Design of Participatory Virtual Reality System for Visualising an Intelligent Adaptive Cyberspace

N. SESHAGIRI*

Abstract

The concept of 'Virtual Intelligence' is proposed as an intelligent adaptive interaction between the simulated 3-D dynamic environment and the 3-D dynamic virtual image of the participant in the cyberspace created by a virtual reality system. A system design for such interaction is realised utilising only a stereoscopic optical head-mounted LCD display with an ultrasonic head tracker, a pair of gesture-controlled fibre optic gloves and, a speech recognition and synthesiser device, which are all connected to a Pentium computer.

A 3-D dynamic environment is created by physically-based modelling and rendering in real-time and modification of existing object description files by a fractals-based Morph software. It is supported by an extensive library of audio and video functions, and functions characterising the dynamics of various objects. The multimedia database files so created are retrieved or manipulated by intelligent hypermedia navigation and intelligent integration with existing information. Speech commands control the dynamics of the environment and the corresponding multimedia databases.

The concept of a virtual camera developed by Zelter as well as Thalmann and Thalmann, as automated by Noma and Okada, can be applied for dynamically relating the orientation and actions of the virtual image of the participant with respect to the simulated environment. Utilising the fibre optic gloves, gesture-based commands are given by the participant for controlling his 3-D virtual image using a gesture language. Optimal estimation methods and data flow techniques enable synchronisation between the commands of the participant expressed through the gesture language and his 3-D dynamic virtual image. Utilising a framework, developed earlier by the author, for adaptive computational control of distributed multimedia systems, the data access required for the environment as well as the virtual image of the participant can be endowed with adaptive capability.

**National Informatics Centre, A-Block, C.G.O. Complex, Lodhi Road, New Delhi-110 003.*
Journal of Information Science and Technology, Vol. 3, No. 3, April 1994.
ISSN 0971-1988

1 Introduction

Virtual Reality System in which a participant interacts with computer-generated objects, in an artificial fully-immersive environment, is a visionary technology which was forecast more than a quarter century ago by Sutherland [18, 19]. Virtual Reality System, a name coined by Jerone Lanier [10], emerged from the earlier concept of a visually-coupled system as a computer-simulated environment that is continuously updated with respect to the head position. Since 1989, persistent demands have been made for developing realistic space and time-based multimedia solutions for certain environments into which participants can 'enter' and virtually participate in the events of the animated environment. The interactive virtual environment can be a computer simulation of a real world situation or an abstract form of a real world event or an abstract world. In the limiting case, the animated world can display synthetic cues which would be indistinguishable from those displayed in the real world. Attempts are being made to render virtual objects look, sound and feel like the real objects that they represent and display dynamical and behavioural patterns obeying the laws of the real world like Newtonian mechanics. Though this ultimate goal is beyond contemporary technology, the increasingly emerging gadgetry prodded by new developments in computer hardware, software, control systems and transducer technologies, among others, are giving an increasing level of realism to the virtual environment. The emerging commercialisation of virtual reality systems is indicative of the strides made in the past four years since it became an organised technology.

Hendersen [7] introduced virtual reality through the concept of cyberspace. Our experience in interactive multimedia simulation is applicable for designing virtual reality. Similarly, work in multimedia database is applicable for designing the cyberspace environment. In this sense, virtual reality is a step in the technological evolution of interactive multimedia.

The approach adopted by the author in creating the Virtual Environment Laboratory of National Informatics Centre (VELNIC) is to conceptualise, explore, adopt, modify and develop the approaches to virtual reality technology which tend to minimise the system cost while developing applications which are relevant to the problems faced by countries like India in areas such as training in specialised subjects and national defence. A complete schematic diagram of VELNIC is given in Fig. 1 which is being realised with an investment less than 80,000 US dollars through the introduction of concepts and techniques which tend to minimise the system cost. An array of input and output devices, each device serving a sensory channel, is considered. A conscious effort is made to minimise the tactile components through the introduction of such concepts as 'virtual participant', 'virtual camera' and 'virtual intelligence' Fibre optic gloves can be used for the recognition of hand gestures and orientation of fingers on each hand. Though costing considerably less, the VELNIC virtual environment system is highly integrated through maximally substituting hardware by software by the creation of virtual hardware like the virtual camera in place of the real one. VELNIC has the following components at present:

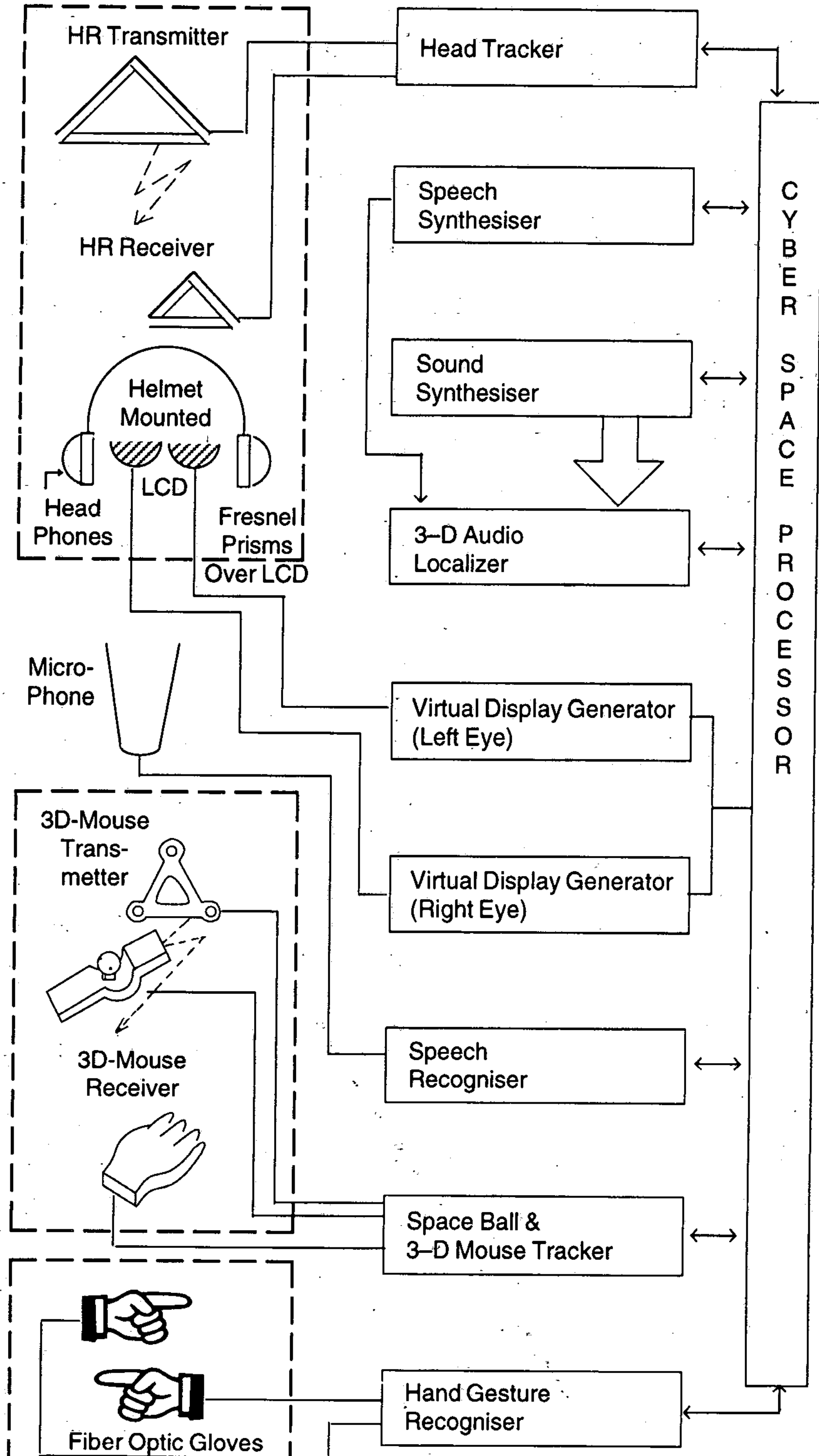


Fig. 1 *Components of the Virtual Environment Laboratory of the National Informatics Centre (VELNIC)*

Visual Components

* A head-mounted LCD display with Fresnel prisms
* Head tracking system based on ultrasonic triangulation
* Flying Mouse along with 3-D tracking system
* SpaceBall device

Auditory Components

* Digital sound synthesis system
* 3-D audio localisation system
* Speech sensory system
* Speech recognition system

Tactile Components

* Low cost fibre optic gloves for hand gesture recognition

Interface Unit and Development Tools

* MultiGen Genlock Adaptor
* Audio-Video (n × m) distributor
* Software like MediaDB, VISTAPRO 3.0, VISTAMORPH, CHIPS WINVIDEO, Cyberspace Kit, KNOWLEDGEPRO, KNOWLEDGE-MAKER, Autodesk 3d STUDIO, SMILE, ANIMATOR, MATHEMATICA, etc.

A Pentium Computer with 32 MB main memory and 2.4 GB modules of disk are used under DOS/WINDOW-NT platform.

In this paper, Section-II describes the simulation of the 3-D dynamic environment created by a physically-based modelling and rendering in real time. The media DB multimedia database software is utilised along with appropriate animation and fractals-based Morph software for creating the virtual environment. An extensive library of audio and video functions and clippings, as well as primitives characterising the dynamics of various objects, support the simulation and rendering process.

Section-III describes how the multimedia database files so created are retrieved or manipulated by intelligent hypermedia navigation and integration with the existing information. Speech commands can be used for controlling the dynamics of the virtual environment and the corresponding multimedia database.

The participant can have a porthole view of cyberspace with full immersive effect through a head-mounted LCD display with a pair of Fresnel prisms and a head tracking system based on ultrasonic triangulation. In Section-IV, a low cost and yet highly immersive system is described based on a mathematical treatment of the stereoscopic effect underlying the use of the Fresnel prism actuated LCD. Other auxiliary tracking equipment like the Flying Mouse and SpaceBall are also described.

For situating the virtual participant in the virtual environment, it is necessary to create as realistic a 3-D image of the participant as possible. Whereas, well-known software packages which can construct a 3-D image of an object, given the front view, side views, back view and the top view of the participant, or by morphing an idealised construct, are commercially available, the portrayal of the head (including face) of the virtual participant in a form recognisable as a true image of the physical participant, is a much more exacting job. As described in Section-V, this can, however, be realised by the use of the wire-frame technique and synchronisation of the front view, side views, back view and the top view through an automated virtual camera.

The movements of the virtual image like walking, running, sitting, standing and turning, are facilitated by geometrising and quantifying the degrees of freedom at various joints in the human anatomy. Instead of the participant actually feeling tactile sensations of the hands, legs and head for enabling him to feel the actions performed by his virtual image, our proposal is to keep the participants substantially static and develop methods for him to mentally feel the actions of his image through a psychological adjustment to the mental perceptions of involuntary movements conveyed purely as an information to the brain. The psychological basis of the present theory has a parallel in the process of speech formation. When we speak, we are not consciously aware of the movements of the tongue, lips and other facial muscles. The will to express something, brings out the sound while the involuntary muscular movements of the face, lips and the tongue automatically fall into the preset patterns. In the same way, the psychological adjustments take place when gesture-based commands are given by the participant for controlling his virtual image using a sign language delivered through preset gestures made by the fingers through the fibre optic gloves.

To create a powerful illusion, the costlier VR systems enable the computer to sense the location and actions of the participant's body in space so as to accurately represent the participant's virtual body and, turn specific body movements and actions into commands for the computer. To give a high degree of realism and a highly immersive virtual environment, the multimillion dollar systems use sophisticated and cumbersome tread mills as well as exoskeletal attachments to the body with tactile sensory feedback channels and a fibre optic body-suit. As described in Section-VI, a low-cost alternative is being developed in which optimal estimation and prediction methods based on the Kalman Filter theory and Data Flow Techniques are used for the synchronisation between the commands of the participants expressed through a finger gesture language via an optical fibre glove and the action performed by his 3-D dynamic virtual image. Conversely, the will of the participant to effect certain actions of his virtual image, would seem to act directly without being consciously aware of the gesture language expressed through the fingures, if the participant has practised the same adequately. Experiments with the VELNIC set up have shown that there exists a tendency for the mind of the participant to directly attach itself to his virtual image.

Utilising a framework developed earlier by the author [16] for adaptive 'computational' control of distributed multimedia systems, virtual environment as well as the virtual image of the participant can be endowed with adaptive capability.

Within the ambit of the above described system, a concept of 'virtual intelligence' is proposed in the last section as an intelligent adaptive interaction between the virtual environment and the virtual image of the participant in cyberspace.

2 Simulation of The 3-D Dynamic Virtual Environment

The software for the animation of the 3-D dynamic virtual environment is based on a physical-based modelling and rendering in real time. It consists of a unified protocol for object/scene description rendering and manipulation along with enabling hardware for lighting and texturing. It embodies lights, light models, colour sub-primitives and drawing sub-primitives along with administration token for file maintenance and readability. A rendering of an object can be carried out by pre-render parsing of the ASCII file into dynamically-allocated structure of object definition opcodes, pre-render definition of lights and light models and traversing opcode list, drawing only the graphic primitives and selecting the currently active light, light models or colour definitions. There are features such as real time animation of a legion of 3-D Icons, graphical editor for specifying articulated motion control characteristics of the object, and features which give the object physical characteristics and mechanisms to govern the object motion given the list of known internal and external forces acting on the object.

Endowing the object with physical characteristics requires the use of the dynamics of a rigid body and careful reconciliation between geometric and algorithmic complexities (See Wilhelms) [22]. It is emphasised that the use of the framework to simulate Newtonian mechanics in real time is essential for most forms of animation dynamics and should be set in a manner relevant to the modelling of the physical interaction of objects as allows:

$$[p_{x_0} \cdot p_{y_0} \cdot p_{z_0}] \Rightarrow \begin{matrix}\text{Current}\\ \text{Position}\end{matrix}; \quad [\theta_{x_0} \cdot \theta_{y_0} \cdot \theta_{z_0}] \Rightarrow \begin{matrix}\text{Current}\\ \text{Orientation}\end{matrix}$$

$$\left[\frac{\delta}{\delta t}(p_{x_0}) \cdot \frac{\delta}{\delta t}(p_{y_0}) \cdot \frac{\delta}{\delta t}(p_{z_0})\right] \Rightarrow \begin{matrix}\text{Linear}\\ \text{Velocity}\end{matrix};$$

$$\left[\frac{\delta}{\delta t}(\theta_{x_0}) \cdot \frac{\delta}{\delta t}(\theta_{y_0}) \cdot \frac{\delta}{\delta t}(\theta_{z_0})\right] \Rightarrow \begin{matrix}\text{Angular}\\ \text{Velocity}\end{matrix}$$

$$\left[\frac{\delta^2}{\delta^2 t}(p_{x_0}) \cdot \frac{\delta^2}{\delta^2 t}(p_{y_0}) \cdot \frac{\delta^2}{\delta^2 t}(p_{z_0})\right] \Rightarrow \begin{matrix}\text{Linear}\\ \text{Acceleration}\end{matrix};$$

$$\left[\frac{\delta^2}{\delta^2 t}(\theta_{x_0}), \frac{\delta^2}{\delta^2 t}(\theta_{x_0}), \frac{\delta^2}{\delta^2 t}(\theta_{x_0})\right] \Rightarrow \begin{matrix}\text{Angular}\\ \text{Acceleration}\end{matrix}$$

If m is the mass of the object, then (F_{xyz}, T_{xyz}) which is the net (force, torque) directed at the centre of mass of the object is obtained by multiplying the corresponding acceleration by the mass. A modified Euler integration attributed to Spiegal [7] is found to be ideal for animation dynamics. This leads to the following final linear velocity, given as:

$$\frac{\delta}{\delta t}(Fp_{xyz}) = \frac{\delta}{\delta t}(Ip_{xyz}) + \left\{\frac{\delta^2}{\delta^2 t}(Ip_{xyz}) \times \delta_t\right\}$$

and the final linear positions, given current positions/velocities and predicted velocity averages at sample time.

$$Fp_{xyz} = Ip_{xyz} + \left(\frac{\delta}{\delta t} (Fp_{xyz}) \times \delta t \right) + \left\{ 0.5 \times \frac{\delta^2}{\delta^2 t} (p_{xyz}) \times (\delta t)^2 \right\}$$

Similarly, the final angular positions are obtained. Here Fp_{xyz} is the final position component and Ip_{xyz} is the initial position component and δt is the time interval since the previous integration. These are the basic equations for animation dynamics during the 3-D simulation.

A deforming force influences the object in various ways [24]. Each polygon in the object has an associated breake and bend threshold token specified. Using the relationship that a force dissipates its kinetic energy inversely over the square of the distance from the force origin to the polygon, the value of the dissipated force per unit polygon surface area is calculated. If the force is strong enough to break the polygon, the original polygon token is removed from the object token list and replaced with a list of smaller triangular polygonal shard tokens. Triangles are used to ensure planar polygon. The shards are initially determined by snipping off the corners of a multisided convex polygon, thus spiralling inward until the remaining quandrangle is divided into two. The remaining shards are broken along the hypotenuse as required. If the force is only strong enough to bend the polygon, the polygon token is scavenged from the object token list and replaced with a new flexible polygon that tracks the impact of a moving point of bending force. The bending force is modelled using Hooke's Law and a spherical spring. If the force is neither strong enough to break or bend a polygon, then it is portrayed as only pushing a polygonal shard.

We can provide a structured mechanism for controlling the behaviour of the object by allowing varying degrees of involvement of the user and the designer with the simulation. To the end user, the prime objective is to achieve maximum realism. An art design team creates a specific object hierarchy with a realistic shape, colour and orientation data. A technical design team would then add the physical attributes of the sub-objects like mass, centre of mass and elasticity, and the impinging force description like the force position or point of impingement in the object frame of reference, force direction unit vector, description of the force, etc. The analytical design team would then specify mappings between sub-object movements and the force affected by such movements. The participant would then be able to control a given object in a realistic manner with realistic results by manipulating a set of control sub-objects linked to local forces. This would give an adjustable focus in specifying high-level object motion in a range of control modes—either directly or indirectly through local force control or more indirectly with sub-object control. A suite of software tools enable a user to rapidly design and test an object's physical characteristics.

In the creation of virtual reality environments, we strongly recommend the use of physically-based modelling even though it may be difficult at present as an adequate set of software tools is yet to be developed for making the modelling efforts simpler. An action control layer will be required to be introduced as the generation of the mapping function matrices could be achieved easily by selecting the object/pair and taking snap shots of a series of object forces and couplings. Each coupling would then be displayed in a 2-D graph with any desired function smoothing or modification. The evolution of the system will have to be in such a direction as to increasingly refine the integration processes by including parallelism of the force sampling process and the addition of adaptive algorithms for more accurate positioning of objects subject to rapidly fluctuating forces.

The above animation method is supplemented by a cynamic qualitative editing capability by including 3-D 'cut and paste' methods of audio, video and animation clippings, primitive graphs, raster images, etc. as well as quick and automated use of the Morphing techniques. The animated segments for video in real time can be produced with animation programmes like Autodesk AnimatorPRO or 3-D Studio. Microsoft Video for Windows, enable 256 colour images in resolutions up to 640 × 480 giving efficient video capture and import of single-frame images in a variety of formats and sizes. The digital Morph software utilised is VISTAMORPH running under VISTAPRO 3.0. Through 3-D Morphing, the transformation of one 3-D object model to another can be carried out. The objective of using VISTAMORPH in the present context is not so mush as to transform or distort 3-D objects, but to edit 3-D objects for realising more realisitc rendering. VISTAMORPH script building and editing utility which gives full control over the VISTAPRO 3.0 extended script language command, allows for the control of virtually every function in VISTAPRO from a script. VISTAMORPH allows the building of VISTAPRO frame by frame simply by appropriate functions clicked with the Mouse.

An important requirement in the creation of a virtual reality environment or cyberspace, is the availability of a multimedia database system. For VELNIC, MediaDB has been chosen which is a multimedia information management system with powerful tools for managing information in the form of texts, images, sounds, graphics and video with capability for easy manipulation of multimedia information. MediaDB is based on object-oriented database technology with the capability for managing extremely large pieces of information or objects. MediaDB server consists of storage, transaction and communications managers whereas the mediaDB client library consists of the following managers: database storage, transaction, object, type documents and communications. The MediaDB client library links into the end user client applications with the client and server capable of receiving either on the same machine or on different machines.

A number of software tools are available commercially which can assist in the building up of a reasonably realistic dynamic environment in cyberspace along with the realisation of 3-D depth with clarity.

3 Intelligent Hypermedia Navigation and Integration

The multimedia database file so created can be retrieved or manipulated by intelligent hypermedia navigation and intelligent integration with existing information. The goal of artificial intelligence in the context of the above described framework for the virtual environment is the creation of artifacts that can emulate reasoning faculty of the participant in his virtual image in cyberspace with the AI domain covering such symbolic reasoning functions as navigation, diagnosis of the situation in cyberspace and planning the next few moves. For this, it is necessary to construct explicit, declarative knowledge-base which, in turn, are operated by general, formal reasoning mechanisms. An intelligent system in this framework should be able to anticipate the changing pattern of the environment in cyberspace and respond appropriately through some type of adaptive behaviour. The adaptation should be as automatic as possible for responding to the exigencies imposed by the environment by making certain inferences based on how the virtual image is interacting with the cyberspace environment. As the intelligence in the virtual image of the participant still comes from the physical participant, adaptation and intelligence derived without reference to the physical participant, is mainly to overcome such effects as the 'lag-effect'. Such an intelligence displayed by the virtual image in cyberspace beyond the cognisance of and interaction with the physical participant, is termed as 'virtual intelligence' in this paper.

Virtual intelligence also provides a means of flexible navigation through complex process in cyberspace and programme functions underlying them, so as to lead to appropriate solutions. This enables the virtual image of the participant to take advantage of several non-linear pathways that offer the necessary level of depth or breadth needed to explore this cyberspace at its 'virtual will'. In this sense, the control of a programme is determined by the environmental conditions in cyberspace and the discretion of the virtual image which is not always pre-determined by the physical participant. To achieve this type of behaviour, virtual intelligence requires an appropriate hypermedia implementation along with other navigational aids.

The virtual intelligent system should also take advantage of the massive multimedia information stored in magnetic and optical disks. The mere separation of relevant information from the not-so-relevant one, with respect to the stimulus provided by the cyberspace environment and retrieving the critical information needed in solving the momentary problems faced, or exigencies of momentary decision-making by the virtual participant without reference to the physical participant, are tantamount to an exhibition of virtual intelligent behaviour. For deriving a perceptible extent of virtual intelligence in the design of a virtual reality system, no more tools are required than the well-tested techniques in AI including expert systems and Neural-Nets as well as hypermedia systems. More sophisticated virtual intelligence may call for the development of newer techniques which, however, are not attempted here.

In a knowledge-base that is created, containing production rules for decision-making or problem-solving as above, it is necessary to understand how the multimedia database and the corresponding knowledge-base component communicate with each other. Typically, how the expert system decides which of the rules, decision trees or objects in its knowledge-base correspond to the situation described in its database should be resolved by the inference mechanism. It is assumed that the knowledge-base has been created in the form of a combination of production rules, decision trees and object-oriented networks. The problem posed here is resolved by searching for patterns of logical relationship between the knowledge-base and the database. In an object-oriented network, the pre-determined patterns of inheritance within the classes of objects, trigger certain actions or events to occur within the expert system. These objects in turn, might activate specific functions or rules to lead to a specific result or conclusion. For such a process of testing rules in an optimal sequence, the standard inferencing techniques of forward or backward chaining are utilised (Fig. 2). To infuse the virtual intelligence in the virtual participant for efficient information management, the hypermedia tool for linking graphics, video, sound, animation clippings and other types of media associatively, is essential for enabling navigation through information in a non linear manner. Hypermedia also simulates the physical participant's ability to organise and retrieve information by referential links in the form of relational object-oriented network. The non-linearity helps in moving or jumping from one point in the programme to another based on the patterns of relationships that are explicitly defined.

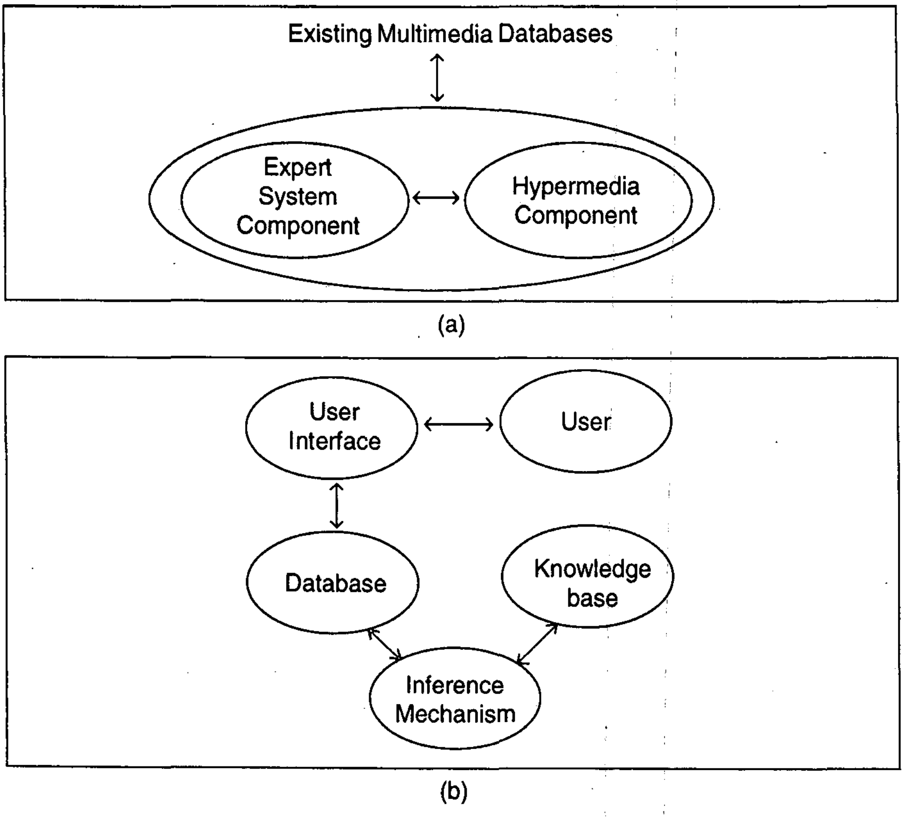


Fig. 2 *Intelligent hypermedia with an imbedded Expert System*

Though a powerful tool for intuitively interconnecting information, hypermedia cannot render intelligent applications. While providing a flexible context for representing knowledge, the application of the hypermedia tool does not lead to reasoning or inference ability. For this reason, there would have been compulsions to take hypermedia tool only as a supplement to AI techniques like expert systems and neural nets. However, there are other more important considerations which prompted us to take the reverse approach, i.e. designing the intelligent system based primarily on the hypermedia, but including an imbedded expert system component [15, 9]. In the creation of a virtual environment in cyberspace, existing information resources in the texts of original documents, collection of graphics, video animation and sound clippings, in massive quantity, are drawn to create the backbone structure of the system. The expert system components are required only to provide specialised local functions. AI technology is increasingly veering towards text-based approaches than intelligence-based approaches in the form of rules as systems can be built faster by side-stepping the laborious knowledge engineering process. This will emphasise hypermedia as the core of the intelligent system with the expert system components carefully imbedded in it to infuse intelligence.

A development tool that can combine hypermedia technology with expert systems is KNOWLEDGEPRO windows which works under the Microsoft Windows environment [9]. This tool was chosen because of its features like field hypermedia handling, excellent string manipulation ability, use of inheritance windows library of screen design tools and a powerful programming environment for expert system development, apart from the basic requirement of a high level of integration between expert systems, hypermedia and windows component. The supplementary use of another software product, KNOWLEDGEMAKER, can introduce a rule set from a group of examples that developers can directly import into KNOWLEDGEPRO or a number of other development tools. Although KNOWLEDGEPRO's inference mechanism is primarily backward chaining, it is possible to programme to chain in a forward direction also. However, KNOWLEDGEPRO does not implicitly have uncertainty handling features and therefore it forces the developer to ask the more sophisticated questions than to hedge the system's advice by associating confidence factors with its recommendations.

If there is a pre-set course of the dynamics of the environment which has not been exposed earlier to the physical participants, the adaptation behaviour of the virtual participant would require a virtual intelligence response due to 'lag effect', among others. Granting the flexibility of manipulation of the cyberspace environment through the commands from the physical participant, it is possible to introduce intervention-based modification of the environment through an appropriate communication modality from the physical participant. As can be seen from Section-VI, the hand gesture-based communication via the pair of fibre optic gloves, is reserved solely for the purpose of controlling the gait, orientation, movement and expressions of the virtual participant. The interventionary commands from the physical participants for modifying the environment therefore, requires some other distinguishable type of communication like speech commands. For this, a speech recognition sub-system is utilised taking due precautions from the analysis of such systems for VR applications as made by Duchnowski and Uchanski [5]. For a low-cost solution, it is necessary to train the system for the speech of the physical participant, take care to issue discrete word by word commands and limit the command vocabulary to the number of works the system can handle. For the purpose for which speech recognition system is used here, these limitations do not pose insurmountable problems.

4 Immersive Porthole View of Cyberspace

The physical participant can have a porthole view of cyberspace with full immersive effect through a head-mounted LCD display used along with a head-tracking system based on ultrasonic triangulation. For realising a low-cost and yet highly immersive system, a specialised Fresnel prism is recommended, a pair of which is oriented over the pair of LCD eye-pieces. A mathematical treatment of the stereoscopic effect underlying their use leads to the optimal orientation of the prisms [20]. For obtaining the correct stereoscopic images of a virtual environment, the position, size and shape of the simulated object should not change as the head moves. In virtual reality systems, a head-mounted display and a head tracker are used to compute the head position with speed and accuracy, and create an image for each eye depicting the instantaneous view. With this it is possible to see simulated objects from different points of view as the head moves.

The computation of the correct stereoscopic images in photography requires orthostereoscopy of occur when the perceived size, shape and relative position of object images match those of the physical objects in front of the camera. However, in virtual reality, there is no physical counterpart for the simulated virtual environment leading to the definition of orthostereoscopy as an invariance of the perceived size, shape and relative position of the simulated object as the head moves around. For calculating orthostereoscopic images, the display code must accurately represent the geometry of the display system on which the image will be viewed. This includes the relative positions of the display screens, the optics and the eyes, apart from the modelling of the relation between the screen and the virtual image. Such a problem has been extensively analysed by a number of investigators [4, 11].

An optics model is constructed to specify the computation necessary to create orthostereoscopically-precise images for a head-mounted display. It is necessary to identify the parameters of the system which are required to be measured and incorporated into a single-eye optics model shown in Fig. 3(a).

The mathematical treatment of the optical distortion relates to radial position r_s of a pixel on the screen to the radial position r_v of the virtual images of that pixel. Let k_{vs} describe the amount of distortion present and let

$$r_{sn}^2 = x_{sn}^2 + y_{sn}^2; \; r_{vn}^2 = x_{vn}^2 + y_{vn}^2$$

Then the position of the virtual image of the pixel can be derived to be

$$(x_{vn}, y_{vn}) = \{[1 + k_{vs}(x_{sn}^2 + y_{sn}^2)]\, x_{sn}, [1 + k_{vs}(x_{sn}^2 + y_{sn}^2)]\, y_{sn}\}$$

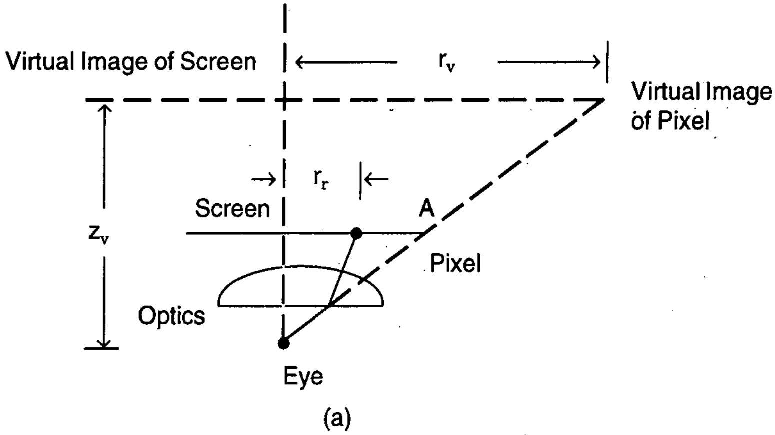
from the position (x_{sn}, y_{sn}) of the pixel on the screen. If D is the distortion function

$$(x_{vn}, y_{vn}) = D(x_{sn}, y_{sn})$$

then the inverse D^{-1} gives what is needed to predistort the image on the screen. This inverse can be approximately determined to be

$$(x_{sn}, y_{sn}) = \{[1 + k_{sv}(x_{vn}^2 + y_{vn}^2)]\, x_{vn}, [1 + k_{sv}(x_{vn}^2 + y_{vn}^2)]\, y_{vn}\}$$

Figure 3(b) depicts the stereoscopic optics model. Points A_1 and A_2 are the pixels



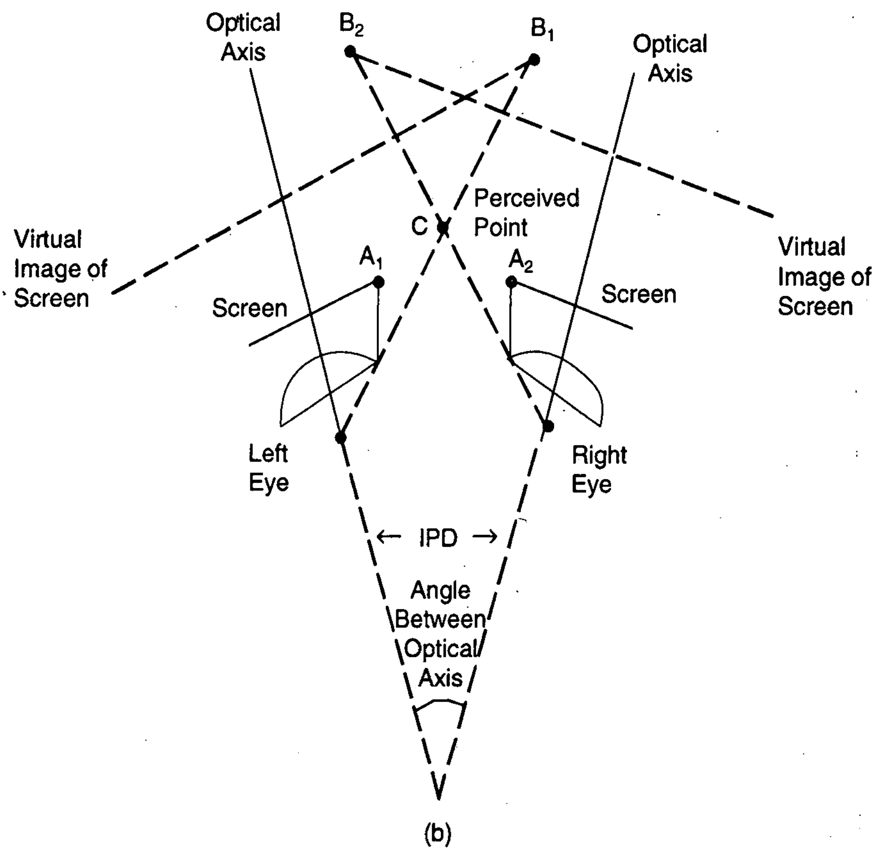


Fig. 3 *(a) Single eye optics model and (b) Stereoscopic optics model for the head-mounted LCD display*

illuminated on each screen while B_1 and B_2 result from a line of sight drawn from each eye to the virtual image of its corresponding picture. These two lines intersect at C, a 3-D point perceived by the user. The Inter Pupillary Distance (IPD) is the base line from which the participant makes the judgement based on the convergence angles to the perceived points. If the optical specifications and the relative positions of display screens, the optics and the eyes are a priori known, it is possible to correlate and calculate several important parameters needed in the model. Utilising these calculations, the design of the head-mounted display can be worked out in which the computer generates stereo images that the participant can view on colour LCD displays through wide-angle optics.

For enhancing the clarity of the 3-D video effect, the above stereoscopic optics is superposed by the hologram effect of the Fresnel zone plate. In traditional optics, Fresnelz one plate is an interference device similar to the defraction grating and is a system of alternating transparent and opaque contours whose width diminishes according to a special formula, as the radius increases. In this sense, it functions as a hologram of a point object as well as a planar equivalent of a lens [20]. However, for the present objective, a special design of a pair of Fresnel prisms are designed such that they bend the light from the LCDs so that the eyes of the participant can converge the image without using any software image shifting. With Fresnel prisms, the head-mounted display can accept a single video input delivered to both eyes. These prisms need not be removed when viewing dual, computer shiftable images.

The head tracker has a stationary triangular transmitter which emits ultrasonic signals with track movement of the receiver fastened on top of the helmet. The receiver, which moves about 3-dimensionally within the active area receives ultrasonic signals and relays them back to the control unit. The control unit converts all the signals into data on the position and orientation of the head that can be processed by the host computers. The head tracker also allows incorporation of voice recognition and sound applications. The 3-D mouse is similar to the head tracker except that when held with the hand, it tracks the position of the hand in 3-D space and when moved on the table, it functions exactly like a 2-D mouse. The 3-D receiver has, apart from the three top buttons of the 2-D mouse, two suspend buttons on the right and left sides for enabling operation in the 3-D mode.

5 Imbedding the Virtual Participant in Cyberspace

With the creation of the virtual environment or cyberspace as above, the next step is to embed the virtual image of the participant in the cyberspace by creating a realistic 3-D image of the participant and endowing this image with dynamic characteristics. This can be carried out in the following three steps:

(a) From the 2-D photographs of the front, back, left side, right side and the top of the still profile of the participant, a 3-D still profile is constructed.
(b) The 3-D still profile is programmed to exhibit dynamic characteristics by identifying all the operative joints with appropriate degrees of freedom to achieve a gait of movements as realistic as possible.
(c) The head (including face) of the virtual participant is portrayed in a form

recognisable as a true image of the physical participant including a coordinated movement of the facial muscles and the lips while speaking.

The first and second steps are carried out in a straight-forward manner utilising the familiar software like Autodesk 3-D Studio and Autodesk Animator. With the existence of graphic workstations which are able to display complex scenes containing thousands of polygons at interactive speeds and with the availability of VR interactive devices, it is possible to create applications based on a full 3-D interaction in which the specifications of deformation or motion are given in real time. The 3-D animation of the profile of the participant can change every time in accordance with the laws of motion subjected to physical constraints. The problem of expressing time dependence of the virtual participant and making it to evolve over time, is a complex one to manage. Once the initial human shape has been created, this shape should change during the animation. This complex problem of ensuring continuity and realism of the deformed surfaces can be carried out by MORPH software packages like VISTAMORPH [6]. As the human animation is complex, it should be split into two parts—body motion control and facial animation. The virtual participant is structured as articulated body defined by a skeleton. The skeleton animation of joint angles can be carried out by parametric key frame animation or physics-based animation. The animation of the human facial anatomy is, however, more complex.

The 3-D shape creation for the profile of the virtual participant can be carried out by the sculpting approach based on the so-called, 'ball and mouse metaphor'. Here, the motion parallax consists of the brain's ability to render a 3-D mental picture of an object through its motion relative to the eye. For an uninterrupted display of the object movement, the hardware should be capable of very high frame rates. To acquire this depth perception and mobility, a device called the 'SpaceBall' is used. When holding the SpaceBall in one hand and the common 2-D mouse in the other, full 3-D user interaction can be achieved. The SpaceBall is used to move around the object being manipulated to examine it from various points of view, while the mouse carries out the picking and transformation work on to magnifying image in order to see every small detail in real time like Vertex creation, primitive creation, surface deformation and cloth panel position. With this facility, the operations performed while sculpting an object closely resemble the traditional sculpting. One of the best software available todate for sculpting a human shape, was developed recently by Paouri, M. Thalmann and D. Thalmann [14]. The sculpting process may be initiated either by Morphing a simpler shape or by starting ab-initio. For example, we can use a sphere as a starting point for the head of a participant and use cylinders for limbs. Subsequently, we add or remove polygons according to the details needed and apply local deformation to alter the shape using Morph software [6].

The computer animation of the human facial expression requires the synchronisation of eye motion, expression of emotion, lips and facial muscle movement during speaking, etc. One of the best software available for facial animation is the SMILE software package developed by Kalra, Mangili, M. Thalmann and D. Thalmann [8]. The system is based on a multilayered approach for specifying facial animation. Each successive layer defines entities from a more abstract point of view. Starting with muscle deformation, one works up to phonemes, words, sentences, expressions and emotions. With a wire-frame model, a 3-D model for the face (including lips and teeth) has been developed by Aisawa, Harashima and Saito [1] and refined by Morishima and Harashima [12].

One of the most convenient tools for computer animation required for virtual reality is the automated virtual camera. From the basic ideas proposed by Zelter in 1985 [23] for developing an integrated view of the 3-D computer animation and the development of virtual movement camera for special cinematographic effect during the same year by M. Thalmann and D. Thalmann [21], the technology of the virtual camera has matured adequately to play a nodal role in VR animation, culminating in the development of automated virtual camera by Noma and Okado [13]. Several real-time direct metaphors have been developed for controlling camera motion. For the non-physical based motion control a 3-D interactive programme called, ANIMATOR [2] has been developed for allowing the creation of several entities like objects, cameras and light. For each entity, a 3-D path may be interactively generated using the SpaceBall whose trajectory is generated using a spline. The ANIMATOR may build complete hierarchy of entities using only the Mouse, after which the animation is created by defining the objects. These paths are generated in 3-D using the SpaceBall while temporal information is defined by control points. The trajectory is then generated using B-splines. A physical-based camera control model provides a powerful, general purpose metaphor for controlling virtual cameras in interactive 3-D environments. The virtual camera has seven degrees of freedom three for camera position, two for line of sight, one for twist angle and the other for zooming. Automating these features has provided a powerful tool in VR animation of the virtual participant.

6 Gesture Control of Virtual Participant

After animating the virtual participant endowed with a capability of natural limb movements and facial expressions, the next essential step is to develop an efficient modality of interaction between the physical participant and the virtual participant. Unlike the multimillion dollar solutions offered with sophisticated tactile feedback systems, the present approach advocates a low-cost solution in which the physical participant is required to be almost static. A subtle psychological metaphor is utilised which allowed the physical participant to control the movements and facial expressions of his virtual image. By using low-cost optic fibre gloves, the various possible configurations of the fingers can, in principle, yield 5! (= 120) possible commands on one hand and 10! (= 3628800) possible commands by completely opening or closing the subset of fingers on both hands. The left-hand fingers and right-hand fingers can be allocated respectively to the main functions and the subsidiary functions within the main functions. The finger binary coding structure for the desired number of commands can be worked out keeping ergonomic conveniences in view. In the costlier designs of the fibre optic gloves, the opening and closing of fingers are done at the distal, middle, proximal and metacorpal levels leading to a ternary coding scheme.

Vaanan and Bohn have built a "Gesture-driven Interactions in Virtual Environment", called GIVEN [3]. This is a tool kit with the following features:

input device independence, application and object-type independence, open extendable architecture, object orientedness, individual behaviour of objects and Gesture dialogue facility.

When the gesture command is issued, there is a finite time delay between cause and effect, and the virtual image responds with a 'lag effect'. Lags greater than 100 milli seconds in the response to the gesture can cause motion sickness to the physical participant because of the continuously perceived discrepancy between head motion and the rendering. This is more pronounced in the case of delays in sound than the visual delays. To overcome this, a method is required for accurately predicting the sensor position to synchronise the motion and rendering the sound arising from the virtual participant. The delay is introduced by various factors such as the intrinsic delay in the sensing device, the CPU time required to calculate the proper response and the time spent in rendering output images or minimum time required for generating appropriate sounds. Unexpected interruptions may also introduce delays. Borrowing from control system theory, Kalman Filter Estimation and Prediction methods can be utilised for obtaining an optimal linear system of the state vectors of dynamic models for predicting the state vectors ahead in time. As an approximation, it is adequate to apply this method to treat only the translational components (x, y, and z) output by the sensor. The basic method for gaining lead time for synchronisation through adaptive Kalman Estimation and Prediction is outlined:

A 'dynamic' process

$$X_{k+1} = f(X_k, \Delta t) + \xi(t)$$

and an observation process

$$Y_k = h(X_k, \Delta t) + \eta(t)$$

where the function f models the dynamic evolution of state vector X_k at time k, the sensor observations Y are a function h of the state vector and time, ξ and η are white noise processes having known spectral density matrices.

In the sensor device used in VELNIC, the state vector X_k consists of the true position, velocity and acceleration of the sensor in each of the x, y, z coordinates and the observation vector Y_k consists of the position readings for the x, y and z coordinates. The function f and h will describe the dynamics of the user's movements and measurements respectively in terms of the state vector.

Using Kalman's result, we can then obtain the optimal linear estimate $\hat{X}_k$ of the state X_k by using the Kalman filter:

$$\hat{X}_k = X_k^* + K_k(Y_k - h(X_{k,t}^*))$$

provided that the Kalman gain matrix K_k is correctly chosen. The prediction of the state vector X_{k+1} at the next time step is obtained by combining the optimal state estimate X_k and the first equation.

$$X_{k+1}^* = \hat{X}_k + f(\hat{X}_k, \Delta t)\, \Delta t$$

In our 3-D dynamic animation, this prediction allows us to maintain synchronisation with the user by giving us the lead time necessary to complete rendering, Morph, sound generation, etc.

Figure 4 shows the processes and communication path used to filter and query each sensor. Since it cannot be readily ensured that the application control process will need query sensors on a regular basis and since it is not desirable to let Kalman filter to enter into the processing pipeline, we bring in two small processes to constantly query and filter the actual device. The application control process will consequently have the opportunity to make a fast query to the filter process for realising the most uptodate filtered sensor position. Using the shared memory between these two processes makes the final queries optimal.

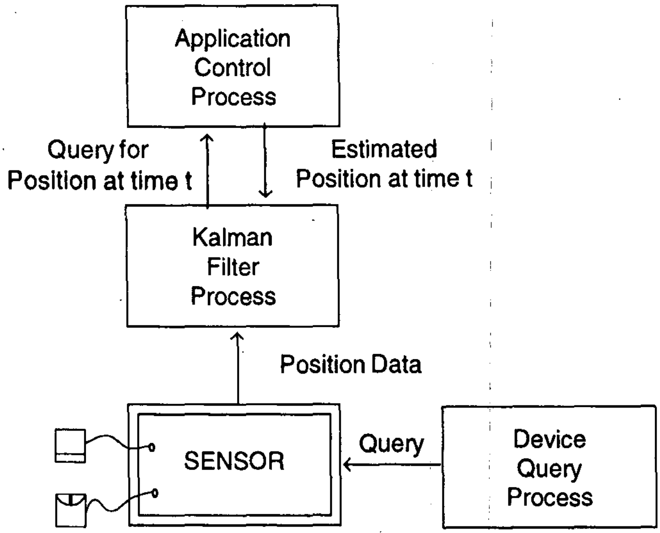


Fig. 4 *Communications for control and filtering of a sensor device*

7 Concept of 'Virtual Intelligence'

The design of a participatory virtual reality system for visualising an adaptive interaction between the simulated 3-D dynamic virtual environment and the dynamics of the 3-D virtual image of the participant in the cyberspace described in the foregoing leads to the formulation of a new concept which may be called 'Virtual Intelligence'. The psychological dynamics of the subconscious and conscious levels of the mind, create certain powerful illusions in the physical participant, making him loose conscious awareness of certain intermediary processes in the command gesture control of the virtual participant. There is an analogy for this in the process of speech formation. When we speak, unless consciously observed, we tend to forget about the synchronous movements of the tongue, lips and other facial muscles in order to produce the speech sounds. The will to express something verbally brings out the sound, while involuntary muscular movement of the face, tongue and the lips automatically assume pre-set patterns. This act of the will directly translates the idea in the brain into the speech sounds. Similar psychological adjustments take place when gesture-based commands are given by the participant for controlling his virtual image using sign language transmitted from the pre-set gestures of the fingers via an optical circuitry in the fibre optic gloves to the computer which translates the commands into appropriate dynamic changes in the virtual participant. It is as if the will of the participant brings about certain actions of this virtual image directly without being consciously aware of the gesture language expressed through fingers via the fibre optic glove. This will be so if the participant has practised the gesture language adequately enough for gestures to become involuntary. Experiments with the VELNIC set up have shown that there exists a tendency of the mind of the trained participant to create an illusion of directly 'attaching' itself to his virtual image. This effect gets especially accentuated because of the VR helmet covering his eyes preventing visual and audio contact with his physical body, including his gesturing commands.

It is in the above context that the 'lag effect' creates a psychological trauma. While being not consciously aware of the gesture, but keeping track of the same subconsciously or involuntarily, the delay of the corresponding action in the virtual participant will bring the mind of the physical participant from the subconscious level of awareness of the gestures to the conscious level and fly back to the subconscious level regarding the subsequent gestures when the eyes perceive the action after the delay time. This in effect creates a mild oscillation of awareness between the conscious and the subconscious resulting in a possible psychological trauma. The use of Kalman Filter-based Estimation, Forecast and Correction overcomes this situation by minimising the effect of the delay. The adaptive behaviour of the system incorporating Kalman Filters happens outside the purview and mental process of the physical participant. In the same way, the intelligent hypermedia-based system which carries out the adaptive navigation and integration also operates outside the purview and mental process of the physical participant. As the basic intelligence in the gesture-driven virtual image still comes from the physical participant, the adaptation and intelligence derived without reference to the physical participant is mainly to overcome such effects as 'lag effect' and flexible navigation.

As pointed out in Section-III, such autonomous intelligence displayed by the virtual image in cyberspace beyond the cognisance of and interaction with the physical participant, is named 'Virtual Intelligence'.

The circumstances for the exhibition of virtual intelligence described in the foregoing are not the only ones possible. In the interaction between the virtual participant and the virtual environment, intelligence can be autonomously induced in the virtual participant beyond the cognisance of the physical participant. For example, the virtual participant may confront a situation in the virtual environment for which there is a multiple choice of decisions. The virtual participant may be directly conditioned by a decision-optimisation algorithm in the computer to enable it to take the best decision. When the physical participant is confronted with the multiple choice problem, his intuitive solution may be non-optimal. Yet, he conveys the command in the action by the virtual participant through the gesture language. If the criteria is built in the virtual participant to disregard a non-optimal solution, the earlier implemented optimal solution will prevail. In such a situation, the virtual participant would be deemed to have acted intelligently outside the command of the physical participant. The kind of feedback impact such a virtual intelligence may impose on the physical participant is a problem for future research by techno-psychologists.

REFERENCES

1. Aizawa, K., Harashima, H., Saito, T., "Model-based analysis synthesis image coding system for a person's face", *Signal Processing: Image Communication* **Vol. 1**, No. 2, pp. 139–152, Elsevier, 1989.
2. *ANIMATOR—User's Manual*, Swiss Federal Institute of Technology, CH-1015, Lausanne, Switzerland.
3. Bohm, K., Hubner, W., and Vaanan, K., "GIVEN: Gesture-driven Interactions in Virtual Environments—a toolkit approach to 3D interactions", *Proc. Interfaces to Real and Virtual Worlds Conference*, Montpellier, France, 1992.
4. Chung, J.C., et.al, "Exploring Virtual worlds with head mounted displays", *SPIE Proc. Vol. 1083*, 1989.
5. Duchnowski, P., and Uchanski, R., "Speech Recognition", N. Durlach (ed.), *Virtual Environment Technology*, BBN Systems and Technologies, Report No. 7661, 1992.
6. Heckbert, P.S., "Digital Image Warping: A Review", *IEEE Computer Graphics and Applications,* Jan. 1991.
7. Henderson, J., "Designing realities: Interactive media, Virtual realities and cyberspace", *Virtual Reality: Theory, Practice and Promise*, ed. by S. Kelsel, Meckler Publ., Westport, Conn. pp. 65–73.
8. Kalra, P., Mangili, A., Thalmann, M., and Thalmann, D., "SMILE: A multilayered facial animation system", *Proc. IFIP Conf. on Modelling in Computer Graphics*, Springer-Verlag, pp. 189–198, 1991.
9. 'KnowledgePro Upgrade blends hypertext and expert systems', 1989, *PC Week*, p. 5, December 18, 1989.
10. Lanier, J., 1990, "Virtual Reality", in article by Hall. T., *New York Times*, p. 1, July 8, 1990.
11. LEEP, *Cyberface II Applications Note*, LEEP Systems Pop-Optix Labs., 241 Crescent St., Waltham, Mass, 02154 USA, 1990.
12. Morishima, S., and Harashima, H., "A media conversion from speech to facial image for man-machine interface", *IEEE journal on Selected Areas in Comm.* **Vol. 9**, No. 4, 1991.
13. Noma, T., and Okada, N., "Automatic Viewing for Computer Animation", *Proc. 42nd Annual Convention IPS Japan*, pp. (2) 367–368, 1991.
14. Paouri, A., Thalmann, M., and Thalmann, D., "Creating Realistic 3-D Human Shape Characters for Computer-generated Films", *Proc. Comp. Animation*, Geneva, Springer-Verlag, pp. 89–100, 1991.
15. Santarelli, M., "Expert Database Systems", *Proc. of the First Intl. Conf. on Expert Database Systems*, Charleston, SC, pp. 109–120, 1986.
16. Seshagiri, N., 1992, "Adaptive Distributed Multimedia: A Concept for characterising Co-cognitive Virtual Reality Systems", *Multimedia Computer and Communication*, Ed. by Seshagiri, N., and Akopov, A., Tata McGraw-Hill Publ. Co. Ltd., New Delhi, pp. 36–56, 1992.
17. Spiegel, M., *Applied Differential Equations*, III Edition, Prentice Hall Engelwood Cliffs, NJ, pp. 1–26, 1988.
18. Sutherland, I., "The head-mounted three dimensional display", *Proc. Fall Joint Computer Conf.*, pp. 757–764, 1968.
19. Sutherland, I., "The ultimate display", *Proc. IFIP Congress*, pp. 506–508, 1965.

20. Tarasov, L.V., *Laser age in Optics*, Mir Publ., Moscow, pp. 75–96, 1981.
21. Thalmann, M., and Thalmann, D., "Single and Multiple Virtual Movie Cameras", Kunii, T.L., (ed.) *Computer Graphics*, Springer-Verlag, pp. 271–283, 1985.
22. Wilhelms, J., et. al., "Dynamic animation: Interaction and control," *Visual Computer* 4(6), pp. 283–295, 1988.
23. Zelter, D., "Toward an integrated View of 3-D Computer Animation", *The visual Computer*, 1(4), pp. 249–259, 1985.
24. Zyda, M., and Pratt, D., 1991, "NPSNET: A 3D Simulator for Virtual World Exploration and Experimentation," Society for Information Display, *Intl. Symp. Digest*, pp. 361–364, 31 May 1991.

Calendar of Events

April 6–8, 1994
7th European Conference on Machine Learning (ECML 94)
Location: Sicily, Italy
Contact: Mr. Luc De Raedt, Department of Computer Science, Katholieke University, Leuven, Celestijnenlaan 200A, B-3001 Hererlee, Belgium

April 18–21, 1994
The 21st Annual International Symposium on Computer Architecture
Location: Chicago, USA
Contact: Mr. Janak H Patel, Coordinated Science Laboratory, University of Illinois, 1308 W. Main Street, Urbana IL 61801, USA

May 2–4, 1994
1994 International Conference on Data and Knowledge Systems for Manufacturing and Engineering
Location: Shatin, Hong Kong
Contact: Prof. Nelson Chen, Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, Tel: (852) 859-2582, Fax: (852) 858-6535

May 4–6, 1994
Modelling Techniques and Tools for Computer Performance Evaluation
Location: Vienna, Austria
Contact: Mr. Gunter Haring, Department of Applied Computer Science, University of Vienna, Lenaugasse 2/8, A-1080 Vienna, Austria, Fax: (43) 1 408 04 50

May 16–20, 1994
1994 ACM SIGMETRICS Conference on Measurement and Modelling of Computer Systems
Location: Nashville, Tenn., USA
Contact: Mr. Rick Bunt, Department of Computer Science, University of Sakatchewan, Saskatoon, SK, S7N OWO Canada, Tel: (804) 221-3458

May 23–27, 1994
IFIP SEC 1994—Tenth International Information Security Conference
Location: Aruba, The Netherlands
Contact: IFIP SEC '94 Secretariat, PO Box 1555, 6200 BN Maastricht, The Netherlands

May 24–27, 1994
KR '94: 4th International Conference on Principles of Knowledge Representation and Reasoning