing alternative constructions upon them. A grid conversation encourages a participant to make clear what, for him or her, distinguished one experience from another. A construct is a way in which some things are construed as being alike and yet different from others. It is therefore inherent in the nature of a construct that it is bipolar. Constructs give us the dimensions of personal meaning, the poles of a construct are the limits of its dimension.

2.2. Method

The process of eliciting personal constructs from a participant using the repertory grid technique proceeds by identifying a collection of experiences. These determine the scope of the ensuing conversation. In the cases described below, participants experienced a series of virtual environments, usually six different ones, these formed the elements discussed during the grid conversations. The chosen items of experience constitute the set of elements to be compared and contrasted with one another to elicit personal constructs. Any distinction that is important to the participant is a valid construct. The conversation proceeds as follows. Elements are grouped into threes (triads), since at minimum a construct is a way in which at least two elements are similar and contrast with a third (op.cit. p.61). The participant is asked to compare the elements in the triads, to consider their similarities and differences and to describe them.

Constructs are elicited by considering different combinations of three elements until no more constructs seem to be emerging and a good cross-section of the possible combinations has been considered. A 'raw' grid is drawn up in which elements – the experiences being compared form columns and each construct elicited forms a row. Descriptions of the poles of each construct – which satisfy the participant as accurate – label each end of each row. The 'raw' grid is subsequently 'focussed' by encouraging the participant to assign a rating to each construct for each experience where that is possible. In the work described below, for example we used a five-point scale to indicate where an experience lay with respect to the poles of each construct. (One pole is arbitrarily assigned a rating of 1, its opposite 5). Elements which are assigned to similar poles can then be clustered, constructs which discriminate between elements similarly can also be clustered. By this means a pattern of personal meaning emerges. Conversation with the participant should continue throughout grid focussing and the participant should be an active participant throughout, as focussing routinely leads participants to refine their the constructs.

The technique relies on good conversational technique and critically on the choice of experiences that form the elements. If a representative selection of elements is used at the end of the process each construct should finally represent an important dimension of the participant's construing. In short, a repertory grid expresses something about the way a person looks at things, using terms which they choose and revise themselves – but which they may not find it easy to express through other means.

3. Pilot Trials

Two pilot trials were carried out to investigate the applicability of the RGA. The first pilot took a set of environments with a very wide range of modeling and behavioural descriptions with three presented immersively, three non-immersively. The second pilot used a set of environments that were all presented on an immersive projection system.

3.1. 1st Pilot

In this pilot, we were primarily interested in whether RGA would yield useful comparisons between virtual environments. Since virtual environments are typically presented for short periods, it was not clear if participants would be able to make useful or consistent comparisons between them. The pilot was purely exploratory in nature and used a series of environments constructed by students on the MSc Anon at Anon. The environments varied greatly in design and interactively. Three were presented on Virtual Research V8 head-mounted display and three on a desktop display. All environments were constructed in the DIVE software and were run on an SGI Onyx.

- W1: Odessey, a spacecraft from 2001 A Space Odessey (HMD presentation)
- W2: Space station with three galleries with alien artifacts (HMD presentation)
- W3: Womb world, an exploration and puzzle solving world (desktop presentation)
- W4: Global sports, a futuristic sports center (HMD presentation)
- W5: Restaurant, a cartoon restaurant (desktop presentation)
- W6: Island, a tropical paradise surrounded by shark infested waters (desktop presentation)

Three participants experience each environment. Each participant generated 7 or 8 constructs and subsequently refined the poles and ranked each environment. Lack of space prevents us from discussing the analysis in detail. What was important about the results is that participants were able to make many constructs concerning the environments despite the fact that they saw each environment for 5 minutes or less. Given the wide variety of environments, many constructs concerned the design of the environments. Some very similar constructs were found by more than one participant. A selection is given in Table 1. A rating of one meant

that the element was associated with the left-hand pole, and a rating of 5 meant that the element was associated with the right-hand pole. A construct that isn't applicable to an element is marked with an X.

Table 1: Selection of constructs from 1st pilot

	Environment						
	W1	W2	W3	W4	W5	W6	
Triggers natural body movement	5	1	3	5	4	2	No intuitive reaction
Urges discovery of familiar (seen before) places	X	X	4	4	4	1	Meets expectation – no surprises
No engaging task	3	1	2	4	5	5	More involved and unaware of outside world
Voyeuristic	1	2	3	4	4	5	Interactive
Brighter	3	5	1	3	3	3	Darker

We have used the WebGrid II software (WebGrid), to analyse the constructs. Constructs can be analyzed to extract similar clusters, and a principle component analysis can be done to extract those constructs that best explain the difference between two elements. Lack of space prevents us from giving details results of the such analyses, but please see http://Anon/Anon for example materials.

3.2. 2nd Pilot

From the 1st pilot we determined that the method would generate interesting constructs concerning the participants' experience. In the 2nd pilot we used a series of environments that we have studied in more depth in other experiments. The study was more focused this time since each environment is self-consistent and they are all of a fairly realistic design, and are not obviously of "fantastic" places. We were thus investigating whether the RGA would be sensitive if there was less variation between the elements of experience. Each of the environments was presented on Anon's Trimension ReaCTor, a CAVE™-like system, with three back-projected walls and a front-projected floor. The environments and the tasks were:

- W1: Pit, take a series of items from one room to and place them in the adjoining room on a chair.
- W2: Talk, give a short talk to an audience.
- W3: Virtual City, locate a statue and remember its position so you can locate it on a map.
- W4: Mansion, find a collection of possible murder weapons inside a mansion.
- W5: Cubes, build the coloured cube from the eight pieces with each face being one solid colour.
- W6: Poster, unscramble phrases written on posters.

Five participants undertook the experiment. Each saw the environments in a random order, and after each environment they completed the Slater-Usoh-Steed questionnaire (SUS) (see Slater and Steed, 2000 for an example). The SUS questionnaire contains a number of questions related to presence that are ranked on seven-point Likert scales. Figure 1 shows the average rating on the presence questions for each participant and each environment.

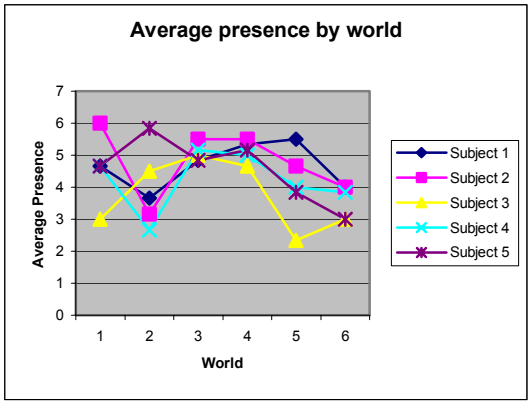

Figure 1 Participants' average rating of presence for each environment in the 2nd pilot

The variability across environments, especially W1, W2 and W5, means that even with large numbers of participants, it would probably be very hard to classify the environments into high and low presence generating. There may be two significant effects: W6 is rated lower on average and W3 and W4 are rated highly. With so few participants, nothing concrete can be determined, but the RGA allows us to generate hypotheses about possible causes that can be studied in further experiments. For example, two participants rated W2 as nerve-racking, and this was distinct from most of the other environments, except W6 where there was pressure of a complex task with a time limit. On the other hand, a construct from another participant was "felt inside the world working with something" v. "standing there not in the world", where W2 was rated 5, and all the other environments were rated as 1 or 2. This leads us to believe that there is a variation in threshold of believability between participants, and that the this participant didn't feel any emotional connection to the presented avatars. The construct that most eloquently describes the difference of W3 and W4 from the others was "somewhere you could really be" v. "can tell its VR" (W3 and W4 were rated 1).

Other constructs that are intriguing included "seemed unlikely" v. "more real place", "fantasy" v. "a place I've been", "able to move about" v. "dealing with a screen" and "can easily imagine as a world" v. "can't easily imagine as a world". These all suggest that the participants are being quite sophisticated when comparing environments and may be interpreting the virtual environment experience as places they *could go to,* not places that they *have been to*.

4. Discussion

The RGA has helped us start to investigate what aspects of the virtual environments cause people to feel more or less present in the virtual environment. We are currently using it in a series of trials concerned with agoraphobia. Given the wide range of possible environments we could build, we decided to start by comparing several of our existing demonstrations environments. These involve a range of types of spaces from cramped interior to wide exterior and from unpopulated to densely crowded. In the design of these environments there is a trade-off between using "obvious" features such as crowds and detailed, expansive models, and then making sure that the environments are considered consistent by the participants. The RGA has helped us identify the important role of avatar representations and behaviours, and this will be a focus of more detailed follow-up studies.

Aside from use as a complement to other methods, we imagine that RGA might be useful in constructing or refining questionnaires. It should be noted that users often spontaneously use concepts related to standard questionnaires such as SUS or Witmer and Singer presence questionnaire (Witmer and Singer, 1998). The RGA could complement other approaches such as factor analysis (Schubert et al., 1999).

Acknowledgements

Thanks to David Swapp who managed the ReaCTor system and helped with two sets of trials. Thanks also to students on MSc Vision, Imaging and Virtual Environment (1999/2000) who took part in the 1st pilot by building worlds and using the RGA as part of the their coursework. This work was supported in part by the EPSRC Interdisciplinary Research Collaboration, EQUATOR, GR/N15986/01.

References

Bowman, D., Gabbard, J., Hix, D. (2002) A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. Presence: Teleoperators and Virtual Environments, vol. 11, no. 4, pp. 404-424.

Draper, J. V., Kaber, D. B., Usher, J. M. (1998) Telepresence. Human Factors, 40(3), 354-375.

Kelly, G.A. (1955) The Psychology of Personal Constructs, vol 1 and 2. Norton, New York.

Meehan, M., Insko, B., Whitton, M.C., Brooks Jr., F.P. (2002) Physiological measures of presence in stressful virtual environments. ACM Transactions on Graphics 21(3): 645-652

Schubert, T.W., Friedmann, F., Regenbrecht, H.T. (1999) Decomposing the sense of presence: factor analytic insights. Presented at the Second International Workshop on Presence, University of Essex, Colchester, 6th and 7th April 1999.

Slater, M., Steed M. (2000) A Virtual Presence Counter, Presence: Teleoperators and Virtual Environments, 9(5), October 2000, MIT Press, ISSN 1054-7460.

Tromp, J.G., Steed, A., Wilson, J., (2003) Systematic Usability Evaluation and Design Issues for Collaborative Virtual Environments, Presence: Teleoperators and Virtual Environments, MIT Press.

WebGrid III http://tiger.cpsc.ucalgary.ca:1500/WebGrid/WebGrid.html (accessed July 20th 2003)

Witmer, B.G., Singer, M.J. (1998) Measuring Presence in Virtual Environments: A PresenceQuestionnaire, Presence: Teleoperators and Virtual Environments, 7(3), 225-240.