VR-based psychological therapy is largely used in different clinical approaches and a variety of studies can be found in the literature about the experience of VR-therapy sessions. Usually, such studies are focused on a user's private experience. With this contribution, we present a perspective to improve VR-session definition towards a "co-defined reality", abandoning both artifact and user–artifact centered interaction in favour of a "situated and context sensible" interaction analysis. Theoretical background is based on an ethnomethodological approach: this perspective gives evidence of how people, in specific social situations, produce shared meanings. Within such theoretical framework, an analysis has been conducted with reference to two different dimensions. From a strictly ergonomic point of view, main results are related to the improvement of VR training sessions and to patient integration in the framework of the therapeutic protocol, to avoid interruptions and breakdowns. With reference to the sense of presence, we investigated how therapist and patient can populate with “others” and “meanings” the co-defined environment and how a VE’s level of interactivity influences the quality of the immersive experience. Main findings highlight the sequential and dialogical process of VR fruition and could help in VR technical design process.

Introduction

Use of Virtual Reality (VR) technology in psychological therapy is largely documented in the literature: virtual environments (VE) have been used by therapists within different backgrounds, and within several theoretical approaches (Krijn et al., 2004; Rothbaum et al., 2001; Rothbaum et al., 2004). Besides such richness, it's not easy to find a common shared definition of Virtual Reality, as pointed out by Riva (2004).

We can identify three different approaches to defining VR: (a) definitions focused on technical elements, (b) definitions based on user perceptions and (c) definitions based on user experiences. Tech-focused definitions refer to hardware and software elements of VR, such as the head-mounted display (HMD) or head-tracker. According to Ellis (1994), VEs can be defined "...as interactive, virtual image displays enhanced by special processing and by nonvisual display modalities, such as auditory and haptic, to convince users that they are immersed in a synthetic space" (Ellis, 1994). In recent years, definitions evolved to considering design topics: for example, Brey defines VR as "a threedimensional interactive computer-generated environment that incorporates a first-person perspective" (Brey, 1999). As stated by Steuer (1993), a VR definition focused only on technological aspects is inadequate, and research needs to focus on the analysis of system factors (vividness and interactivity) that have an influence on human experience. User perception-based definition focuses on what the user can perceive of VEs while considering perceptual realism, spatial sounds and multi-sensorial perceptions: “We define a virtual environment as a set of data displayed in such a way as to create in users the impression of objects in a space” (Nunez & Blake, 2001). Recent works underline the importance of user experience, like immersion, presence and social presence: “defining characteristics of virtual reality systems is to create presence and fool the user into believing that one is, or is doing something in the synthetic environment” (Dongsik et al., 2003). This position is shared by Riva who defines the core element of VR as a mental experience, which makes the user believe that ‘he or she is there’, that he or she is present in the virtual world (Riva, 1999).

In this essay, we want to stress how these three aspects (technical elements, user perceptions,
user experiences) could be compared to situated-action theory elements: object, subject and process (Galimberti & Riva, 2001). Within this metaphor, virtual reality, like any other artifact, could be studied looking at its objective characteristics (performance, functionality, hardware and software, for example), at users and actors that use it, and finally looking at dynamic processes fired by VR use, as communication, immersion, presence and mediated communication. Therefore it could be defined as a social artifact: users can interact with other users, processes and living experiences, using artifact elements (Galimberti et al., 2006).

This consideration requires analyzing VR social dimensions of interactional data. Allport defines social psychology as “an attempt to understand and explain how the thought, feeling, and behaviour of individuals are influenced by the actual, imagined, or implied presence of others” (Allport, 1985). A question we want to raise is: how is the “other” in VR? How can a user imagine, imply or interact with an actual “other” during VR sessions? Theoretical background is based on an ethnomethodological approach: this perspective gives evidence of how people, in specific social situations, produce shared meanings. From an ecology of state-oriented perspective we expanded the focus to the concept of ecology of process. Ecology of state includes both an ecology of context, which has been exhaustively defined and conceptualized by different research streams, and an ecology of situation. The wider ecology of process concept introduces the value of dialogical perspective, extending the social dimension of the studied context. Within the dialogical approach, VR-based therapy sessions could be studied as a social co-construction of meanings, where therapist and patient negotiate what is going on, how it's going and who is present. From this point of view, both patient and therapist are interacting within a medium, with a medium and with the other in the medium. The last is an emergent actor that therapist and patient co-define within their interaction and conversation; its emergence allows us to study the VR-session as a social context where a new, more flexible way of producing and interpreting data is needed, since it is originating separately from therapist, patient and their interaction with/within virtual environments.

**Method**

In order to explore social dimensions of VR experience, we analyzed 54 session transcriptions, collected during the MIUR-FIRB NeuroTIV project, with 9 patients and 3 therapists. The protocol used was the ECT (Experiential Cognitive Therapy), which integrates VR experiences with the traditional CBT (Vincelli et al., 2002). Virtual environments used were the Panic Disorders and Agoraphobia VR modules developed in the framework of the project: an underground station, a town square, a supermarket and a lift (Belloni et al., 2006). All environments contained static avatars, represented with 3D models (see Figure 1) and 2D sprites. We used a qualitative software analysis called NVivo 2, developed by Qualitative Solutions and Research (QSR) International in 2002, to compile and analyse the data.
We used a qualitative content analysis approach to underline how therapist and patient interact within virtual reality, and how they co-define the existence of the “other”. Content analysis is “the use of replicable and valid method for making specific inferences from text to other states or properties of its source" (Krippendorff, 1969), and it has two major approaches: inductive category development and deductive category application. The main idea of the inductive procedure is to formulate a criterion of definition, derived from theoretical background and research questions. Following this criterion, data is analyzed and categories are considered tentative and deduced step by step. Within a feedback loop those categories are revised, eventually reduced to main categories and checked in respect to their reliability. If the research question suggests quantitative aspects (e.g. frequencies of coded categories) it can be analyzed (Mayring, 2000). Such a decision implies that data is approached with the goal of allowing the categories (nodes) to emerge from the reading of the qualitative data (Gibbs, 2002).

In order to document the process of category development, we present a "documentational table" (Constas, 1992), representing the three procedural elements of categorization (origination, verification and nomination) and the temporal aspects of the process. According to Constas (1992), origination identifies the locus of category construction, verification details the strategies used to support the creation and application of categories, and nomination is concerned with the naming of categories; while temporal designation addresses the temporal characteristics of such a process.

Categories generated during this study are related to the 1) type of conversational action (1.1 Request, 1.2 Description and 1.3 Scaffolding action and attributes of action: 2) Locus (Inside VE and Outside VE); 3) Focus (Place, Other Agnition, Intention Attribution) and 4) Center, as defined below. Table 1 shows the components of categorization domain on the vertical axis and the temporal designation domain on the horizontal axis. Categories are represented by numbers located according to the components and temporal characteristics associated with a given category.

The origination component is based on three sources for generating the categories. The Scaffolding action category is based on both literature and investigative perspective: Steiner introduce the idea to use VEs in order to “scaffold the students' pre-existing knowledge of story structures” (Steiner, 1994) and in a psychosocial approach, we decided “a priori” to investigate how the scaffolding function could be related to therapist’s actions. In this perspective, the therapist could support the patient’s description of VE and its experience. Others categories are based on an interpretative position, and they emerged during data analysis and have been designated in a iterative confrontation by the investigators.
The verification component was applied both iterative and at the end of the coding process in a technical way: once a draft of categories definition was prepared, the next step was to check that different coders could replicate each other’s work. To pretest categories, we randomly selected 11 sessions (representing 20% of data) and two coders independently coded the selected sessions. To quantify the degree of agreement during the preliminary pretest and final intercoder reliability, we used Cohen's kappa method as suggested by Dewey (1983). As reliability levels in the pilot test ranged from
.65 to .72, we proceeded to code the full sample: according to the exploratory nature of this study and to considerations presented by Rietveld and van Hout (1993), we judged this range acceptable. The nomination component was accomplished in an “a priori” literature-based way for Scaffolding category (Steiner, 1994), and “a posteriori” for other categories, with labels based on an interpretative position.

Results
We’ll briefly explain each category, providing an example.

Type of conversational action: patient and therapist could perform three types of actions. They can do a Request, a Description, and a Scaffolding Action.

A Request action means that the actor (usually the therapist) asks the other to do something (to describe something, to perform a movement, to speak about something):

T117: “Can you describe what you see?”
T8: “Could you go upstairs?”

A Description is usually a reply done by the interlocutor, while describing an environment, an action, or a sensation.

P116: “There is a map on the wall … It’s colourful”

More interesting is the Scaffolding action, done mainly by the therapist: he could confirm and support the description provided by the patient, enhancing his impressions.

T118: “Could you tell me more about that map?”

Locus: the analysis of the qualitative data suggests that each action, besides the therapeutic process, could refer to two different “loci”: Inside VE and Outside VE. This node refers to “where” the action is performed, and “where” the actor is perceived; the patient can describe an action done inside the VE, for example:

P12: “I’m going downstairs”

Performed during an exploration of the Subway VE. Or an action done in the real world:

P34: “I’m pulling the joystick, now”

Focus: this attribute refers to the element perceived, described or interacted with. An actor can describe a place referring to its physical, spatial or functional elements (Place), or he can speak about who is present (Social), and what they are doing. The Social node could be deeply divided into: Other agnition, when speaker focuses on a specific “other” and Intention Attribution, performed when an avatar is being driven by a purpose.

Center: this node codes an action centered on the speaker. We can find many descriptions:

P73: “I’m in the middle of a large square”

or, more often, descriptions of social environment:

P54: “There are a lot of people around me”

This node is usually related during post-hyperventilation phase (as requested by the protocol), or while the patient is describing the VE without any previous request to do it.

Such dimensions could be found mixed in a wide range of actions: therapist could ask to perform an
action inside the VE, and the patient describe an action done outside it; or therapist could describe a corrective (while solving an hardware issues) action outside the VE, and patient would speak about an external sensation (for example, speaking about a previous experience). We present in Table 2 the frequency of different nodes.

### Table 2

<table>
<thead>
<tr>
<th>Category</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Action</strong></td>
<td></td>
</tr>
<tr>
<td>Action</td>
<td>1860</td>
</tr>
<tr>
<td>Request</td>
<td>864</td>
</tr>
<tr>
<td>Description</td>
<td>672</td>
</tr>
<tr>
<td>Scaffolding action</td>
<td>324</td>
</tr>
<tr>
<td><strong>Locus</strong></td>
<td></td>
</tr>
<tr>
<td>Inside VE</td>
<td>339</td>
</tr>
<tr>
<td>Outside VE</td>
<td>103</td>
</tr>
<tr>
<td><strong>Focus</strong></td>
<td></td>
</tr>
<tr>
<td>Place</td>
<td>281</td>
</tr>
<tr>
<td>Social</td>
<td>391</td>
</tr>
<tr>
<td>Other agnition</td>
<td>295</td>
</tr>
<tr>
<td>Intention Attribution</td>
<td>96</td>
</tr>
<tr>
<td><strong>Center</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>442</td>
</tr>
</tbody>
</table>

### Discussion

Analyzing conversation during VR-based sessions with a psychosocial approach, researchers can underline some elements: each action could refer to an in-world or out-of-world target, and could be performed inside or outside the VE; places and actions could be described by their “physical” dimensions, or actors can choose to focus on their social meanings; therapist and patient can reciprocally support or detract their shared representation of the VE.

From a preliminary analysis, we can notice that users describe social elements of VE more frequently than physical (space, objects, colors and light, for example): 391 descriptions are based on social features while 281 on physical elements. Social description could be focused on general characteristics, or on a specific “other”: 295 social descriptions referred to an avatar, its appearance, its position or its imagined actions (as stated above, in our VEs avatars where mere 3D static models). Moreover, 96 of these descriptions are about intentions, expectations, feelings of such avatars; an important element of the Intention Attribution code is that we can find it more frequently after a Scaffolding action by the therapist: actors co-define the existence of others inside the VE. A typical sequence points out, from a social perspective, that there is an “emergent other”, acting in the VE: to a
request done by the therapist to describe the spatial VE, the patient can reply by describing a social element of it; on the next turn, if the therapist performs a scaffolding action, supporting such social description, patient focuses on the Other Agnition, and then attributes an intention to the avatar. After that, the patient will speak directly to the avatar:

P143: “Hello, miss... Could you let me pass, please?”

while trying to go around an avatar blocking his way to the subway train.

This sequence has some rules: (a) the therapist could easily stop this sequence describing an action performed outside the VE, or asking to describe a previous experience; (b) after such a sequence, the patient would usually perform descriptions of the social VE; (c) following description will be focused on the self in VR. In the same way, therapists could modulate the sense of being in the VE by performing specific actions: by using (or avoiding) a scaffold act; supporting (or not) patient descriptions; diverting the focus of an action from inside the VE towards outside; by introducing external elements, for example a request to think about a previous, real-world experience; or request to describe the VE “around you”, positioning the patient inside the VE.

Regarding presence and social presence concepts, this framework could help the understanding of a user’s experience of being inside the VE. Users often (442 on a total of 672 description) speak about events and elements as related to their position (“around me”, “near”, “I’m on a train” and so on), and 339 times they described elements inside the VE. We can therefore trace a continuum based on the co-defined position of the patient and the existence of others: from “I’m here, in the research room, alone, moving around a picture on my HMD using a joystick” to “I’m here, in a mall, trying to find a way out avoiding the crowd”. We argue that it’s possible to attribute a presence level regarding dialogical position assumed by the actors on this continuum.

**Conclusion**

We opened this essay with a question: are VR-based sessions a social context that can be analyzed with psychosocial tools and methods?

In order to find an answer we analyzed 54 sessions and defined some characteristics of conversational actions performed. By conversing, patient and therapist co-define and negotiate a shared world. This shared world is not a social vacuum, but it’s inhabited by a plurality of actors: the patient and therapist themselves but also the other they are speaking about and the other they are speaking to them.

This consideration could help researchers to understand dynamics, interactions and experiences within Virtual Environments: by studying the conversational negotiation of the sense of being there, we can find clues about presence, social presence; as well ideas to design and implement better VE for therapy. Further studies should compare the sense of presence attribution based on conversational actions to some other presence measurement methods, such as behavioural elements or standardized questionnaire.
Bibliography


Riva, G. (1999). From technology to communication: Psycho-social issues in developing virtual environments. *Journal of Visual Languages and Computing, 10*

Riva, G. (2004). *Virtual reality for health care: The status of research*


Steiner, K.E., & Moher, T.G. (1994). *Scaffolding story construction with interactive multimedia*. Association for the Advancement of Computing in Education.
