

# Display Techniques and Methods for Cross-medial Data Analysis

Luciano Gamberini\*, Anna Spagnolli

Ergonomics and New Technology Labs  
Department of General Psychology  
University of Padova

---

## ABSTRACT

Various kinds of resources (physical, digital, local, far), settings (real and mediated, single or multiuser) and mediating tools are simultaneously active during the interaction with digital environments. In conducting research on human-computer interaction is then vital to work with cross-medial data collections, namely with data which derive from different collection procedures addressing various aspects of the interaction and which are combined according to an overarching methodological rationale. The present paper intends to describe some techniques for the collection and displaying of cross-media data, integrating them with some methodological considerations. Three procedures will be illustrated, namely the split-screen technique, that allows the synchronized visualization of different environments on the same screen; the action indicator augmented display, that allows to enrich the visual recording with signals notifying the occurrence of a particular event; the pentagram, which allows to transcribe multiple sequences of events in their reciprocal temporal relationship. The basic characteristic of these techniques are described and illustratively applied to the interaction with virtual environments.

---

Keywords: *virtual environments, data collection, interaction analysis*

Received 5 June 2003; accepted 8 June 2003.

## 1 The need for cross-medial data collections

Over the last couple of decades, the recourse to multiple devices to collect data on a same phenomenon has spread for technical and theoretical reasons. First, technological and ergonomic improvements have produced versatile, affordable devices with friendlier interfaces, so that high skills in computer science or engineering

---

\* Corresponding Author:

Luciano Gamberini

Ergonomics and New Technology Labs, Department of General Psychology, University of Padova

via Venezia 8, 35131 Padova, Italy

Tel: +39-049-827-6605

Fax: +39-049-827-6600

Email: [luciano.gamberini@unipd.it](mailto:luciano.gamberini@unipd.it)

are no longer needed in order to operate them. Second, a conceptual preference has developed for studying a phenomenon in the context of its actual occurrence and in its natural appearance, so that different levels (from the detailed operations on a device to the norms regulating an activity type), modalities (from gestures to speech) and components (different resources that actor is working with simultaneously) of such a phenomenon need to be recorded. This is particularly true for studies on media usage and human-computer interaction, where participants' action draws on various kinds of resources (physical, digital, local, far), is distributed across different settings (real and mediated, single or multiuser) and operates with different mediating tools. It is then vital to combine multiple means of data gathering and create *cross-medial collections*.

A research using 'cross-medial data collections' requires a good design strategy in order to be reliable. It is for this reason that qualitative methods are looked at with renewed interest; here, the concern with preserving the structure of the phenomenon under study has made customary the combination of different recording techniques, from videorecordings to field notes, from journals to drawings and pictures. In the same vein, qualitative and quantitative data are often combined in order to obtain a more comprehensive analysis (see for example Gamberini et al, 2003), in a mixed quantitative and qualitative research design, which Creswell has distinguished into sequential ('the researcher tries to expand the findings of one method with another method'), concurrent (the researcher converges qualitative and quantitative data in order to provide a comprehensive analysis of the research problem') and transformative ones ('the researcher uses a theoretical lens as an overarching perspective within a design that contains both quantitative and qualitative data') (2003, p.16). Under appropriate methodological conditions, the use of cross-medial collections sets a new standard of accuracy in research. The possibility of inspecting the original patterns of data repeatedly and of sharing them with other scholars increases the transparency and accountability of the analytic process; the access to several aspects of an event helps highlighting phenomena that may otherwise escape our perception.

The present paper intends to describe some techniques for the collection and display of video-data in the field of human-computer interaction, illustratively applied to the interaction with virtual environments. The approach we suggest for the interpretative analysis of video-recorded events is discourse/interaction analysis, centred on the qualitative examination of action sequences (Heath and Hindmarch, 2002; Goodwin, 2000; Jordan, Henderson, 1995), which has sensibly influenced the solutions we elaborated. Quantitative analysis, not addressed in this paper, starts from the sequences of human-interface events, namely from the users' operations on the computer interface that can be collected automatically (Fisher, Sanderson, 1996).

## 2. Some preliminary considerations

Let's demystify two commonsensical beliefs that may plague a cross-medial procedure, namely fidelity and triangulation.

Still or moving images are taken as objective renditions of the events portrayed. This is easily disconfirmed in the experience of people using images in their own research, well aware that each and every shot requires -at least- a perspective, a framing and the exclusion of some features from the picture (Suchman, 1995). Instead of putting this down to the limitations of the recording system, semiologists (Barthes, 1964), media scholars (Evans, Hall, 1999; Berger, 1995) and visual culture researchers (Walker and Chaplin, 1997; Mitchell, 1994) underline that choices are intrinsic to any image; visual representations, even direct visual perceptions, always need cultural practices and pragmatic resources to be made sense of and are therefore 'interpretations'. Consequently, video images do not provide direct glances on bare events, but are necessarily shaped by specific situation and cultural practices that make them meaningful (Latour, Woolgar, 1986). This may be extended to any other kind of rendition that seems to neutralize the intervention of an arbitrary observer, such as the automatic recording of outputs from a computer system or any analogical representation of a phenomenon, acoustical, psychophysical or similar. Collecting data on a phenomenon does not amount to reproduce it objectively, no matter how many sides of it we try to cover or how 'un-mediated' it looks to us.

Another misleading assumption is that *triangulation* among different sources of data, namely the adoption of several methods for data collection (or several sources of data on the same phenomenon or several researcher in the same project), may erase subjectivity and partiality from the data. To be sure, any scientific endeavor needs to come to terms with the issue of subjectivity and try to handle it in some way. However, sociology and philosophy of science reminds us that we cannot defeat subjectivity, but rather increase intersubjectivity and transparency. Right from the start, when the material is prepared for a subsequent analysis, the natural occurrences are 'domesticated' according to methodological conventions: the transcription of a videorecorded interaction, notwithstanding its emphasis on fidelity and accuracy, is actually the first step of an analytical treatment (Ochs, 1979). Endowed with critical awareness, we can go on to the description of three techniques for collecting and displaying cross-medial data.

## 3. Split-screen technique

A rich array of modalities of human-computer interactions are available today: 'real' environments augmented with digital information or reachable via telecommunications;

artificial environments overlapped with physical ones or embedded within them; mediated environments for social or individual navigation. A common characteristic of all these settings, authors start to recognize, is that they are partly digital and partly physical, partly artificial and partly real (for example, Hayles, 1999; Kellerman, 2002; Gamberini and Spagnolli, 2002; Spagnolli and Gamberini 2002). In addition, the action they host may intersect the action on other settings, which the person is simultaneously engaged in (Heath, Luff, 2000).

In particular, immersive virtual environments while placing the user in a three-dimensional virtual scenario depend on real, physical aspects as well. On the one hand we have the “real” body, its movements and the events taking place in the physical room hosting the virtual equipment; on the other hand, we have the “virtual” body, its movements and the events taking place in the virtual media. The simultaneous involvement of the user in both mediated and ‘natural’ environments produces a double source of data for the researcher.

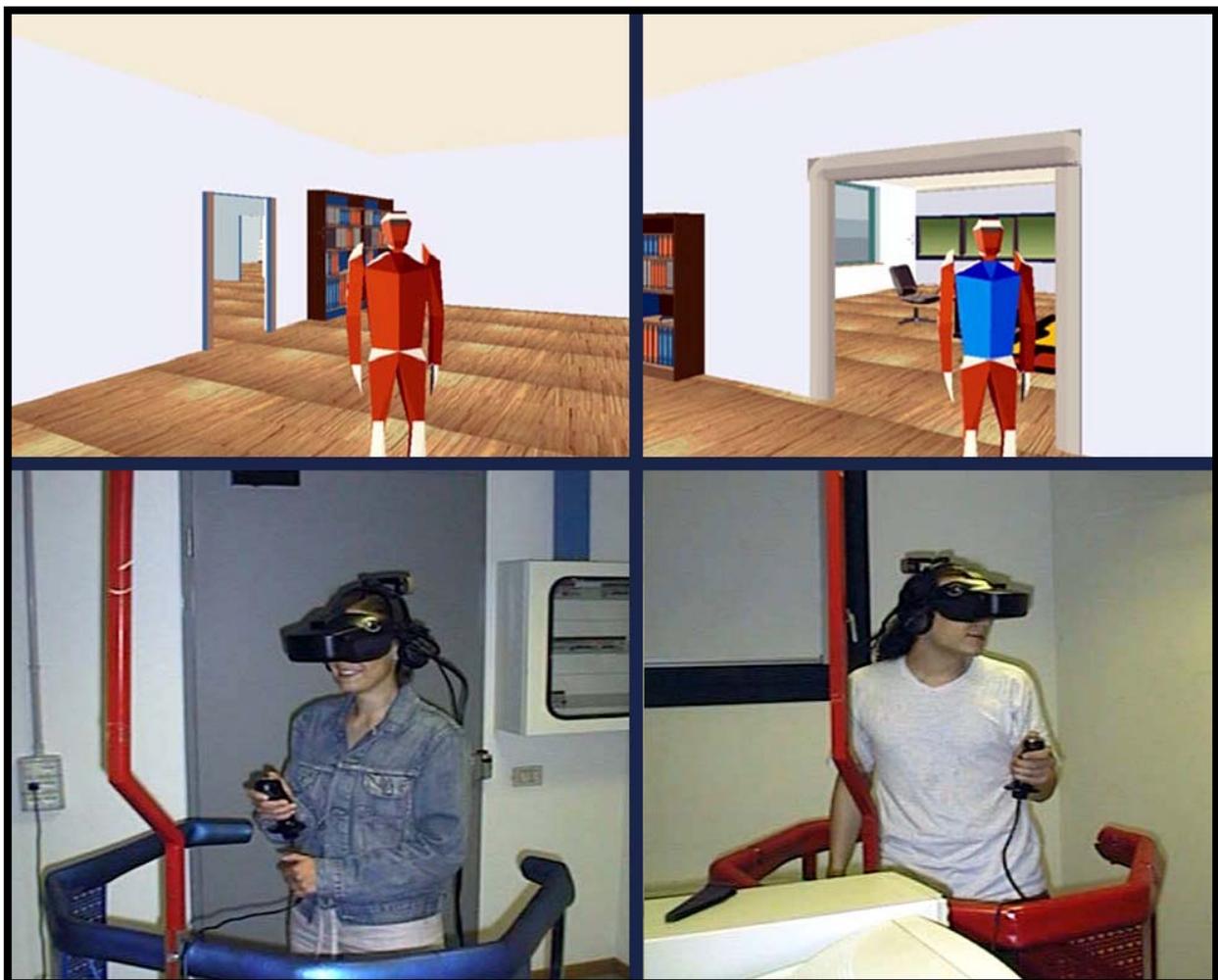
The *split-screen technique* allows to consider the situation in its complexity. In the simplest case, when one user is immersed in a virtual environment, the screen is split into two portions. One half of the screen shows the real environment (figure 1, on the left) with the action performed on the interactive devices (e.g. head mounted display and joystick), gesticulation, talk with other people and so on. The other half of the screen shows what the avatar (or, more generally, the virtual body of the user) is doing in the virtual environment, the feedback received and other events in the simulation.



**Fig.1** The split screen technique applied to a single user during navigation in a VE.

The sense-making process (Rosson & Carrol, 2002; Norman, 1986) in which the user exploits the affordances of the environment to structure his/her action is highlighted by the data offered by this synchronized double videorecording of the events. In fact, we encourage not to treat the events in the real and virtual environments as necessarily separate, but to consider them as components of a hybrid setting hosting a unitary course of action. Users' posture and movements, such as head rotation or joystick manipulation, can be directly analysed in conjunction with the events occurring in the virtual environment to understand the reason why they are produced.

We have referred so far to a single user in an immersive virtual environment, but this technique can obviously be “multiplied” to support the analysis of multi-user environments as illustrated in figure 2 where the split-screen technique is applied to a collaborative virtual environment with two participants (Gamberini et al, 2003).



**Fig. 2** A split-screen with four synchronized images showing two participants in the virtual and the real environment

Obviously, all video sequences come with related, synchronized audio tracks. The digital recording permits to collect some information on the acoustic events, such as their start, length, source and pattern. We suggest to set separate audio channels for different acoustic sources (for example the talk recorded by the microphones in the physical room and the sound effects in the simulation), so as to facilitate their discrimination during the analysis.

#### 4. Action Indicator Augmented Display (AIAD)

The observation of the events during human-computer interaction may be particularly difficult when participants' actions are too fast and overlapped to be detected by watching them. For example, a rapid sequence of actions on the button of a joystick can be difficult to be captured by observing the hand: quick movements may be irremediably lost, and with them the possibility of a fine analysis at this micro level.

With the purpose of facilitating the analyst's work and eliminating gross misinterpretations on what goes on, we used (Spagnoli et al, 2002) a symbolic graphic indicator, called "Action Indicator Augmented Display" (AIAD). This graphical monitoring system is visualized in a corner of the monitor and is activated by a pre-defined set of participants' actions on the interface. A simple version of it is shown in figure 3, where the movement forward, backward, pause and action on virtual objects are indicated.



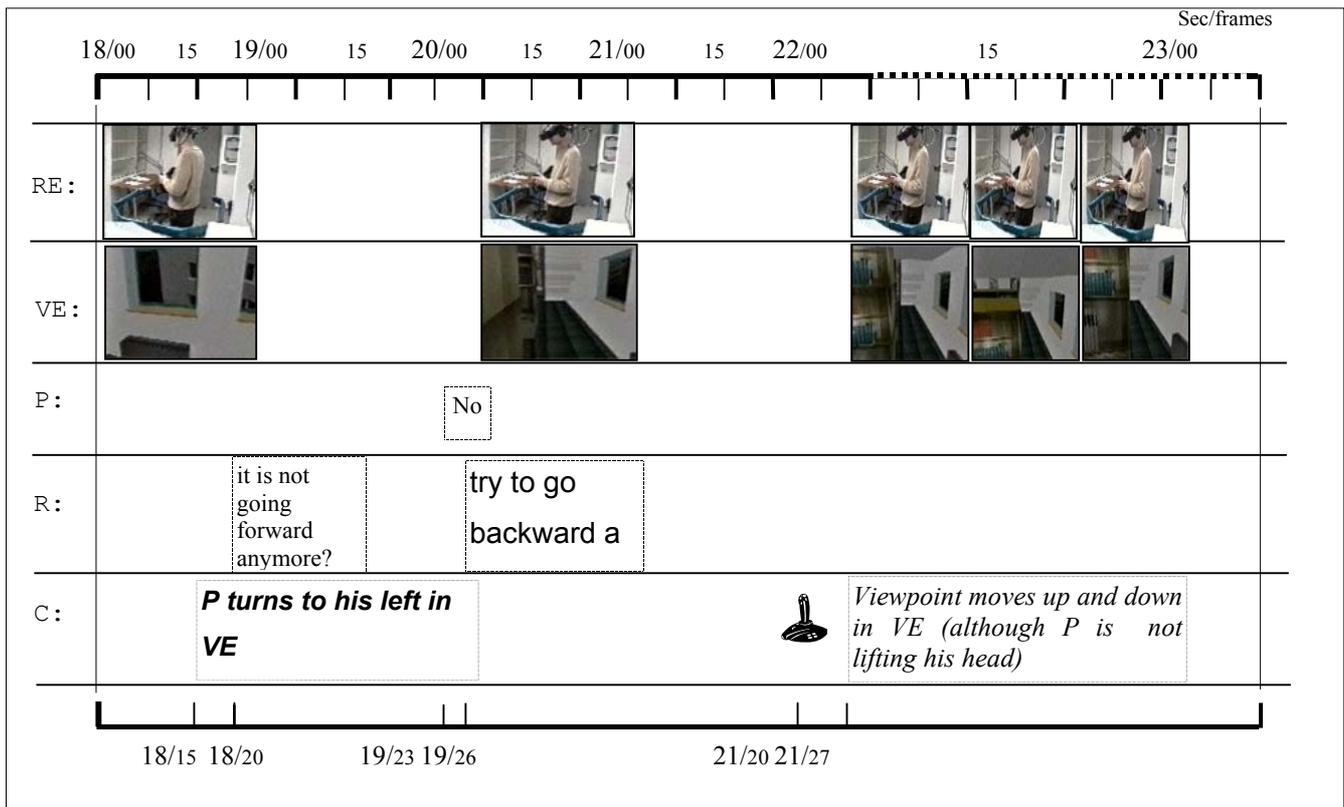
**Figure 3:** An AIAD, at the bottom of the screen, shows an arrow indicating that the button for the movement forward is pressed .

An AIAD can be easily realized by programming a graphical output of any event of interest, such as any avatar's collision against the virtual objects, a head movements, the appearance of particular object in the visual field, etc.: they can be automatically recognized by the program and translated into graphical symbols on the screen. During the interaction each symbol blinks when appropriate, like the arrow in the figure 3 and "augments" the information provided by the images. AIAD output must be synchronized with the flow of events, matched with other automatically recorded data and the overall timeline of the session. Researchers can organize their appropriate AIAD by selecting the events that are most relevant to their study.

### **5. Pentagrams: representing multiple sequences of events**

To analyse simultaneous sequences of events interrelated to each other we adopt a representation rationale called 'pentagram'.

Preliminary, different *kinds* of events are defined, most likely non-verbal events, verbal events, actions on various settings (computer mediated and natural, for example), commentary from the analyst (the commentary is needed because more than one action can be shown in the videoframe). Then the events are put on a dedicated line in the pentagram, positioning them with reference to a timeline on the top of the pentagram. The beginning and completion of each event is measured in seconds and/or frames, obtained from the video-recorder or the digital viewer; the granularity of the timeline can be changed according to the desired level of details.



**Fig. 4.** An example of pentagram for the transcription of cross-medial data; from frame 22 through 23 the granularity of the timelines changes to allow the display of events occurring at a short pace

The novelty of this representation rationale with respect to more conventional transcriptions resides basically in the following aspects:

1. the use of a timeline;
2. the attribution of one line to each kind of events;
3. the equal status attributed to the different kinds of events.

Placing each action on the pentagram with respect to a timeline has many advantages: first, it makes the length and overlap of any event appreciable at a glance; second the space occupied by each action horizontally depends on its actual duration instead of on the verbosity of the description.

When the non verbal events represent the majority of the data, as it is common in human-computer interaction, the use of descriptions, conversation and pictures provides a more fitting representation. The natural organization of the different lines of events is preserved, without privileging the verbal one and inserting any other events

into its architecture, as it is customary in classic transcription techniques. The result is a polyphony of events interplaying with each other.

Measuring each and every action and building the pentagram is however extremely time consuming. The timeline pentagram can be the solution adopted since the very beginning of the transcription or it can be used on selected fragments after a rough transcription of the events has already been outlined, their temporal unfolding and interplay been indicated without precise temporal measures.

## Conclusions

In the previous paragraphs we presented the characteristics of three techniques for gathering and displaying cross-medial data. The peculiarity of these techniques is to combine several methods of data collection to address the complex nature of the phenomena under study. The basic structure of each technique can be adapted to specific research goals, provided that a good balance be found between conflicting needs: on the one hand, to develop customary solutions for particular research project; on the other, to take into account conventions already in use so that other researchers can understand and adopt the solutions we propose.

We purportedly chose to coin the expression *cross-medial collection* instead of using similar ones, such as 'multi-medial' or 'multi-modal'. As to the first one, we intended to underline the necessity not only to adopt several devices (as in 'multi-medial') but also to seek a methodological rationale to connect them to each other. As to the second ('multi-modal'), it distinguishes the different data on the basis of their sensorial and semiotic properties (for example visual versus numerical, gestural versus verbal, etc) (Nigay, Coutaz, 1993), while we wanted to distinguish them on the basis of the capturing procedure.

## References.

- Barthes R. (1964). *The rhetoric of the image: Image, music and text*. London: Fontana.
- Berger J. (1995). *Ways of seeing*. New York: Viking Press.
- Creswell J. W. (2003). *Research design. Qualitative, quantitative and mixed methods approaches*. London: Sage.
- Evans J., Hall S. (1999) (eds). *Visual culture: The reader*. London: Sage.
- Fisher, C., Sanderson, P. (1996). Exploratory sequential data analysis: Exploring continuous observational data. *Interactions*, March, 25-34.

- Gamberini, L., Spagnolli, A. (2002). On the Relationship between Presence and Usability: a Situated, Action-Based Approach to Virtual Environments. In G. Riva, F. Davide (eds) *Being There: Concepts, Effects and Measurement of User Presence in Synthetic Environments*. Amsterdam: IOS Press.
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32, 1489-1522.
- Hayles K. N. (1999). The condition of virtuality. In P. Lunenfeld (Ed), *The digital dialectic. New essays on new media*. Cambridge, MA: The MIT Press.
- Heath C. e Luff P. (2000) *Technology in action*, Cambridge, Cambridge University Press.
- Heath C., Hindmarsh J. (2002) Analysing Interaction: Video, Ethnography and Situated Conduct. In T. May (ed) *Qualitative Research in Action*. London: Sage.
- Kellerman A. (2002). The Internet Earth. A geography of information. Chichester, UK: Wiley and Sons.
- Jordan, B & Henderson, A. (1995). Interaction Analysis: Foundations and practice. *The Journal of the Learning Sciences*, 4(1), 39-103.
- Latour B, Woolgar S. (1986). *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press.
- Mitchell W.J.T. (1994). *Picture theory*. Chicago: The University of Chicago Press.
- Nigay L. & Coutaz J. A design space for multimodal systems: Concurrent processing and Data fusion. Proceedings of INTERCHI'93, ACM Press, pp. 172-178.
- Norris S. (2002). The implication of visual research for discourse analysis: transcription beyond language. *Visual Communication*. 1(1): 97-121.
- Ochs E. (1979). Transcription as theory. In E. Ochs, B.B. Schieffelin (eds), *Developmental Pragmatics*. New York: Academic Press.
- Ochs, E., Schegloff, E.A. e Thompson S. A. (eds) (1996) *Interaction and grammar*. Cambridge, Cambridge University Press.
- Spagnolli A., Gamberini L. (2002). IMMERSION/EMERSION: Presence in hybrid environments. *Fifth Annual International Workshop on Presence*. Porto, 9-11 October.
- Suchman L. (1995). Making work visible. *Communications of the ACM*, 38 (9): 56-64.