Key elements of social presence for a psychosocial approach to VR based CBT: an interaction analysis

Gloria Belloni, MA, Matteo Cantamesse, Ph.D. candidate, Carlo Galimberti, Ph.D.
Università Cattolica del Sacro Cuore di Milano, Milan, Italy
gloria.belloni@gmail.com

Abstract
An effective VR based CBT (Cognitive Behavioural Therapy) requires an explicit knowledge of presence and social presence dynamics fired by virtual reality immersion. Most of the literature suggests that CMC does not have the capacity to support social and affective interaction, however, recent reviews of the social presence literature question the extent to which this literature can be generalised to every communication medium and to every application of these media. In the present contribution, we propose a qualitative method based on a psychosocial approach to investigate different dimensions of the therapeutic experience and interaction. The final objective of the study is the improvement of the VR based therapy system designed by our research unit both from an ergonomic and, consequently, clinical point of view according to a perspective that is strongly ‘context driven’.

Social Presence and VR based therapy

The genealogy of the social presence construct can be traced back to Mehrabian’s (1969) concept of immediacy: “Immediacy is the directness and intensity of interaction between two entities”. Short et al. (1976) introduced and defined the term social presence as “the salience of the other in a mediated communication and the consequent salience of their interpersonal interactions”. According to Riva (2004) this principle has significant outcomes on design process and ergonomic issues. The VR system used in our study based on the ECT (Experiential Cognitive Therapy) protocol for the treatment of anxiety disorders, recognises in ‘other’ presence an important therapeutic and functional element thus deserving, in our opinion, a careful investigation in order to allow interventions on two fronts:

- **Ergonomic**: the design of technical functionalities as they are implemented will have to support, and if necessary reinforce, the perception of the highest levels of social presence with specific reference to intentionality and ‘sense attribution’ that implies the ability to understand the other’s intentions.

- **Clinical-therapeutic**: reflections emerging from the study can be at the same time considered relevant aspects from the clinical point of view and consequently lead to a further improvement of the clinical protocol currently in use, stressing other distinctive elements in comparison with non VR based CBT.

The research framework: an analysis respectful of the ecological process

The study reported is part of the NeuroTIV project “Immersive Virtual Telepresence Managed Care for the Assessment and Rehabilitation in Neuropsychology and Clinical Psychology”
Theoretical approach

As stated above, first definitions of ‘social presence’ can be dated back to 1976. More recently, Biocca et al. (2001) propose a clarification of the concept classifying different definitions into three themes respectively centred on co-presence and mutual awareness; on the experience of psychological involvement; on the behavioural interaction through which social presence is realised. Nevertheless the concept is still unclear and mainly used in different ways without specific distinction: with reference both to the medium in mediated communication and to perceptions, behaviours or attitudes of the participants (Rettie, 2003). Related to social presence are the concepts of ‘awareness’, defined as “an understanding of the activities of others, which provides a context for your own activity” (Dourish and Bly, 1992), and more recently the one of “connectedness” (Adler’s 1998) referring to an individual’s attitude and relationship to society and considered as one of the main principles motivating social behaviour (Smith and Mackie, 2000). Schumie et al. (2001), Insko (2003), Riva (2004) propose exhaustive reviews of main methods applied to study the concepts of presence and social presence on the basis of the different theoretical approaches related to the nature of these constructs. This kind of research is generally based on questionnaires or psycho-physiological measures useful to investigate the level of presence and social presence but that aren’t enough to understand what is happening and how interactive context can help the therapist to keep the patient’s “sense of being there” and the patient’s sense of the “other perception” in control. We agree with the approach presented by Spagnolli et al. (2003) that “recognizes that presence is an ambitious concept referring to the user experience in the VE, which is complex, contextualized and dynamic. It stresses the reciprocal contribution of both environment and its inhabitants in configuring each other and the central role of local action in shaping presence”. The authors identify some focuses which must be considered in presence research: the process through which presence is constituted and changed, the problems of the virtual body, of the VE boundaries and of the objects recognised in the simulation, the denial to consider a single perspective on the concept of social presence as the only possibility rather than allowing the coexistence of different configurations. The qualitative methods applied in the study presented are strongly based on perspectives that consider the action as the focus of analysis. In particular, Situated Action Theory (SAT) is an approach developed within the context of socio-cognitive research known as “cognition in practice” and introduces a change of perspective: the action is not considered as the execution of a ready-conceived plan,
but as a process of adaptation to the context. Suchman underlines that “instead of separating actions from the circumstances in which occurs as the execution of a carefully thought out plan... [SAT] tries to study how people use circumstances to develop an intelligent course of action” (Galimberti, Riva, 2001, p.167). We think that such a perspective can be generally applied to presence research and to social presence research more specifically.

Methodology

The study, aimed at discovering some key concepts related to social presence, is based on the interaction analysis (Suchman and Trigg, 1991, Goodwin, 2000). This is a qualitative method that has its focus on the ‘action’, at the same time observable as an object, and produced in a local and specific context. Within a 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 is 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 the virtual environments. Within the framework of qualitative approaches it would be of central interest to develop the aspects of interpretation, the categories, as near as possible to the material, to formulate them in terms of the material. For that scope, qualitative content analysis has developed procedures of inductive and deductive category development, which are oriented to the reductive processes formulated within the psychology of text processing (Ballstaedt, Mandl, Schnozt & Tergan 1981; van Dijk 1980). Deductive category application works with prior formulated theoretical derived aspects of analysis, bringing them in connection with the text. The qualitative step of analysis consists in a methodological controlled assignment of the category to a passage of text.

Procedure

The specific object of the study is to verify some key aspects connected to social presence and their relevance for therapy. The study includes a first analysis of occurrences in quantitative terms and the consequent in context evaluation that is the analysis of the specific situation in which a verbal exchange occurs, a thought is expressed, or an action done. The reference context for the coding definition is represented by the specific aims and therapeutic actions foreseen in the VR based ECT protocol for each one of the sessions. Main therapeutic actions have been coded through the identification of an ideal interactional sequence constituted by: therapist act/request, patient act/answer, therapist act/resumption. Both therapist and patient have then defined a series of therapeutic acts/requests and possible reactions/answers. These last ones are specifically based on the social presence level determination proposed by Riva (2004). Special codes called modulators have then been added: they allow further qualitative specifications (positive/negative evaluation; answer coherent/non coherent to the specific request). Therapeutic interventions not strictly connected to the VR phase of the session aimed, for example, at the rational understanding of the cognitive restructuring process have not been deeply considered because they are not necessarily connected to the VR use.

Categories used for the analysis are exhaustively described in a parallel study conducted in the framework of the same project (Cantamesse & Galimberti 2007) and are synthetically described in the table below.
<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action</td>
<td>Actions performed by the actors</td>
</tr>
<tr>
<td>Request</td>
<td>The actor (usually the therapist) asks the patient to do something or to speak about something</td>
</tr>
<tr>
<td>Description</td>
<td>A reply given by the interlocutor (usually the patient): he can describe an environment, an action, a sensation</td>
</tr>
<tr>
<td>Scaffolding action</td>
<td>The therapist can confirm and support the description provided by the patient enhancing his impressions or interpretations</td>
</tr>
<tr>
<td>Locus</td>
<td>“Where” the action is performed and “where” the actor is perceived</td>
</tr>
<tr>
<td>Inside VE</td>
<td>Action performed in or description referred to the VE</td>
</tr>
<tr>
<td>Outside VE</td>
<td>Action performed in or description referred to the real world</td>
</tr>
<tr>
<td>Focus</td>
<td>Element perceived, described or interacted with</td>
</tr>
<tr>
<td>Place</td>
<td>Actor describes an action or a sensation in a place referring to its physical, spatial or functional elements</td>
</tr>
<tr>
<td>Social</td>
<td>Actor describes an action or a sensation in a place referring to who is present</td>
</tr>
<tr>
<td>Other cognition</td>
<td>The speaker is focused on a specific “other”</td>
</tr>
<tr>
<td>Intention Attribution</td>
<td>An avatar is judged being driven by a purpose</td>
</tr>
<tr>
<td>Center</td>
<td>Actions or descriptions are centered on the speaker</td>
</tr>
</tbody>
</table>

Table 1. Categories’ description

In order to evaluate the reliability and the validity of the coded data, 12 sessions were randomly assigned to two judges. Each judge received brief training as well as a codebook, with a definition and an example for each node. Cohen’s Kappa statistics were also computed for each node. Kappa coefficients ranged from .69 to .82.

The analysis procedure is based on ‘content analysis’ carried out on transcriptions of interactions between therapists and patients within the VR based Experiential-Cognitive Therapy (ECT) treatment protocol for Panic Disorder and Agoraphobia. The present version (Vincelli et al., 2000) is based on the classic CBT protocol developed by Clark (1986). For VR based ECT, our research unit developed the Virtual Environments for Panic Disorders with agoraphobia (VEPDA) virtual reality system (Belloni, Cantamesse, Galimberti, & Gatti, 2006). VEPDA is a 4-zone virtual environment developed using the GameStudio A6® engine toolkit. The four zones reproduce different potentially fearful situations - an elevator, a supermarket, a subway ride, and a large square. In each zone the therapist can set up the characteristics of the anxiety-related experience. In particular the therapist can define the duration of the virtual experience, its end, possible exploration paths and the number of avatars to be included.
Data corpus is derived from transcriptions of 54 VR based ECT sessions (3 patients for each of the 3 therapists involved in the project for a total of 9 patients). Subjects are consecutive patients, aged 18-55, seeking treatment in one of the institutions involved in the study. They met DSM IV criteria for panic disorders and agoraphobia for a minimum of 6 months as determined by an independent clinician on clinical interview, according to the SCID (Structured Clinical Interview for DSM IV) model (Spitzer et al, 1992; William et al., 1992). The VR based ECT protocol has been carried out by three independent therapists after an adequate training phase.

As to the type of data produced, the reference to the ecology of process (Galimberti, Belloni, Cantamesse et al. 2006) may be presented as follows:

<table>
<thead>
<tr>
<th>Focus of attention</th>
<th>Type of data</th>
<th>Interaction level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Therapists</td>
<td>Observation of therapists’ use of VEs</td>
<td>Personal experience of VE use</td>
</tr>
<tr>
<td></td>
<td>Data concerning therapist-patient interactions during VE experience (transcriptions)</td>
<td>Situated interaction between users (therapist-patient)</td>
</tr>
<tr>
<td>Patients</td>
<td>Observation of the patients’ use of VEs</td>
<td>Personal experience of VE use</td>
</tr>
<tr>
<td></td>
<td>Data concerning patient-therapist interaction during VE experience (transcriptions)</td>
<td>Situated interaction between users (therapist-patient)</td>
</tr>
</tbody>
</table>

Table 2. Typology of data produced

**Key aspects and examples of results**

In this work, categories suggesting that the VR can be considered as a psychosocial experience
have been used to analyze and interpret therapeutic sessions with a special focus on social presence with the aim to find out possible solutions to improve VR environments and protocol. On the basis of the categories described in the previous paragraph, some further enriching elements have been identified. In particular: 1) the existence of reality check dynamics. This means that in specific situations the therapist invites the patient to compare what is happening in the virtual world with a possible, or simply imagined, real situation. This procedure seems to be connected to specific therapeutic phases and purposes and will be further investigated; 2) Scaffolding actions: the therapist can confirm or support descriptions provided by the patient, enhancing his impressions and comprehension of the situation. They are considered significant elements assuming what we defined sense making function. Therapist and patient tend to give sense to the situation without using a narration in the proper sense. 3) Social and centered descriptions seem to be connected to the anxiety level reported by the patient. Some examples of results deriving from the content’s analysis are now presented.

When does the patient focus on social environment?
One of the first emerging key points concerns the coherence between requests to describe the general VR environment made by the therapist (“Tell me what you see” or “Try to describe the plaza”) and the patient’s reply. According to the category previously defined, requests by the therapist and answers of the patient are usually of two types: related to general descriptions of the VE and mainly focused on objects, or related to descriptions focused on other persons (“There are a lot of people. The lady is smiling”).

Considering the following excerpt, we see that the focus of the patient’s answer is on the description of other people present in the VE:

Therapist 81: “Ok ... What do you see around? Describe me all the objects you see around”
Patient 82: “Well, now...now you have the coach with the straps. It’s not...very crowded...there are men and women, with classic and casual dresses and a...a woman”
T82: “Perfect.”
[....]
T83: “Ok, stop for a while G., how are you feeling now?”
P84: “I’m not…I’m not at my ease”

From a preliminary quantitative analysis of the occurrences based on a frequency calculation, it seems to emerge that when the anxiety level increases, patients’ answers and descriptions are more focused on social elements of the environment than on objects, even if the therapist request specifically refers to the description of the objects in the environment. This kind of answer has been marked as non coherent (to the specific request). The anxiety level can be defined and checked thanks to frequent feedback (an average of one every 3 minutes) solicited by the therapist from the patient about his or her sensations or feelings. It is evaluated on a subjective scale from 1 to 10.

On the contrary, when the anxiety level is lower, the patient answers the request of the therapist concerning a general description of the environment focusing on the objects around him in a more coherent way. This does not happen only when the anxiety symptoms are voluntarily provoked as foreseen by the clinical protocol (in this case the request to describe objects is usually connected to the distraction strategy the patient is expected to apply), but also if the frightening stimulus emerges while the patient is freely exploring the environment. A successful way to study this specific aspect is then to further investigate in which specific phases of the
session this happens and to look for possible connections with the interaction style of each therapist.

**Perception of Self in VR and social environment**
A second finding consists in the fact that, in general, when the patient describes the environment focusing on social elements such as detail referred to persons present in the environment (code defined as description of social VE), he is inclined to describe himself as effectively present in VR, for example referring other actions or intensions to himself (higher level), or simply including himself in the social space (lower level) more than when he describes inanimate objects or physical circumstances. This modality has been defined as self in VR.

T13: “Ok...have a look around, then concentrate on your sensations, on your thoughts, ok? How is it? The anxiety?
P13: “Mhh...fine”
T14: “From one to ten?”
P14: “Yes...three, four”
T15: “What are you thinking of?”
P16: “We are too many in a big space...”

This element can be connected to the activation level in the sense that when the patient is immersed and recognizes others’ presence, her or she is inclined to consider and define himself as reference point.

**Is it possible to identify a ‘therapeutic style’? How to use VR stimuli**
A third consideration concerns how different therapists use VR stimuli or reality elements to help the immersion. One key code to help identify behaviors and strategies connected to this aspect has been defined as reality check. It identifies all the passages in which the therapist invites the patient to compare his experience in the real world to the one in the virtual environment. This sort of ‘strategy’ can be used for different purposes: for example to force the immersive experience a bit or, on the contrary, to support the use of the imaginative technique.

P65: “For example, I was on the tube, in Paris...”
T67: “No we are on the tube, now”
P66: “On the tube in Paris I was saying”
T68: “No, we are on this tube, here, now. Let’s try to have a look around for example”

In this case the therapist insists in order to keep the patient immersed in the VE. In other cases the therapist uses some cues of the VE to invite the patient to imagine “what would happen in the real world if...”. The analysis of the interaction suggests that these different strategies are applied by different therapists or by a single therapist in the same session. It is possible to identify a proper ‘style’ when the number of occurrences of one type is considerably frequent for a specific therapist and independently by the patient. It is important to underline that this strategy, even when it can be defined as a style, could depend on the aims of the different therapeutic phases: the need to “bring” the patient in the VE, or, on the other side, the necessity to teach specific relaxing techniques to control the symptoms. Surely further research is needed on this topic.

**Narration as “sense making”**
Different from the initial hypothesis formulated by our researchers about the opportunity to use
narrative elements in order to help the therapist to enrich the virtual experience, it seems that therapists rarely feel the need to create a story to facilitate the immersion. More often, detailed descriptions of the elements already existing in environment and support action towards the patient when he tries to do the same (code defined as *scaffolding action with reference to the immersion*) are enough to “give sense” to the situation and to the immersion in general. Narration and “sense making” can be considered two ideal levels set on a *continuum* that includes different ways to enrich the immersive experience through the co-creation of a credible, or at least possible, situation.

**Conclusions**

In this work we have illustrated a method for analyzing therapist-patient interaction with a VE and with the ‘other’ emerging in VE even when no interaction, for example with other avatars, is foreseen. VR-based sessions can be considered a social context that can be analyzed with psychosocial methods. Information gathered with this kind of analysis, after appropriate studies and further investigations, could help the researchers to become more aware of the dynamics emerging during the therapeutic interaction in a correct *process ecology* perspective. The focus of the research is on those elements more strictly connected to concept of *social presence*, to its multi-componential nature that includes both human and technological elements, literally emerging from the complexity of the interaction and the value it can assume from the therapy efficacy point of view.

Considerations deriving from the study will be used not only to improve VR environments, but also the clinical procedures by formalizing - if necessary - those interaction aspects which are not visible without an in-depth selective approach which takes into account all possible elements and levels of this special therapeutic relation. To mention some examples with reference to the VR environments: 1) *Focus on the social environment.* ‘Social situations’ will be improved and better characterized. The social dimension will have to be immediately perceived by the patient (queues at the supermarket, more people in front of shop-windows); 2) *Self-perception.* The presence of the patient must affect the environment and above all the social environment: for example the patient can have the possibility to push his way through the crowd and see the people change their place.

With reference to the therapeutic protocol: 3) *Therapeutic style.* Each VR environment can be better structured in different zones with different situations to facilitate the therapist in his narration purposes and according to personal styles. Possible narrative paths can be suggested and formally included in the protocol; 4) *Narration as sense-making.* Useful elements that the therapist can autonomously activate and use to focus patients’ attention can be included in the VR environments such as, for example, sudden rumors, prototypical characters (a crying girl, a policeman, a person wearing a strange dress and so on). The therapist could use these cues to build short stories according to clinical purposes, but attention must be paid in order to create a correct balance of elements because they do not have to distract from the therapy. As for point 3, these new possibilities can be formalized in the protocol.

**References**


