KUNGL TEKNISKA HÖGSKOLAN

CID-111 • ISSN 1403-0721 • Department of Numerical Analysis and Computer Science • KTH

Round Table: A Physical Interface for Virtual Camera Deployment in Electronic Arenas

Michael Hoch, Kai-Mikael Jää-Aro, John Bowers

CID, CENTRE FOR USER ORIENTED IT DESIGN

Författare: Michael Hoch, Kai-Mikael Jää-Aro, John Bowers

Round Table: A Physical Interface for Virtual Camera Deployment in Electronic Arenas

Report number: CID-111
ISSN number: ISSN 1403-0721 (print) 1403-073X (Web/PDF)
Publication date: 2001
E-mail of author: mpunkt@zkm.de [kai,bowers]@nada.kth.se
URL of author: http://www.nada.kth.se

Reports can be ordered from:

CID, Centre for User Oriented IT Design
NADA, Deptartment of Numerical Analysis and Computer Science
KTH (Royal Institute of Technology)
SE-100 44 Stockhom, Sweden
Telephone: + 46 (0) 8 790 91 00
Fax: + 46 (0) 8 790 90 99
E-mail: cid@nada.kth.se
URL: http://cid.nada.kth.se

Round Table: A Physical Interface for Virtual Camera Deployment in Electronic Arenas

Michael Hoch

ZKM | Institute for Visual Media,
Karlsruhe, Germany

mpunkt@zkm.de

Kai-Mikael Jää-Aro, John Bowers

Royal Institute of Technology (KTH),
Stockholm, Sweden
[kai,bowers]@nada.kth.se

Abstract.
In this paper, we describe a physical input device for the control of virtual cameras. The so called *RoundTable* has a round projection area where physical icons are used to stipulate the position of virtual cameras. With this scenario we propose a hybrid mixed reality environment for use by production personnel for real-time camera control during a live-broadcast. We present first results of using the RoundTable to support the managing of events in electronic arenas and compare them with traditional interfaces for camera control. We also comment on findings from a scenario in the field of sound mixing and sound composition.

1 Introduction

A number of applications of Virtual Reality (VR) technology exist where multiple cameras are deployed and switched between to take in the virtual environment (VE) from a number of viewpoints. Notably in so-called 'inhabited television' (iTV) applications, where a multi-user VE provides some of the input to a broadcast program [2, 7], virtual cameras may need to be deployed and selected in real-time in a manner similar to the direction of TV. This work takes as its starting point the argument that *finding the action* is one of the core problems in producing events in large-scale participatory settings, especially live events. When many participants are involved and the action can become distributed and is not necessarily governed by a strongly constraining script, it is easy for events of interest to be missed. Accordingly, effort has been devoted to develop a tool which might enable production personnel to gain an overview of the action in an electronic arena and, through this, to support the deployment of resources such as virtual cameras and virtual microphones to pick up happenings of interest. In this paper we describe an approach using a round table as a physical input device to support event management in such electronic arenas via a notion of activity oriented virtual camera deployment and control. That is, it is the activity of participants in an electronic arena which serves as a resource to guide the deployment, direction and control of the views which are made available for broadcast or other forms of dissemination.

It has been an important emphasis of this work that technical systems be proposed and developed which are capable of real-time operation in settings which have a consider-able degree of unpredictability. The application is able to receive and visualize a real-time stream of position, orientation and activity data from an electronic arena. An application, for example, which would require time consuming activity at the interface selecting from hard to access menus or with physical interfaces which could not be used dexterously would be unlikely to be acceptable.

This paper, describes our initial development work of a novel *physical interface* (embedded within a *room-sized environment*) for the control of virtual cameras in applications in electronic arenas. Several usability issues have prompted this proposed solution, including the following.

1. Interaction using conventional desktop input devices such as mice, joysticks and keyboards is often too slow when time-critical selections are required (to move icons or make selections with a mouse requires the user to first grasp the mouse, then make a controlled movement on screen to the target, engage with the target, and then execute the appropriate function).
2. Image direction in the settings of interest to us is a cooperative activity, where multiple users (directors, camera operators and other production members) need to sustain awareness of each other's gestures around shared artifacts—such forms of mutual awareness being very commonly documented as an essential feature of cooperative work in time-critical settings (see, e.g., [9]).
3. Real-world space needs to be recognized and reserved for participants to bring freely whatever real-world documents and other artifacts they wish allowing interaction with these to be interleaved with technically-mediated interaction. These phenomena tend to speak against fully immersive solutions.

2 Mixed Reality / Shared Environments

Several approaches to integrate physical and virtual space in a shared environment have been proposed, for example, DigitalDesk [13], Bricks [6] and phicons [8]. Based on these foundations some applications have been shown to successfully integrate physical interaction handlers and virtual environments or tasks, as in the System BUILD-IT [10], where engineers are supported in designing assembly lines and building plants, or in URP [12] where a physical interface is used for urban planning, or the concept of 'Embodied User Interfaces' [5] where the user physically manipulates a computational device.

In the table environment BUILD-IT described in [10] a menu area is proposed for object selection that, thereafter, can be placed on the virtual floor plan by moving the interaction handler. This approach uses the physical object as a general interaction device. The physical objects that are used in [12] for the urban planning example are mostly used in a less generic but more specific way which lowers the chances of errors due to user input, e.g. a building phicon would less likely be used as something else than a generic brick object. Another approach is reported in [11] where physical objects, the so called 'mediaBlocks', are used as digital containers that allow for physical manipulation outside of the original interaction area.

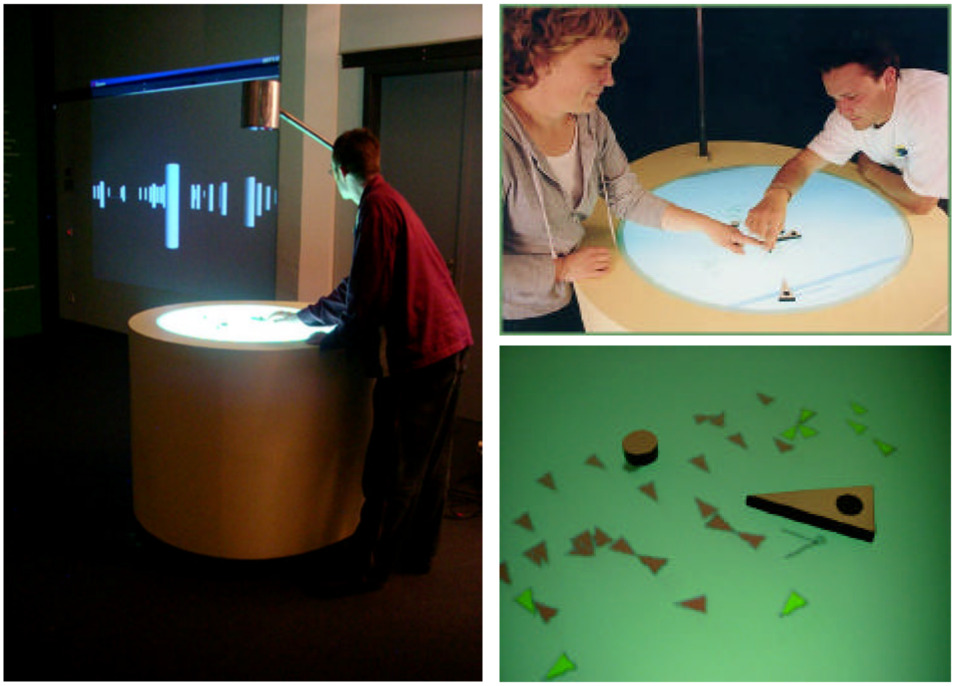
The input devices proposed in this chapter extend these approaches in a number of ways. First, we introduce a context sensitive functionality to the physical objects a user interacts with. That is, the exact significance of an action on a physical object can change in relation to the context in which the action is performed. This enables us to support several different kinds of user action without proliferating the number of phicons which need to be used and identified. Second, we propose a setup that combines physical interaction with abstract visualization in an application that is not concerned with the off-line design of an environment, but real-time intervention in an electronic arena. This combination of physical interface with abstract visualization and real-time consequences of interaction gives our work uniqueness within exploration of physical interfaces. Finally, we emphasize the overall working ecology in which the physical interface we have prototyped is designed to fit. We imagine a room-sized cooperative environment where physical interfaces might enhance and add to traditional interfaces and work activity. This concern for realistic cooperative working environments is rarely emphasized in the design-led demonstrations of physical interfaces and tangible bits which are commonly reported.

3 Round Table with Interaction Blocks

A round table with a projection screen in the middle is used to display a map of the electronic arena (see Figure 1a). The image on the table-screen is rear-projected—that is, projected from underneath the table using a projector and a mirror. The projection screen is approximately 80cm across with a table height of approximately 95cm.

Physical objects are placed upon the table-top projection screen to deploy cameras (figure 1b), select cameras for transmission (TX in broadcasting terminology), enable zooming of the display and the other operations we shall shortly describe. On a second projection screen next to the table (to the rear of Figure 1a), a 3D rendered scene can be displayed from the perspective of the deployed camera. Alternatively, the camera view, as well as the TX view, can be shown on additional monitors in a room-sized environment.

A pole mounted on the table holds a real camera with infrared light. It is used for tracking blocks that can be placed on the table screen (see Figure 1). These can signify the position of virtual cameras by means of positioning the interaction blocks on a representation of the virtual scene. For robust segmentation of the blocks on the projection table, we use retroreflective color attached to each block (made available by 3M, Neuss, Germany) and an infrared filter on the camera to eliminate visible light. In Figure 1c, a visualization of participant activity is shown with participants depicted as shaded triangles together with the, larger triangular, camera phicon and a small circular *probe* phicon, that are used to make selections in the display (indicated by darker, red highlighting).



a) b), c)

Figure 1: *a) Table with rear projection, b) deploying cameras on table surface, c) Top view of table with floor plan, abstract visualization of avatars and interaction blocks*

In the following we will describe two different application scenarios where the RoundTable has been applied to. We also evaluated a traditional interface approach using mouse and/or keyboard.

4 Application Scenario1: Camera Deployment

We have developed a prototype implementation of the principles we have described above, called SVEA (Sonification and Visualization for Electronic Arenas). SVEA implements the Spatial Model described in [1], it computes and displays awareness maps, and the support for simple mouse-based interaction. We place isosceles triangle shaped markers representing the participants of the electronic arena on a 2D projection of space with the color of these markers displaying a measure of the activity they show. Areas of high activity will thus be conspicuous as large, brightly-colored areas.

Once a new camera phicon is detected on the RoundTable, a virtual camera is assigned to the location indicated and pointing in the direction suggested. The view onto the electronic arena from this virtual camera can be selected for transmission (TX) by placing a small round *camera selector* phicon into the non-reflective hole in the middle of the triangle. Using the camera selector phicon is proposed as an intuitive and simple way for selecting cameras. The *probe* phicon is used to select a group of avatars in the projected visualization. Rather than attempt—using phicons—to replicate the mouse gesture of clicking and dragging used to select groups in the screen-based application, we decided to exploit and extend the awareness model underlying the visualization to enable context-sensitive selections to be made [4], i.e. exactly which avatars it selects, how many and in which configuration, is dependent upon the avatars' orientations and proximity with respect to the probe. We describe this use of the probe phicon as being 'context-sensitive' because exactly which avatars it selects, how many and in which configuration, is dependent upon the avatars' orientations and proximity with respect to the probe. Finally, the *zoom* phicon allows one to zoom into the visualization in a similar context-sensitive fashion. The group of avatars is determined corresponding to those which an object placed in the electronic arena at the location corresponding to the zoom phicon would be aware of. The display is rescaled so that it shows this group plus an area around them.

In the desktop version of SVEA, algorithmic camera deployment is actioned by dragging the mouse over the display to select a set of markers (i.e. participants). As soon as the mouse is released a camera is deployed to the algorithmically computed optimal location for that group according to whichever algorithm is set as a preference. In this, algorithmically enhanced way, camera deployment can be efficiently actioned by a single interface gesture. Also in in the desktop version of SVEA, the 2D display can be magnified up to 16 times by selection of a menu option to zoom in on a selected group of interest with the center of that group being the center of the zoom. Zooming rescales the relative separation of markers in screen distances but not the size of the triangles themselves. Zooming, therefore, can clarify the relative positions of participants that are so close to each other as to appear overlapping when in a 'wide angle' view.

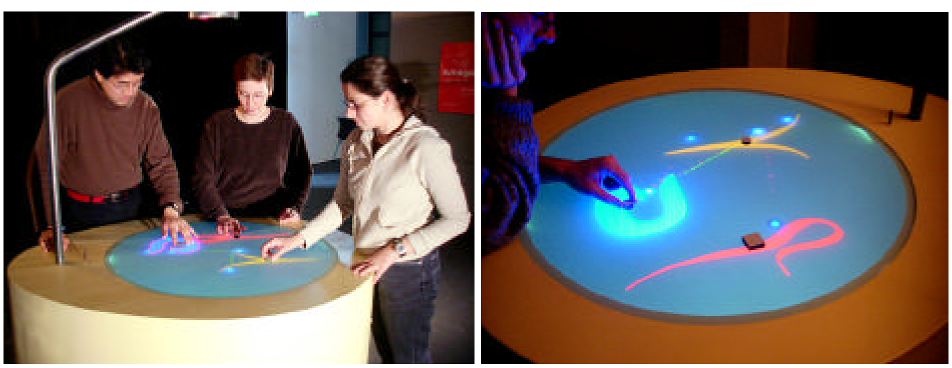
In the RoundTable production environment, we have had to make important design decisions over issues to do with the relationship between inserting a new camera phicon into the vision system's field of view and the 'lifecycle' of a corresponding virtual camera. Does deploying a phicon create a new camera at the designated location? Or does it merely re-deploy an available camera? Are there as many camera phicons as virtual cameras (and no more)? Or can virtual cameras be created without upper limit? Does the removal of a camera phicon cause the corresponding virtual camera to pass out of existence? Or does it cause the virtual camera to return to some default behavior? Ultimately, we feel that such questions have to be answered with respect to particular applications.

Our prototype, then, arbitrarily restricts the number of cameras to a user-preferred limit but does so without prejudice to alternative possibilities. Within this set of cameras, the default behavior is an autonomous one. That is, if a camera is not deployed through activity at the round table, it maintains behavior which is entirely algorithmically determined. This default behavior is to rove the space, following the gradient of increasing awareness, while avoiding other cameras. The introduction of a camera phicon will 'claim' the first available camera. Which camera is 'first' is determined in numerical order—thus retaining a sense of 'Camera 1', 'Camera 2' and so forth which could be explicitly referred to by users to individuate cameras. The assignment will pass over already assigned cameras. Amongst other desirable features, this method has the consequence that cameras already selected for TX will not be suddenly and mistakenly cut to another location.

Our introduction of a phicon controlled zooming functionality has some important consequences. After zooming, the projection of the virtual space will be transformed such that the position of a camera phicon may no longer correspond to the virtual camera previously associated with it. There are a number of design options here. Zooming could (i) forcibly deassign virtual cameras from phicons or (ii) the phicon and the associated camera become disjoint in their spatial location on the table or (iii) virtual cameras 'snap' to the new relative locations occupied by their phicons. In discussion with users, each of these options have been revealed as having benefits and drawbacks as general solutions, so again ultimately the choice should be made on application-specific grounds.

5 Application Scenario2: Interactive Compositional Machines

For this application we ported one piece of the CD-ROM "Small Fish" that has been created by Kiyoshi Furukawa, Masaki Fujihata and Wolfgang Münch at the ZKM in Karlsruhe to the RoundTable. Small Fish allows the user to manipulate a predefined set of musical algorithms, that stipulate what kind of relationships between visual graphics and musical time exist and how the player is able to influence the sounds via the screen. As a CD-ROM, Small Fish uses a traditional interface with mouse. For porting the application to the RoundTable we modified the mouse interface to actually get driven by the table output. We allowed three phicons, which did not get differentiated in shape, to stipulate the position of three different graphical objects. Hence, allowing three users to simultaneously interact with the system (see Figure 2a).



a) b)

Figure 2: RoundTable with Sound Mixing Application "Small Fish", a) three users interacting simultaneously, b) detail view of the graphical objects

By modifying the position of the phicons on the table surface, the graphical objects position gets stipulated (see figure 2b). Music is generated by producing a Midi stream of notes that gets influenced by the position of smaller circular objects that stick to the larger graphical objects (see small blue points in figure 2b). By moving the larger objects, the user can "catch" or remove the small objects from other objects. Another two small circular objects are special, they move around autonomously, touching the smaller objects in a predefined order. Whenever they touch a small (blue) objects, a Midi note is produced. Tone and the note itself is stipulated by the y-position of the "collision" (i.e. from lower left top upper right in figure 2b). Hence, notes and speed can be stipulated by the user by modifying the position of the objects in y-direction and spatially arranging the objects on the 2D display surface.

Our first results of evaluating the Small Fish Interactive Compositional Machine by showing this application to some ‘players’ were really positively. All users liked the tangible interface to the application. We noted that users did interact in an immediate and intuitive way. The interface allowed a single user to most easily change the position of the graphical objects at different spatial locations on the table. There is no more a point and click operation needed, no positioning of one interface handler to different graphical objects, but more a direct manipulation of graphical objects with the aid of the physical objects on the table surface. Even when more than three people where present, we noted that interaction did switch between users, i.e. showing collaborative behavior. Also users reported it being “fun” to interact with other users simultaneously.

6 Discussion and Results

The applications described in this paper have been trialed upon the RoundTable as well as in more conventional desktop variants. Using the RoundTable indicates how mixed reality technologies (here mixing the manipulation of physical objects with projected computer generated displays) can be deployed in support of production activities in an electronic arena. Our work has been marked by a focused concern to reflect upon its nature and evaluate its advantages and problems throughout design. Where possible we have put our technologies before both members of the public and media professionals. We believe that our emphasis on combining physical interfaces and virtual displays within a shared environment designed to support cooperative work is novel, and that physical interfaces for virtual camera control in electronic arenas is a unique application area.

Currently, we have built a functional prototype for deploying virtual cameras that shows the applicability of our approach. We introduced a novel context-sensitive use of physical objects and presented novel selection and zoom operations using this technique. We have presented the RoundTable to a number of persons independent of the development group so as to evaluate it as a means for presenting production support tools. Users of the RoundTable have included a number of media professionals as well as students and visitors to our lab. Most importantly, several production personnel who were practically involved in the inhabited TV events have been presented with the table and the concepts it embodies. These evaluation sessions have been informal and discussive, rather than based around controlled experimentation. This is appropriate given the kind of technology it is and the prototype standing of the applications we have developed for it.

The main advantages with using the RoundTable that have repeatedly come to light are twofold:

1) Giving both a position and an orientation using a physical block is noticeably faster than using a mouse, since mouse-based positioning requires one operation for indicating the position and a second for indicating the direction – this latter requires dragging and thus takes longer time than just placing an item. (The mouse operations can be combined into one physical movement, so that the position at mouse-down is taken to be the origin of the camera and the position at mouse-up to indicate the end of a direction vector. However, it is very hard to do this with the same speed and precision as with a phicon.)
2) Several people can stand around the table and easily do adjustments or new placements of cameras, thus supporting mutual awareness among the production personnel as they engage with a shared artifact. Conventional screen-based, desktop solutions are more cumbersome in this regard.

Problems we encountered concern the way users touch the objects, i.e. when users grasp objects from above they often partly (or fully) occlude the objects from the camera. This can make the tracking erroneous. However, once the user gets instructed how to handle the objects we do not encounter this as a major problem. Indeed, we have noticed users spontaneously discover that covering an interaction object is a gesture that can be creatively used (e.g. to simulate the momentary removal of an object). Again, if such occlusion problems become excessive, it is possible to consider siting the camera beneath the table. However, this solution can lead to misunderstandings if users think that triangular phicons always represent the presence of a camera in the electronic arena. For our method of camera allocation, this would be an inappropriate 'user-model'. Rather, the camera phicons should be regarded as *tools with which to deploy virtual camera*s, not representations of those cameras themselves. It is clear, though, that this is a less natural model than expecting a one to one correspondence between virtual camera and camera phicon.

It is important to observe that similar issues arise for most attempts to physically interface to the virtual and are not specific to our application or our particular design decisions. Only in the extreme case of a completely strict coupling between physical activity and consequences in the virtual world *(and vice vers*a) would it be possible to think that a physical object could non-problematically represent a virtual one. As soon as the coupling is relaxed the relationship between the physical and the virtual has to be *achieved in users' practical understandings* rather than technically mandated (see [3]). In a sense, this argument presents a limit case on the use of tangible mixed reality interfaces. When using simple 'inert' physical objects like blocks, many simple user-interface operations will cause problems with the significance of the physical objects. Rescaling and scrolling a display will make the objects disjoint from anything they are to represent.

In the case of a production tool intended to be shared by personnel, with individuals possibly attending to different parts of the display at the same, there may be good arguments for disallowing operations such as zooming and scrolling. Changing the viewpoint in the display places fairly strict restrictions on the type of collaboration between production personnel as any zoom or scroll would potentially move an individual's work area out of sight. We conclude therefore that our concept of 'context sensitive zooming' supported by a phicon, while of some novelty (we know of no other work which has such an elegant, single phicon solution to such a fundamental interaction issue in tangible interfaces), is probably not of great practical utility in our target domain (production work for electronic arenas). In our domain, on the occasions when zooming etc are required, conventional desktop solutions seem more appropriate.

One of the directors of the inhabited TV events envisaged problems in using a device like the RoundTable because it suggested to her that she would have to disengage with inspecting view monitors (both of TX and sources) in order to use it. To be sure, in conventional TV practice, a director will be continually attending to her sources and to TX with, if she is also doing vision mixing, fingers poised on pre-programmed buttons to bring about transitions between sources. SVEA on the RoundTable is a very different setting from this. In our current design, the selection of a camera requires the placing of a small peg into a camera phicon. This is not a faster gesture than pressing a button ones finger is already poised over.

A technology like ours could still be used by non-directors. Indeed, typically in a large-scale distributed live real-world event, a director will be assisted by several assistants who submix or pre-select views for her. A RoundTable production suite could well have a specific role in browsing and homing in on action in electronic setting analogous to that one. Selecting a camera with the peg would not necessarily select TX but would sub-edit a shot for directorial consideration.

7 Conclusion

Our initial motivation for the physical interaction solution that the RoundTable embodies was based on the claim (often articulated in the literature) that interaction with phicons is commonly faster than with conventional desktop widgets. This claim is critically interrogated since we were able to compare RoundTable with conventional workstation implementations of essentially the same applications. Some operations are noticeably quicker, but others seem slowed. This again enables us to know the limitations of the RoundTable as a solution. Our overall view is that a tangible physical interface (like the RoundTable) enables some interaction capabilities while inhibiting others. And the converse is true for conventional desktop interfaces. The RoundTable enables some kinds of gestures to be made much more swiftly than others (e.g. the simultaneous specification of an orientation and a position). Other gestures may be slowed (e.g. those which require an object to be visually searched for on the table before an action, like selection with a peg, can be initiated). The RoundTable permits multiple viewers to experience its information visualization more readily than a screen on a personal workstation. This facilitates certain patterns of collaboration and inter-working between personnel.

Commands can be actioned from the keyboard, by mouse-gesture, by menu selection and so forth. On the RoundTable, we have tried to avoid simulating or reproducing such widgets, preferring a consistent physical interaction approach throughout. Accordingly, we have a limited number of blocks with basic functionality. This works well. Our attempts to go beyond this and trying to support classic operations such as zooming and scrolling have introduced tensions and inconsistencies in our design approach and given users difficulties in finding the appropriate ‘model’ for understanding what's going on. In short, the RoundTable is an elegant interface for applications that need a small repertoire of commands to be supported.

Our overall appraisal of the RoundTable, then, is nuanced. In contrast to much of the HCI literature at the time of writing, we do not think that physical-tangible interaction methods offer categorical advantages over conventional desktop-workstation-based solutions. At least, not so in all contexts. Our experience has enabled us to articulate the boundary conditions on the applicability of such interaction techniques. We conclude that gestural interaction in relation to an interaction surface is an idiomatic way for supporting participants' engagement within an electronic arena.

8 REFERENCES

1. Benford S, Fahlén L (1993), *Awareness, Focus and Aura, a spatial model of interaction in virtual space*, Proceedings of HCI International `93, Orlando, August 1993.
2. Benford S, Greenhalgh C, Brown C, Walker G, Regan T, Morphett J, Wyver J and Rea P (1998), *Experiments in inhabited TV*, Proceedings of CHI'98, pp 289–290.
3. Bowers J, Pycock J, O'Brien J (1996), *Talk and Embodiment in Collaborative Virtual Environments*, Proceedings of CHI'96, pp 58-65.
4. Bowers J, Hellström S-O and Jää-Aro K-M (1999), *Issues in management of Inhabited TV broadcasts,* eRENA Deliverable 4.3. 1999.
5. Fishkin K, Moran T, Harrison B (1998), *Embodied User Interfaces: Towards Invisible User Interfaces*, Proc. EHCI'98 Conference on Engineering for Human-Computer Interaction, Crete Greece, 1998
6. Fitzmaurice G, Ishii H, Buxton W (1995), *Bricks: Laying the Foundations for Graspable User Interfaces*, ACM Proceedings CHI'95, Denver, Colorado, 1995.
7. Greenhalgh C, Benford S, Tylor I, Bowers J et. al. (1999), *Creating a Live Broadcast from a Virtual Environment*, SIGGRAPH '99, Conference Proceedings, Los Angeles, 1999.
8. Ishii H, Ullmer B (1997), *Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms*, ACM Proceedings of CHI'97, pp. 234-241.
9. Martin D, Bowers J, Wastell D. (1997), The interactional affordances of technology: an ethnography of human-computer interaction in an ambulance control centre. Proceedings of HCI'97.
10. Rauterberg M, Bichsel M, Fjeld M et. al. (1998), *BUILD-IT: A Planning Tool for Construction and Design*, Video Proc. + Summary CHI'98.
11. Ullmer B, Ishii H. Glas D (1998), *mediaBlocks: Physical Containers, Transport, and Controls for Online Media, Proc. SIGGRAPH'98, ACM 1998.*
12. Underkoffler J, Ishii H (1999), *Urp: A Luminous-Tangible Workbench for Urban Planning and Design*, Proc. CHI'99.
13. Wellner P (1993), *Interacting with Paper on the* DigitalDesk, CACM, Vol. 36, No. 7, July 1993, pp 87-96.