



MASTER THESIS

MULTIMODAL USER INTERFACES FOR CAAD

Supervised by

ao. Univ. Prof. Dipl.Arch. MS Arch. Dr.phil. Georg Suter

E259.1

Institute of Architectural Sciences

Dept. of Digital Architecture and Spatial Planning

submitted at The Vienna University of Technology

Faculty for Architecture and Planning

by

Luis Daniel Pozo Torres

0928202

Vienna, 25 February 2013

Acknowledgements

I would like to express the deepest appreciation to Prof. Georg Suter to whom I owed a big gratitude for his time spending with me, for his advices and guidance, for the long discussions and patience, without his persistent help this Thesis would not have been possible.

I thank the professors and Assistants of the Master program in Building science especially to Prof. Adeshir Mahdavi, Prof. Bob Martens, Prof. Peter Ferschlin and Prof. Georg Franck-Oberaspach. Their lectures and work where an important source of inspiration and motivation.

Special thank to my father Juan Manuel, and mother Maria Luisa for supporting my studies in all possible ways, for trusting me and giving me the opportunity of studying abroad , most of all for giving me the privilege of taken the name of Bolivia to far away territories.

Thanks to Michael Lisner for his support in the last stages of the realization of my Thesis.

Finally many thanks to the people that is not been mentioned here, but in one way or another contributed to my academic projects during the last 2 years of studies in Vienna.

Best Regards

Luis Daniel Pozo Torres

Abstract

This thesis examines the potential of multimodal user interfaces for computer-aided architectural design (CAAD) systems. Multimodal user interfaces leverage visual, haptic and sound based channels either for data input or output. The Interest in this field lies in the rich functionality and advanced user interaction which can be introduced in the development of future CAAD systems.

A survey of the technologies and existing applications that use multimodal user interfaces was made. Design tasks such as; the creation of a wall, the placement of a window on a wall element and the change of position the window, were considered as requirements for the definition of a multimodal user interface. Based on the analysis of requirements a design task network is proposed in a diagrammatic way. In addition suitable input and output channels have been defined as the means of communication between users and CAAD systems. The design task network can be used to implement and evaluate the performance of future multimodal CAAD systems.

The thesis concludes with a discussion of the challenges of multimodal user interaction for CAAD and describes an agenda for future research in this area.

Contents

1. Introduction	1
2. Background	5
2.1. Multimodal human computer interaction (MMHCI)	5
2.2. Technologies for interaction	9
2.3. Applications	22
3. Requirements for CAAD systems	28
3.1. Types of data	29
3.2. Workflow description	33
4. Design task network for Multimodal CAAD systems	35
4.1. Design process with a Multimodal interaction interface	35
4.2. Interface definition for CAAD systems	36
4.3. Design process visualization	41
4.4. Interface mockup	45
4.5. Examples	48
4.6. Visualization of information for Multimodal interaction	55
5. Conclusions	58
References	60
Web resource	64
Nomenclature	67
List of figures	67
List of tables	70

1. Introduction

Advances on computing and digital technologies have changed the way many professionals conduct their work. With the introduction of the PC (Personal Computer) along with the various software applications for processing information, into the market, the digital platform has expanded at an increasing rate. In this context engineers and designers have found in these technologies a useful and precise tool that allows them to process large quantities of data; and to represent, visualize and manipulate different kinds of information in a more comprehensive way.

Since the 1980s with the implementation of CAD systems the way of producing and visualizing diagrams, geometries and detailed drawings has changed dramatically. What was once a manual time consuming production it is now being hosted now by a digital platform where the information is processed in the digital space with the use of various software applications which can mimic the same functionalities of a real workspace by providing several tools for digital editing, storage and processing. The means have changed so did the workspaces and big tables were changed for smaller computers and now in days software developers enrich constantly the set of tools and applications that serve for digital design production, graphical representation and data administration.

User interfaces plays a big role now that they make possible the link between human and digital world, without an adequate language between both even the simplest task would become problematic to solve.

Since the introduction of the Pc the interaction between users and computers takes place within a dynamic loop understood as: A trilogy of the keyboard, mouse and the screen. Nevertheless this has changed during the last years with the incorporation of Mixed Reality and Augmented Reality interfaces which allow a more dynamic and

intuitive interaction with the digital, this research will be focus on this type of interface interaction and its application and potentialities towards the development of an innovative design tool for Architects and Engineers.

The various levels of complexity that an Architectural project demands, starting from the diagrammatic concepts until form generation are always though in a spatial way, in a three dimensional world. The representation of this three-dimensionality in the digital world is possible thanks to the software that allow users to interact by manipulating and exploring geometrical configurations. Unfortunately this activity has been relegated to the minimum of human physical expression by the conventional means of Human Computer Interaction, so the entire process of conceiving spatial ideas when developing an architectural project is limited in human expressivity by the conventional interaction methods [1].

The actual scenario of the workstation for professionals that deal with computers (Fig.1), among them architects and engineers, was design to increase productivity and reduce fatigue. This type of body movement restriction for long hours brings along many problems that deal with human health and psychological effects. Therefore the design of future workstations requires serious attention when dealing with human mobility.

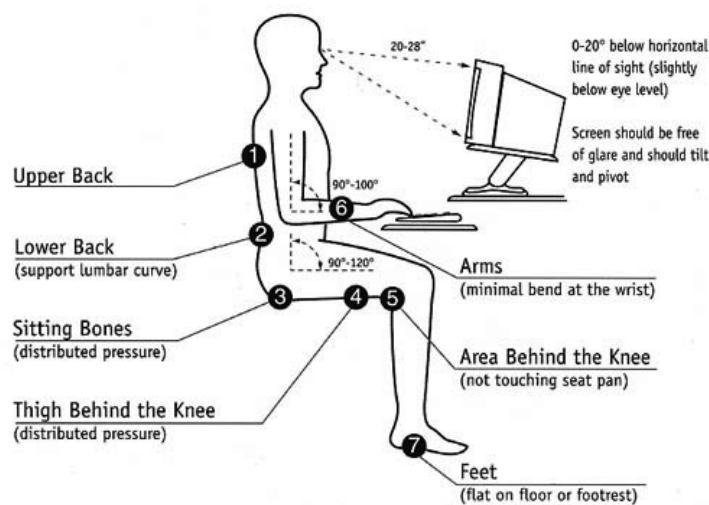


Fig.1 Desktop Computing (Source: web page)

Another issue rises when dealing with the manipulation of the geometric entities and administration of information which has reached a high level of complexity, the two dimensional output conceived as the flat screen of the computers is not sufficient enough for an adequate understanding of a project. This planarity is another way of restricting the users to interact with their real information which is represented as 2d images when they are been though in 3d.



Fig.2 Immersive reality ,CSI Galicia (Source: web page)

Here is where Virtual Mixed interfaces (Fig 2.) offer the user more possibilities for physical tangible interaction [1][2] by merging the digital into the real world, with this the means of communication also expand for example the movement of body parts as well as the speech can be recognized and later be translated into actions and commands. In Architectural production this could be useful due to the different kinds of information that has to be processed. An adequate Interface System that allows this type of interaction could change the ways of production now that in cyber spaces physical restrictions are removed and imagination can flourish so that humans can walk back and forth in mental and digital-virtual spaces as they want [3]. The concept is simple, yet not buildable due to the technological limitations that could make this possible, but this is in any case not an issue because advances on science have always prove to have an accelerate rate, having this assumption the necessary means would be available in the next 10 years, therefore the need of research in this field.

The research consist first in identifying the technologies that made possible human interaction with the digital world, the information was taken from an extensive

bibliographic review supported by digital libraries and Internet. In order to narrow the scope and for purposes of working with actualized information, documentation published after the year 2000 was mostly considered. After identifying the potential of Multimodal Interfaces a specific case of a design task routine based on the generation of a building element, in this case a “wall” , was analyzed to later propose a network system that describes how MMHCI techniques work and could be represented for farther analysis and optimization.

The document is organized in 5 Parts: Part 1 contains the Introduction of the main theme, Ideas and motivation for writing the thesis, Part 2 makes focus on the references and the Background of HCI technologies and applications, in Part 3 requirements for a design routine of 3d Architectural Objects are established. In Part 4 upon the previous analysis a network system for understanding and mapping the way the user interaction happens while working on Multimodal systems is being proposed. Part 5 contains the analysis and evaluation of the results. Finally references and the list of tables and figures are mentioned.

2. Background

2.1. Multi Modal Human computer Interaction (MMHCI)

Human-computer Interaction (HCI) is an area of research and practice that emerged in the early 1980s initially as a special area in computer science, HCI is defined as “a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them” [4].

As the advances in computing technologies grew Human-computer Interaction became an important field when dealing with computer graphics, operating systems, human factors, cognitive psychology, (Fig.3) and many other systems part of computer sciences. All this led to the development of several HCI techniques for different purposes and users.

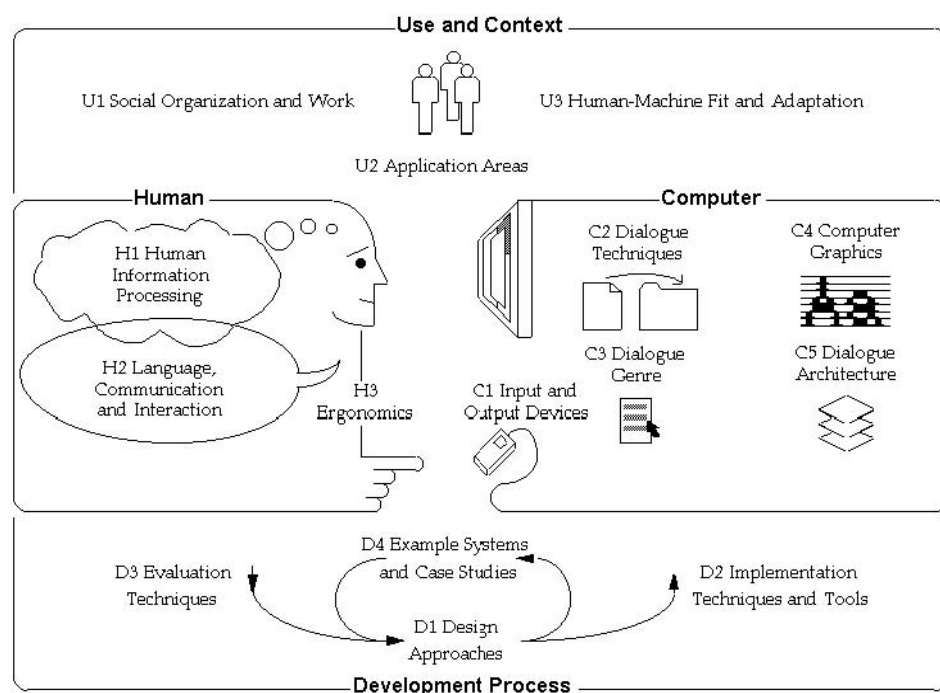


Fig.3. Human-Computer Interaction [13]

The first HCI techniques of input and output were designed for a closed loop system which considers a “response computing time to human interaction times” for multiprocessing and supporting windowing environments to solve the problems of substantial cognitive communication and interaction aspects not previously developed for human needs [7]. The next step was to focus on "user interface management systems" and "user interface toolkits".

The evolution of the technology has lead to a standardization of the interfaces which we can find now on days from the side of both the hardware and software, a clear example the Pc, and the window systems introduce by Microsoft in 1985. Nevertheless the expectations for the future development on HCI systems will have to deal with [4]:

- Ubiquitous communication,
- High functionality systems,
- Mass availability of computer graphics,
- Mixed media,
- High-bandwidth interaction,
- Large and thin displays,
- Embedded computation,
- Group interfaces,
- User tailorability (Tailorable), and
- Information utilities.

In the last thirty years research on sensorial tracking and speech recognition have been characterized by the constant accumulation of small incremental improvements coming aside the advances in computing technologies. In this context multi sensorial HCI has always been placed at the intersection of several research areas where each component also called modalities have been treated by separate but in recent years

with the merging of Mixed Reality technologies and Ubiquitous Media Interfaces [5], Multimodal interfaces have gained interest because of their potential for improving communication with a real time interaction between the digital and the real world.

As computers become integrated into everyday objects, effective natural human-computer interaction becomes critical. The necessity of humans to communicate with computing systems have reached high levels of complexity because of the merging of new types of processing information so the actual means of interaction are not sufficient. Another fact is that many computer applications users need to be able to interact naturally with computers the way face to face human-human interaction takes place, [6] consequence of these may be gifted with a high level of intelligence to support the performing of multiple tasks by humans.

With the introduction of ubiquitous computing instead of using intermediate hardware tools humans could interact with computing systems in a more natural way using the same techniques humans use for communicating between them, this means to make use of speech and body gestures which are consider the everyday fundamental means of communication.

Humans are capable of perceiving their surrounding world thanks to the simultaneous use of 5 senses: Sight, sound, touch, taste and smell. Now when dealing with multisensory also called multimodal systems means that more than one channel or sense takes part on the interaction [25]. The most common ones used for HCI are: sight, sound and touch. Multimodal is considered when more than one channel is used such as Speech and Gesture [7]. Multisensory technologies are being characterized in two groups: one for input which deals with the human point of view and the other the output which comes from the computer retrieval. Table 1 shows the classification of both input ad output channels.

Human senses	Channels for human computer interaction	
	Input to the computer	Output from the computer
Sight	Eye movement	Visual display
Touch	Hand gestures	Tactile feedback
Hearing	Voice commands	Speech synthesis
Taste	-----	-----
Smell	-----	-----

Table1. Multi-sensory interaction channels [9]

One of the problems encountered when dealing with Multisensory systems are the administration of data obtained which comes in large quantities and needs to be combined for processing and recognition according to specific commands, for example when dealing with speech recognition the problem is originated due to the singularities of the way of talking of multiple the users, so the challenge for the system is to classify this input data according to certain affinities, the same happened with gesture recognition. The ideal scenario would be that any input to the system should be recognizable in real time to provide a natural communication bridge between humans and the computer.

Another approach on these technologies deals with affective human computer interaction stated in [8] where the authors discuss the possibility that “affective states” should be linked to function, so computers could be able to recognize emotions and other cognitive states and expressions. This affect recognition could be a way of providing an accurate understanding of Multi sensorial Inputs by adding certain patrons of behavior for solving possible mismatches on the data interpretation.

Finally the problems found in the analysis and Interpretation of individual modalities could be overcome by integrating a multimodal HCI, this has been proved by Chi-Chen Chu [9] where the results show that the user can acquire 80% of usefulness and helpfulness by multiple interaction mechanisms.

Some recent examples of this type of technologies happen at Stanford University where research groups in HCI are using these multimodal systems to generate smart environments. Other types of actual applications can be found in the design of smart meeting rooms for communication, in museums for didactic uses as part of experiencing art and in the robotic industry.

2.2 Technologies for interaction

2.2.1. Voice input

This type of input allows users to introduce data in the form of literal commands and numbers by recalling them with the speech, an example of this takes place on the applications developed by Autodesk where a command line works as a literal mean for recalling certain required tools whenever they are needed. One important functionality of this type of interaction happens when a the command line works as an assistant by suggesting possible commands according to the task being performed or according to the affinity of typing symbols , this can be in the form of questions or visual displays.

In [10] and [11] the authors used a CAD targeted vocabulary as well as CAD-targeted grammars to enhance the practical applicability of a Voice Aided CAD System , they also proposed a novel type of Context-Aware Inference (CAI) to simulate human's intelligent inference by means of context awareness. Commercial applications that offer these functionalities can be installed in conventional Pc's and work with the support of a microphone. This type of software is also known as CSRA or "Computer Speech Recognition Applications":

Windows Speech Recognition is a commercial product developed by Microsoft, it already introduces an engine for Speech Recognition on the version of Windows Vista and the latest versions of Windows 7, is available in 7 languages: English, French, Spanish, German, Japanese, Simplified Chinese, and Traditional Chinese.

Speak4CAD[@Speak4CAD] is another existing commercial application developed for Auto CAD, it comes as a Suite Packet which includes versions for, AutoCAD, Inventor, SketchUp, Solidworks, InDesign and PhotoShop, as a way of promoting this type of functionality

Vox Force[@Vox Force] is a speech recognition engine product with a free General Public License from Open Source; it runs on Linux, Windows and Mackintosh and for mobile applications like Sonic Cloud and Dragon Dictation.

2.2.2. Hand Motion

This type of input works by capturing the movement of body parts, recognizing the gestures of the hands to establish a communication between users and software applications. This method be used either for direct manipulation of digital representations or as a way of recalling tools by using gesture commands [13].

There are two ways for the input to take place; the first one deal with a Vision based recognition method and the second uses wearable devices with sensors incorporated to read visual information (Fig.6).

Vision based Recognition Software Applications work by processing the information that comes from a series of images of a video camera so that a visual memory can recognize the hand posture, the orientation, the center of the movement and the finger tips. These types of algorithms first recognize the contours of the shapes of the hand with information such as skin color (Fig.4), later it analyzes the curvature of the perimeter of the shape. For example a convex hull small enough can be detected as a

finger tip (Fig.5), finally this information becomes a cloud of points of reference that represent the position and the language of hand gesture. Research projects about it can be found at [14], [15] and [16].

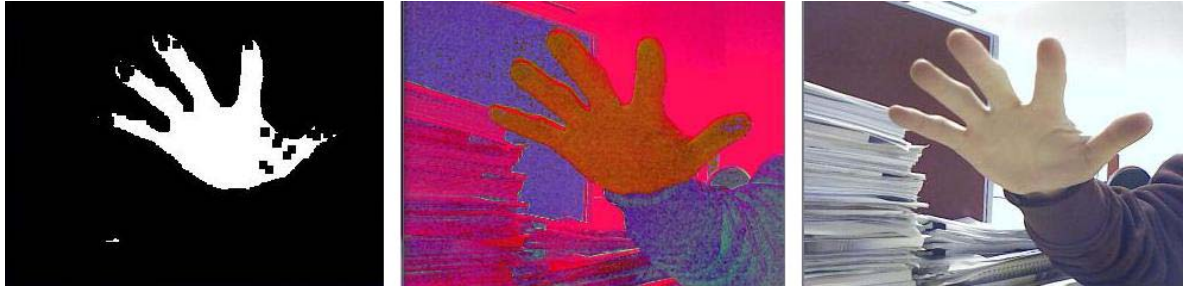


Fig.4. Skin color detection (Source: web page).

Some examples are: EMGU CV Hand Gesture[@Andol], Hand Vu [@HanduVu] and Eye Sight [@Eye Sight].

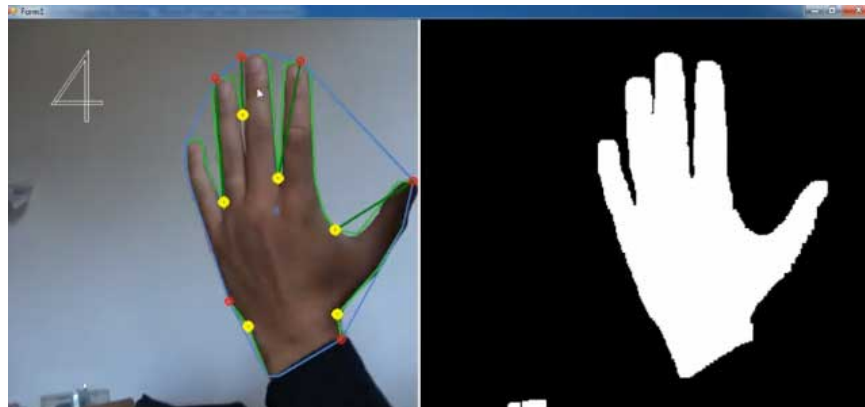


Fig.5. Emgu ,contour representation (Source: web page).

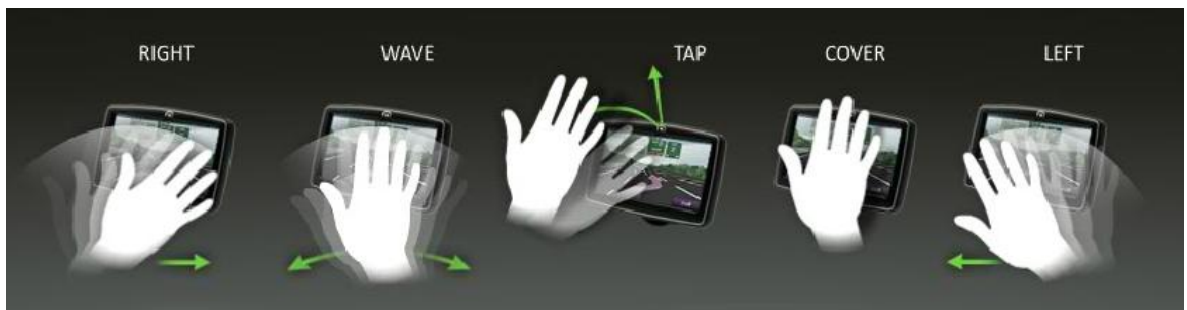


Fig.6. Eye Sight technologies (Source: web page).

A recent research project that applies this type of interaction is 'Sixth Sense' developed by Pranav Mistry from the Fluid Interface Group MIT [SixthSense, 17]. The component elements of the hardware are a camera and a tiny projector mounted in a pendant like wearable device, the software sees what the user see and projects information onto surfaces, walls, and physical objects around the user, so he can interact with the projected information through natural hand gestures (Fig.7), this is an attempt to free information from its confines by seamlessly integrating it with reality, in other words making the world surrounding a computer.

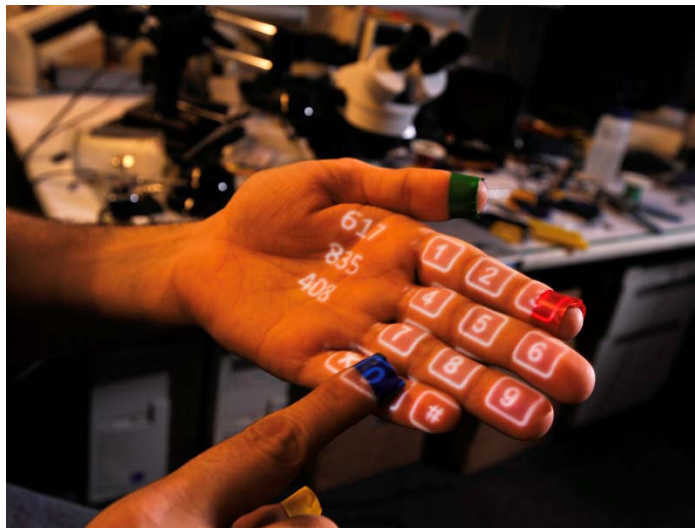


Fig.7. Using palm for dialing numbers (Source: web page).

The Infrared Distance Measure Method [18] consists on placing an array of infrared distance measuring devices on a surface and connected to a central computer, with this the position and direction of the hand can be precisely detected the space (Fig.8).

Infrared distance measuring devices [18] have an effective distance recognition of about 40 cm, with the precision range between 4 cm to 30 cm, this is compatible with general hand movements, nevertheless this can be consider as a limited method because it doesn't generate enough patrons of information for a complex language interaction.

One example of this type of devices is Kinect a motion sensing input device developed by Microsoft, it has been used for as a game controller mainly and it has claimed to be the fastest selling consumer electronic device.

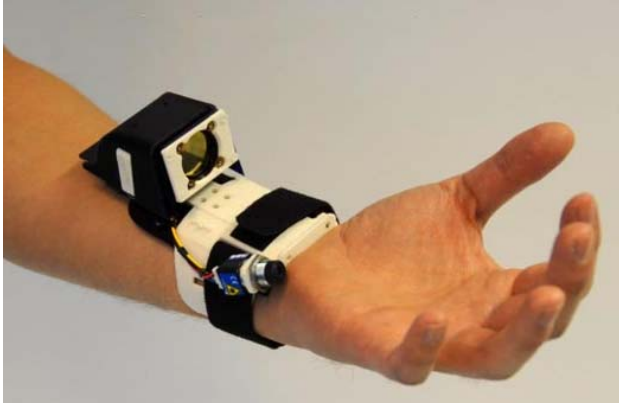


Fig.8. Using Infrared Devices, Newcastle University
[Source: web page].



Fig.9 Cyber Glove -Hand motion capture
[@Cyber Glove].

The third method and one of the most accurate deals with wearable gloves, which have incorporated a number of sensors, in some cases more than 30 (Fig.9). The sensors are distributed on the surface on the hand to capture a high resolution motion data so that the information can be read in a 3d coordinate system and later translated to recognize positional patterns, these devices can be wired or wireless depending on the power source. In some cases the use of an external extra sensor is used to track the global position of the hand.

Some examples are: The Cyber Glove II and III [@Cyber Glove], Shape Hand [@Shape Hand] and 5DT Data Glove 5 Ultra and 14 Ultra [@Data Glove].

2.2.3. Eye Motion

This type of input works with by tracking position of the pupil from an incoming sequence of images from a video camera, this information is then translated in coordinates of specific positions in a 2d coordinate system with the help of a software applications (Fig.10).

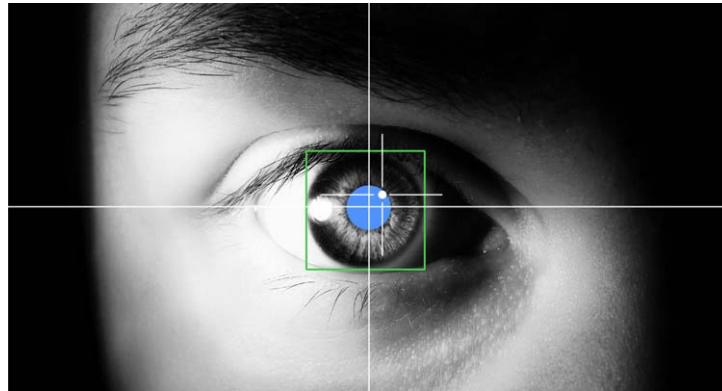


Fig.10 Pupil tracking (Source: web page).

The effectiveness of the tracking relies on a clear and dark contrast of the captured images of the pupil and a proper calibration of the hardware and software. Some methods of calibration use synchronized cameras to identify the position of the user's pupil. [26]

The tracking devices that make this possible can be incorporated as Head Mounted Devices (HMD) or can be Portable Devices located at a short distance from the viewer.

HMD Glasses (Fig.11) consist in a near - infrared LED light that illuminates the eyes and creates a dark pupil effect , then a video camera captures the movement of the eye through mirrors and glasses located to not to interrupt the view of the user. The size and the weight of this type of devices hast to be minimal to provide a more natural feeling for the interaction. Some products are: EyeWriter [@EyeWriter], VisionTrak [@VisionTrak] and Tobii Glasses [@Tobii].

Portable devices are not constrained by the distance and they don't have to be near the user's eye, these devices can be located at a certain distance enough to capture and recognize the movements of the pupil, and because they are not a wearable devices allow the user to move more freely in a motion tasks (Fig.12).

Some products are: S2 Eye Tracker [@Eye Tracker], Tobii X60 AND X120 [@Tobii] and Eye Tech Digital Systems [@Eye Tech].



Fig.11 HMD for eye tracking Tobii Glasses
(Source: web page).



Fig.12 Portable device for eye tracking
(Source: web page).

3ITU Gaze Tracker application is an open-source eye tracking application whose aim is to provide a low cost commercial gaze tracking systems. The software requires a camera equipped with infrared night vision, a standard video camera or a Webcam. It is a product developed by the Gaze Group at the IT University of Copenhagen [@3ITU Gaze].

2.2.4. Voice output

Also understood as Speech Synthesis it refers to the artificial production of human speech provided by an artificial artifact. Usually a computer system is used for this purpose where a TTS (Text to Speech) system converts normal language text into speech, this can also be called speech synthesizer and it can be implemented as an independent software application or as a plugging for an existing one (Fig13).

The synthesized speech requires a recorded data based of speech that could be stored and be recalled whenever needed to formulate sentences or commands for the communication interaction with a user. Important is that in this kind of systems both naturalness and intelligibility should be maximized to provide a more fluid interaction. Research projects can be found at [19] and [34]. Following some application will be briefly described.

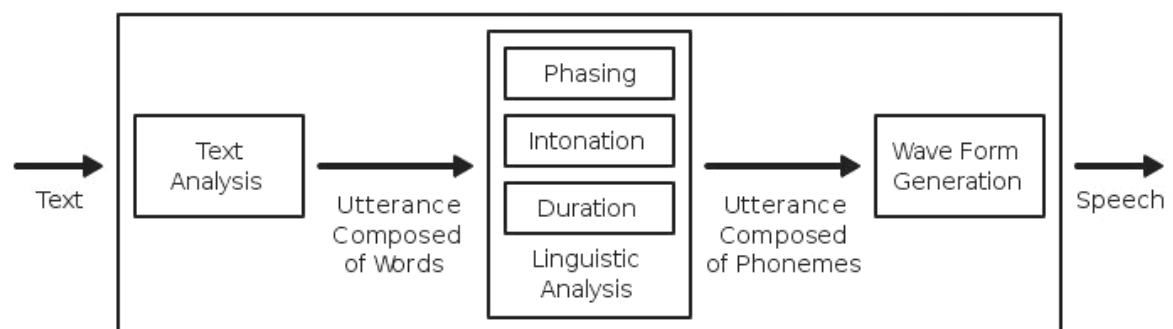


Fig.13 Flow diagram of speech synthesis (Source: web page).

VoiceOver is the speech system integrated to Apple Inc.'s Mac OS X, iOS and iPod operating systems [VoiceOve]. Speech Application Programming Interface (SAPI) is part of the system components of Windows Desktop that support speech synthesis and speech recognition, the products derived from it are: Narrator, Coolspeech, Textaloud, and Ultra Hal which can also be found in websites [MicrosoftTellme]. This all are part of the latest research on Natural User Interface (NUI) from Microsoft which focuses on speech synthesis as a medium for communication of future interfaces [MicrosoftResearch].

JAVA Speech API also known as SAPI is cross-platform and supports command and control recognizers, dictation systems, and speech synthesizers. Android is another application, which offers this type of functionality, developed by Linux-based operating system for mobile devices such as smart phones and tablet computers [Android].

Exist various applications that come in shapes of plugging or gadgets that one can find on the Internet, some of them are designed to assist tasks such as Web Browsing or e-mail administration and are incorporated on the Interface, others can be downloaded so the users can make free use of them for domestic purposes, some examples: Speech Synthesizer [Speech Synthesizer], eSpeak [eSpeak] and Model Talker [ModelTalker].

2.2.5. Haptics

Also know as tactile sensing, it refers to the feedback devices that allow users to receive information in the form of a self sensation on some part of the body. This kind of interaction allows the awareness of physical presence of digital representations and can enable users to feel material elements of an interface even when this doesn't exist on reality such as panels and switches. This doesn't mean that the entire shape and texture of objects has to be simulated but at least some feedback of feeling could be interpreted as the presence of a real object [20].



Fig.14. Cyber Touch Technologies
(Source: web page).

The technologies that can make this possible are wearable devices for hands also called Haptic Gloves; some examples are briefly described as follows:

CyberTouch Systems Developed by Cyberglove which offers a vibro tactile feedback that creates a realistic environment for users enabling them to experience how a virtual object moves and responds to interaction in real time (Fig.14) The CyberTouch system consists of six small, lightweight vibro -tactile actuators; one on each finger and the palm of the glove, the actuators can generate pulses, sustained vibration, or customized vibration patterns.

Along with the CyberGrasp and the CyberForce this devices can simulated real physical conditions of objects to be perceived by the Hands [@CyberTouch].



Fig.15 P5 Glove [@P5 Glove]



Fig.16. CyberGlove - University of Cornell
[@CyberGlove]

P5 Glove [@P5 Glove] is based on bend sensors and remote tracking devices, The software system sends packets of information to each finger to generate the bending force which contains values in a 3d coordinated system. It uses A 6 degrees mode of tracking (X, Y, Z, Yaw, Pitch and Roll) to ensure realistic movement for a more intuitive interaction (Fig.15).

Cyber Glove (Fig.16) is the product of a research project from the University of Cornell which has developed a glove that provides tactile feedback of virtual objects and a sensation of holding virtual objects; this also includes temperature, weight and size of the object [@Cyber Glove].

2.2.6. Visual Display

This is the most important feedback for the HCI, because it can replace the other two modes of interacting feedback by providing enough visual information to the user for understanding the 3d geometries in a 3d digital space.

There are various types of displays available but for purposes of commodity and to provide a more natural interaction for the user the review will review technologies such as HMD and wearable devices that can be consider as an extension to the body.

Suterland's Classification describe this type of displays as Optical See through, they leave the real world perception as it is and display the AR overlay by means of transparent surfaces and lenses. This allows the user to interact with both real environment and digital objects during the design process. One of the advantages is that the user can use its own body to have reference of scale during the design process [10].

These See Through visualization can also be well applied for purposes of Collaborative design considering that two users need to establish work sessions from different locations , In other words this system would provide visual relation for the workflow to happen.

OLED (Organic Light-Emitting Diode) Displays is one of the latest technologies that can be used to create transparent displays (Fig17), these are bright and efficient with a fast response time and a wide viewing angle display. This is made from organic materials (carbon based) that emit light when electricity is applied. This type of surfaces can be made very thin (the thinnest prototype is 50 microns) transparent and even flexible.

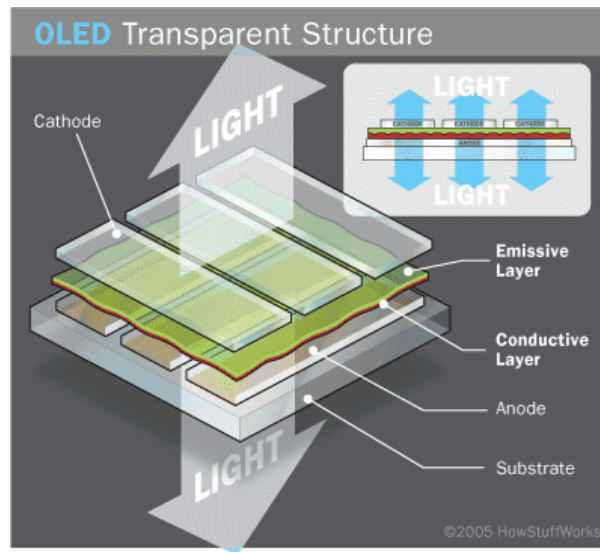


Fig.17. OLED Technologies (Source: web page).

Among the companies that will start implementing this technology in the development of display products are: Samsung which has already developed prototypes of computer Laptop which use these types of translucent screens [31] and Apple with a new product called Apple See-Through Augmented Reality HMD Glasses [32].

The Bionic Contact Lenses is another latest technology developed by a research group at the University of Washington in Seattle; they have created contact lenses with built-in electronics and an LED, powered wirelessly. Much of the hardware is semitransparent so that wearers can navigate their surroundings without visual interference which could cause disorientation.

This Bionic contact lenses (Fig.18) fall into the category of nanotechnology since the manufacturing processes for the circuits require manipulation at the sub-microscopic level but in any case the technologies needed to build functional contact lenses are in place and it will be a matter of time until its performance is improved for future manufacture [30] and commercial distribution [21].

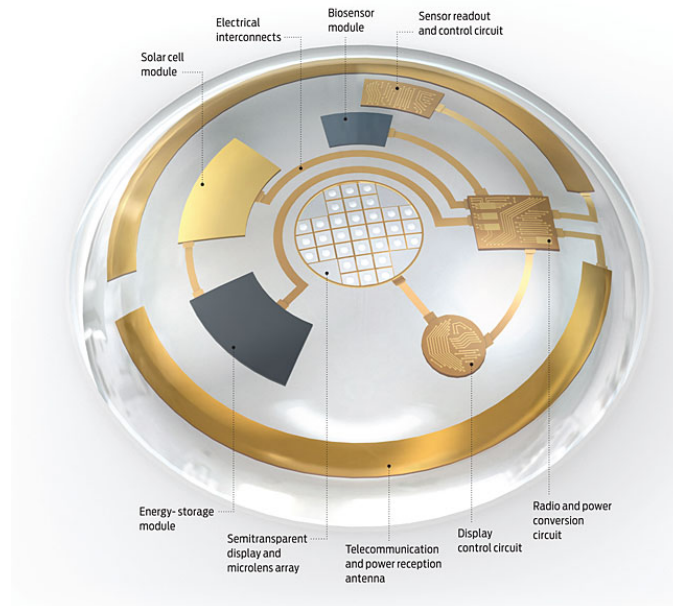


Fig.18 OLED Technologies (Source: web page).

2.3. Applications

Among the most recent and promising technologies that host multimodal interaction are Mixed Reality interfaces, in its general definition they merge the virtual and the real worlds into one new environment where both real and computer generated graphics can be perceived for interaction in real time. In this section I will briefly mention the origins of Virtual Reality (Fig.19) and Virtual Augmented technologies (Fig.20) by assuming this as the new platform on the evolution of Architectural design tools.



Fig. 19 Virtual Reality (Source: web page).

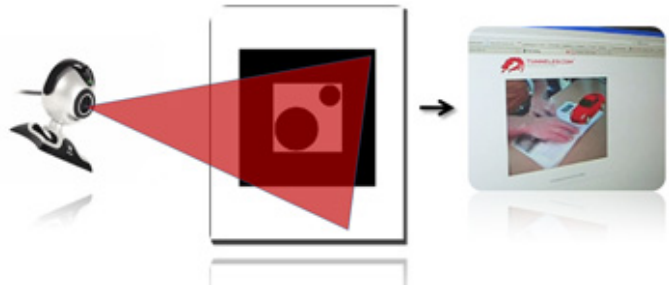


Fig. 20 Augmented Reality (Source: web page).

Virtual reality (VR) in its wide definition is the technology that simulates geometrical objects and environments via three-dimensional computer models, so the feeling of visual realism is achieved through a high-resolution, stereoscopic images [35]. There are two ways of VR, one allows viewers to project themselves into the scene, it is called a Virtual Environment (VE) and in the second one the scene is shown in full scale and viewers are surrounded by 3d images this one is called Immersive Environment.

In VR the sense of presence doesn't consider some of the human senses such as smell or taste, it focuses mainly in virtual displays , so the conditions to provide this feeling of presence deal with visual perception: image, quality, image dimension and view distance [12].

In the recent years Virtual reality has become a tool that can serve for manufacturing processes such as product design as well as simulation. Some other applications such as Electronic Design Automation, CAD, Finite Element Analysis, and Computer Aided Manufacturing are link to the naval, aerospace, and automotive industries. Beyond that 3D computer graphics techniques are currently used in research and development for medical purposes and training.

This interactive digital medium incorporates the use of powerful computer along with displays devices, sensors, electronic headsets, and gloves. The possibilities and applications are limitless now that more and more computing systems are being introduced to the daily routines so the idea of introducing the user in a virtual world has the potential of become the next dominant form interacting with the digital world.

Augmented Reality is part of the general area of Mixed Reality on the Reality-Virtuality continuum by Milgram and Kishino (Fig.21).

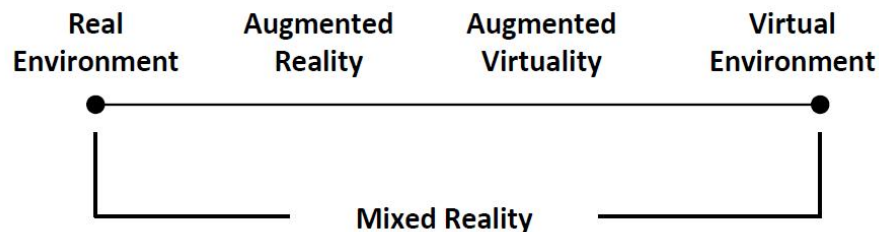


Fig. 21 Reality-Virtuality continuum [8,29]

The Augmented Reality has been classified in fourth major categories: Video-Based AR, Mobile AR, Tangible User Interface AR, and Spatial AR. Each category allows users to experience the virtual world with different methods of interaction; some examples of applications are related to medical uses [4], manufacturing and industry [9], design sector [22] and ubiquitous visualization.

The terms Tangible User Interfaces (TUIs) refer to a system in which the users can grasp, control and manipulate a real-world object as a means of manipulating virtual data. The relationship between the user and the computer generated object is often one to one scaled and happens through shape and positional recognition; the benefits are reducing cognitive load and a shallow learning curve [23]. An example of this type of application is the free software AR Toolkit 2006, where the virtual object is associated with a printed black hollow square marker, so when the marker changes position and orientation so does the virtual object with a constant update to provide the real time interaction.

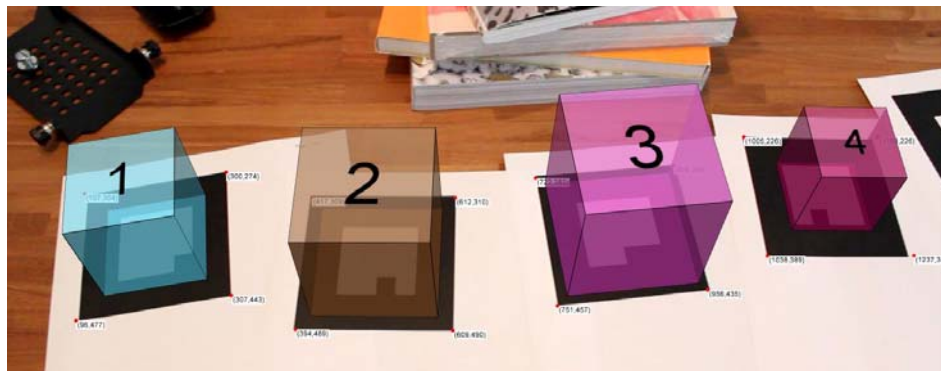


Fig. 22 Augmented Reality Markers (Source: web page).

In Spatial Augmented Reality (SAR) the limitations of using “Traditional” Head-mounted or Hand-Held displays are being dismissed by the use of large spatially aligned optical elements, such as half-silvered glass, mirror beam-splitters, transparent screens, holograms or projectors in order to deal with the technological limitations. These systems are often site specific and expensive, requiring extensive setup and maintenance; nevertheless their increasing interest from part of education and science is changing that scenario.

Since the introduction of AR to the computer science community many organizations were formed and after a few years founded The International Symposium on Mixed and Augmented Reality (ISMAR) which has become the major symposium for industry and research to exchange problems and solutions.

Late AR interfaces are being developed for extending their context, with the introduction of Ubiquitous Computing many of the processes and information that are displayed in the desktop, are being displaced into the objects that make the environment, with this the communication between people and information processes happened in wider field of view where any given surface can be a potential screen and 3d computer generated elements can happen everywhere (Fig.22), all this can be possible with the use of Special lenses.

The gaming industry has become one of the fastest growing and most economically successful kinds of software [24] and is one of the most influential when it comes to HCI, knowing this it's easy to understand the big interest that this industry has to reach various levels of dynamism in order to provide a high level of entertainment to the consumers.

When looking at the entertainment industry one of the first graphical video games, Space War by Slug Russel, included the first computer joystick a two dimensional control device that allowed the users a new physical interaction, since that the evolution of hand controllers changed dramatically introducing various types of joysticks by companies such as Sega, Taito and Atari.



Fig.23 Wii User Interaction
(Source: web page).



Fig.24 Wii Game Controllers
(Source: web page).

The latest game consoles such as the Wii produced by Nintendo provide a highly efficient real time interaction with accelerometers and infrared light sensors that allow a real time interaction with the digital world (Fig.23) (Fig24). This is a sign of the increasing interest in Game developing from the perspective of HCI. Some researches include 3D graphics and 3D sound technologies to enhance engagement and immersion in game playing. With this the race for reaching quality Visual and audio elements for video games are being subject of studies. Video Game developers take that seriously in order to provide the best condition for player performance and addiction [24].

The next generation of gaming interfaces included various alternative means of input and output with devices that incorporate gesture tracking technologies, 3D sound and Virtual environment to enhance engagement and immersion in game playing.

Besides al mentioned above I would like to point some of the inspirational concepts that come from Science Fiction, I consider them as an important visionary reference which have influenced engineers and developers that drive the evolution of the HCI since the beginning of computing technologies. These ideas came from the film industry; I will mention some of them briefly.

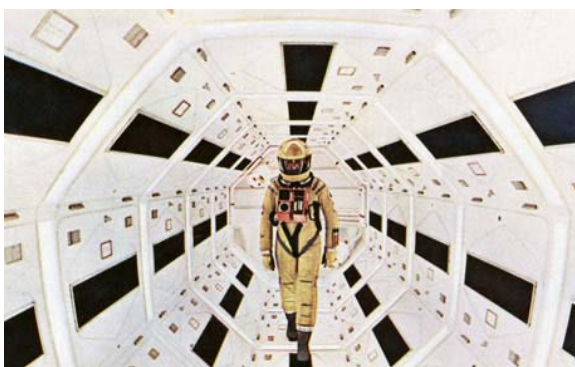


Fig.25 2001: A Space Odyssey, Kubrick, 1968
(Source: web page)



Fig.26 Holodeck, Star Trek ,Roddenberry ,1987
(Source: web page)

In the Science fiction film “A Space Odyssey” by Stanley Kubrick (1968) it is showed a new concept for HCI, in which crewmembers of a Spaceship interact with the Ship’s Computer by talking to it, this one has no mouse or keyboard so the only way of communicating is conversationally (Fig.25), Other example is the famous Holodeck from Star Trek by Gene Roddenberry (1987) where an immersive virtual reality interface allows the characters to interact physically with fictional environments created for running simulated missions (Fig26).



Fig.27 HIC, Iron Man, Favreau, 2011 (Source: web page).

A latest example comes from the Holographic Interface developed by Prologue in the movie Iron Man from Marvel Comics (2011) and Directed by John Favreau. In this film the main character Tony Stark the lead character is able to interact and manipulate digital holographic interfaces which are rich in information visualization, he uses hand gestures and speech recognition to develop, explore and modify engineering designs, (Fig.27) this shows the potential of the communication between a CAAD system and a user in a more intuitive and even enjoyable way. Some other alike UI can be found in films such as “Sunshine” and “Minority report”.

3. Requirements for CAAD systems

Assuming that MMHCI can be integrated into the CAAD systems, the next step is to define how this interaction would take place, therefore it is important to establish first the elements involved as a requirements, from a CAAD point of view.

To define these requirements the approach taken is based on the definition of a specific design routine and the elements involve, this implies the workflow when performing building object generation and modification, for example creating and placing a wall in a 3d file. Defining the types of data and tools can be useful for understanding the communication language that takes place on this interaction. For further development it is important to consider is the use of BIM libraries and form generation based on Parametric Design.

The actual software applications designed for CAAD work with a Graphical User Interface platform which is also the base of HCI methods for personal computers, this standardization allows users to interact with the digital information by visual recognition of the information when working with computer systems. It has been proved in many studies that the recognition of graphics and icons is the easiest way to learn and remember how to use a software program and in architecture that is not an exception. Establishing the basic components of existing applications will be the starting point of defining the requirements.

A previous analysis of Architectural CAD based applications was made previously in order to identifying the types of information involved. The software examined where: ArchiCad [ArchiCad], Revit Architecture [Revit] and Sketchup [Sketchup]. Nevertheless users in the field of architecture also tend to explore other applications beside the mentioned ones, either for developing complex geometries or because of the need of specific functionalities that can not be found or are complicated to perform in a conventional CAAD application, this will also be analysed.

3D model representation is being widely used not only in Architecture , the use of these type of applications is widespread in various fields such as gaming industry, medical visualization , engineering as well as in film and animation industry [@3dMax,@Maya,@Cinema4d,@Rhino3d,@Blender]. Some of these applications are listed below:

Developer	Product	Special Functionalities
Alias Systems Corporation	Maya	3D Modeling, 3D Animation Dynamics & Effects, Rendering & Imaging Pipeline Integration
Autodesk	3d Max	3D Modeling, 3D Animation Dynamics & Effects, Rendering & Imaging Simulation
Maxon	C4D	3DModeling, 3D Animation, Texturing,Rendering
Mc Neel	Rhinoceros 3d	3DModeling, Simulation
Blender	Blender	3D Modeling, 3D Animation , Rendering & Imaging
Last Software	Sketchup	Push/pull or modeling technique, 3d warehouse, Direct Connection with Google earth.

Table2. Software Characteristics

3.1 Types of Data

When working with any of the applications mentioned above, the user deals with different kinds of information organized as follows:

2d Geometry; are the objects represented in 2 Dimensions (Fig.28,29), these are an important element of the WIMP interaction paradigm when working with planar

retrieval such as Computer screens, each of these elements can be organized and place according to user preferences in many software products.

For visual representation of information these are very useful now that these can show only relevant information when working on different functionalities or parts of a project e.g. detail drawing.

- 2d Drawings: Sections, Plans, Elevations, Details, Diagrams.
- Tool Icons
- Interface Windows : Palettes, Ribbons
- Attribute Windows : Material Bitmap, Object Properties

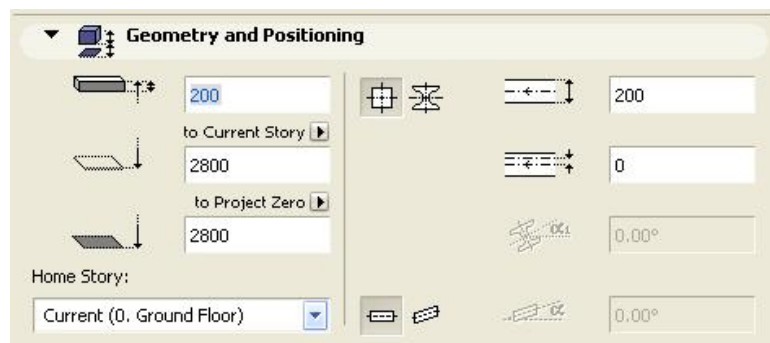


Fig.28 Beam Properties Chart [@ArchiCad]

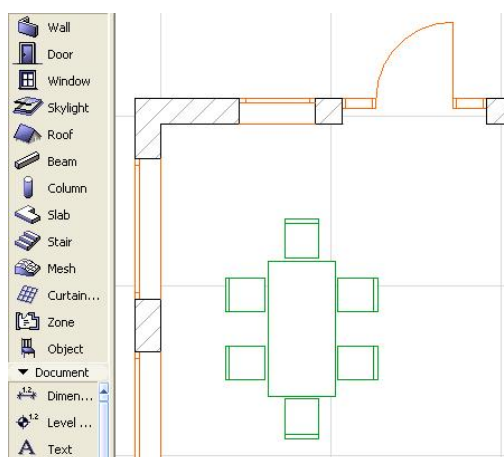


Fig.29 2D Graphical Representation [@Archicad]

Textual Reference, is related to the written language which allows the users to use a set of alphabetical symbols to generate words, commands and Sentences to communicate with the computers (Fig.30); allows the users to identify the different tools provided by the GUI; it is also a method for saving information. From the side of the software itself it allows the possibility to communicate with the user either with the use of status windows or command bars. Some of these elements can be listed as follows:

- Document Administration
- Layer Administration
- Navigator view Administration
- BIM libraries
- Parametric Control
- Object Properties
- Referential coordinates, distances, Location
- Textual Tools representation

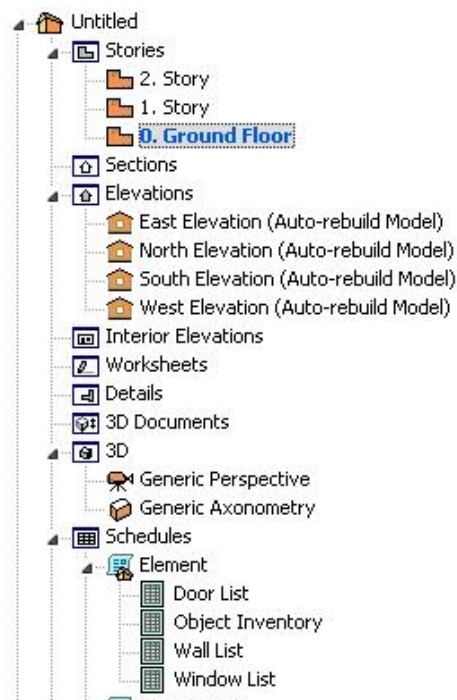


Fig.30 Navigation Project Map - [@ Archicad]

And 3d Geometry, which is the most important now that it constitutes the platform of the interacting process, these are contained in a 3d representation of the digital space (Fig.31,32). This term is understood as a representation because it only exists on the digital world which is only possible thanks to of complex algorithms and scripts which mimic a three dimensional space. The elements can be grouped by its geometrical properties and also according to labeling as in Building Components.

- Geometric Polygonal Components:
Planes, edges, Points
- Geometric Nurbs: Lines, points, control points
- Primitives Objects: Box, Sphere, Plane,etc.
- Building components: Wall, Window, Slab, Column, Beam
- Groups , Families, Parents, Childs

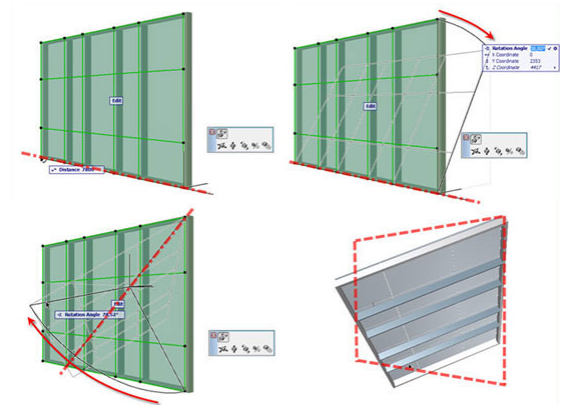


Fig.31 BIM Component [@Archicad]



Fig.32 3D Building Model [@Archicad]

3.2. Workflow description

CAAD workflow is defined by a series of small tasks lined up and sufficient for the generation or modification of a building component, this can be geometrical as well as BIM object. In this definition in order to achieve a complex design the number of steps increases and so does the number of decisions that the user has to take. When performing these tasks in a sequential way even a small variation of a workflow can lead to a different result, this nevertheless depends on the type of CAAD applications and preferences between users. In this document a specific design routine of the generation of a building element in this case a “wall” will be introduced as an example of implementation of MMHCI.

The main goal is defined as the creation of a wall, this consist of a number set of different tasks that the user has to achieve in order to elaborate a complete 3d computer generated model of a building, There are levels of sequential happenings which need to take place to achieve a given goal, this implies that exist an interdependence of each of the elements in play to the whole process and to the adjacent ones, with this even small actions can not be taken as isolated process.

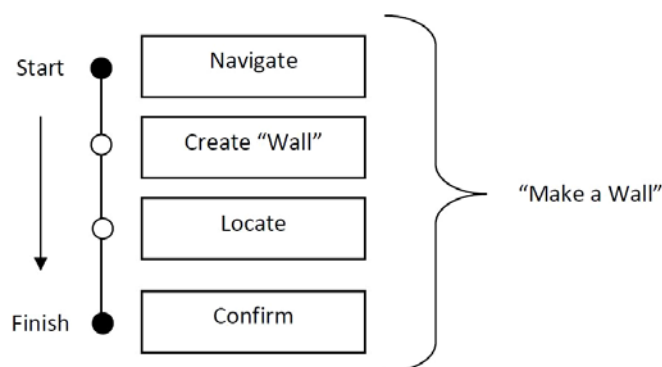


Fig.33 Design routine “Make a wall”

In the Diagram in Fig.33 each of the boxes represent an action to be performed by the user by activating specific tools, by doing so the user can proceed to navigate, create and manipulate geometries in the 3d Space. Here is where multimodal interaction

takes place by allowing the user to use different modes of communication to interact with an accurate 3d representation of the geometric information.

Each of the steps is described as follow:

Navigate: The user locates him self in the space this can happen by moving the working space with the hand, the purpose is to find a preferable point of view on the 3d space for working. If the wall is meant to be place in an existing 3d model the user may turn of some layers, groups or geometries to have better visibility as well as explore possible views by circular orbiting.

Create wall: The user can recall this command with the voice and obtain a default geometry or a default wall element to work with, this doesn't mean that is the final product but just an element that is can be modified, for example by scaling.

Locate: The user places the primitive wall component on the right position via direct manipulation; here the system can automatically activate Snapping on points for a more precise manipulation of the geometries and the positioning on a 3 axes coordinate system. If working within a grid the system can provide with snapping on Vertex or edges of the grid.

Done: This is necessary to inform the system that a specific goal has been achieved in order to start with the next one, by deselecting or deleting the system automatically jump into a different mode which can be the start point for a new process, for instance navigate.

4. Design task network for Multimodal CAAD systems

4.1 Design process with Multimodal Interaction Interface

When working with a Multimodal interface is important to define the design process as a sequential series of steps or Stages that the user has to go through to perform a given design task, these stages can be understood as intersection points between design intentions of the user and the tools provided by any CAD application, for example “Navigate” taken as node is where the user intentions meet the tool that allows exploring a 3d object by using hand gestures and voice commands such as “Orbit around”. Like this many tools can be grouped according to the functionality and be taken as general nodes of the interaction. Recalling each node would imply the starting point on a specific sequence of actions so the number of times they are recalled is greater than the elements each one contains.

When reviewing the approach taken by different authors and research groups in the field of HCI it was evident the variety of interaction methods used to communicate with the computer applications, in some cases like in *Augmented Tangential Reality*, real elements like 2d markers are been used to recognize specific functionalities, these can also be used as referential elements on the 3d space. In other cases the user is able to recall tools by pronouncing commands via Voice Recognition applications. By looking at the examples mentioned, the different kinds of inputs happening in a sequential way, could generate a large number of possible interpretations and generate confusion during to communication process with a CAAD system. Organizing the different functionalities of CAAD application in a Design Task Network would facilitate the reading of this type of interaction and help avoiding possible confusion of communication between users and CAAD applications. This could also be helpful for software developers for analyzing and evaluating Multimodal Interfaces.

4.2 Interface definition for CAAD systems

The interaction between users and the digital information happens within a sequential set of steps, this can be called " design routine"; in each step the user either introduces information or receives information from the system. For example when creating a wall, the user chooses to recall the Command "Create Wall" with the Speech input channel and later with the hands or any other physical input to the system the user can place the Wall geometry on the 3d workspace. Along the process the retrieval could always happened with the Vision Output Channel and also be supported with a speech synthesis report in case collision between objects takes place.

By grouping the various functionalities that CAAD systems offer in a set of fewer elements in this case "Nodes in interaction", each time the users enter to a certain node the interface is able to recognize and provide the correct interpretation of an input and also provide with the right feedback, in this way the communication could take place in a more fluid way. This Nodes of Interaction can be useful for differentiating input interpretations by the computational systems, this would make possible that the movement of a hand which in one node means "orbit around" in another could mean "Navigate" for instance.

In [9] the authors define design task network based on a set of 10 nodes which are consider the most relevant for a product shape design, each one of these nodes allows the user to have access to a more specific set of tools , the user can travel within them to accomplish a given 3d basic design. Based on that definition and considering a previous analysis done on software applications for ACAD a Design Task Network ,containing a set of 9 nodes mapped (Fig.34), is been proposed. This Network is based on standard functionalities and requirements of ACAD systems in order to provide a more coherent way of interacting and reading of the information provided by MMHCI.

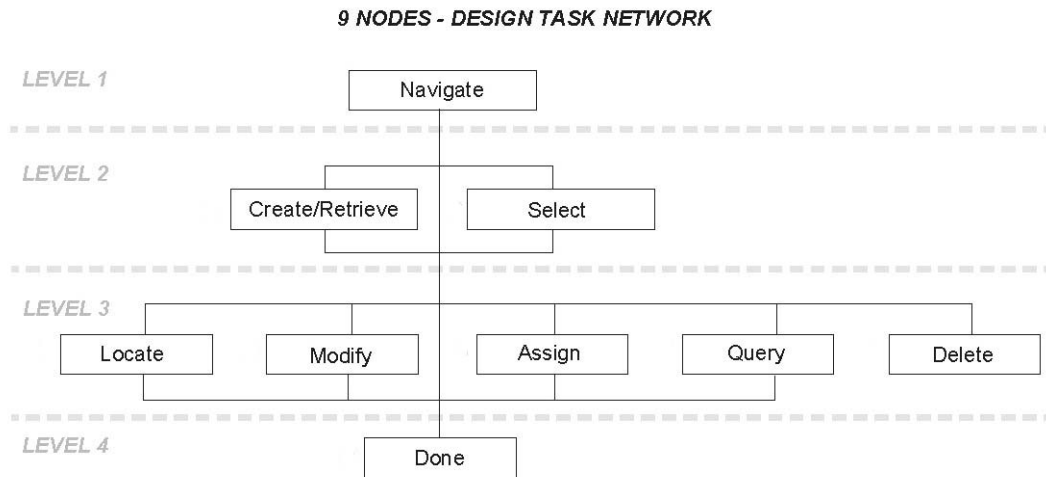


Fig.34 Nine Nodes Design Task Network based on [9]

The starting point at Level 1 is located the node “Navigate” so that the user can first locate himself on the workspace or search for the specific location on a given project to work on.

In the next Level 2 are either “Create/Retrieve” or “Select” a 3d existing object, this means the first geometrical input to the system.

Level 3 is where most of the design decisions take place because is where the biggest set of tools is located for the example “Modify” allows the user to perform all kind of geometrical modifications this also means that from each node a new branching of tools can be categorized in groups as in : resize, duplicate, Boolean Operations, Polygonal editing among others; “Locate” allows the user to move and place elements on the working space; “ Assign” refers to label or specify an ID to the elements created; “Delete” eliminates a selected element form the working space; “Query” is where the user can look at specific properties of the elements such as dimensions or construction materials

In Level 4 “Done” stands for confirmation of an action and turn off the selection to work with a different 3d object or start a new navigation.

The Design Network (Fig.34) has been developed to contain different types of design routines within its 9 Nodes of Interaction; each possible path allows the user to achieve total control on 3d modeling routines. The example at (Fig.35) deals with the creation of a building element in this case a “Wall”, a path that travels between some of the nodes of the Design network can be traced as follows:

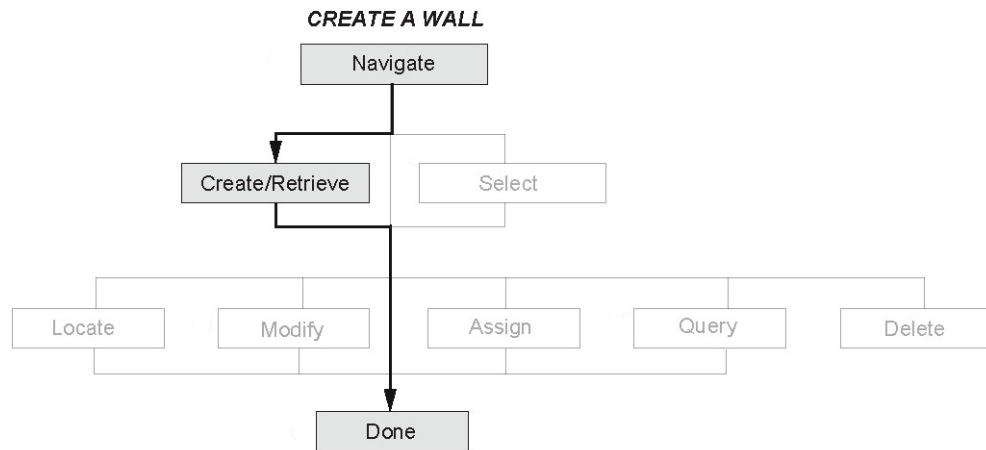
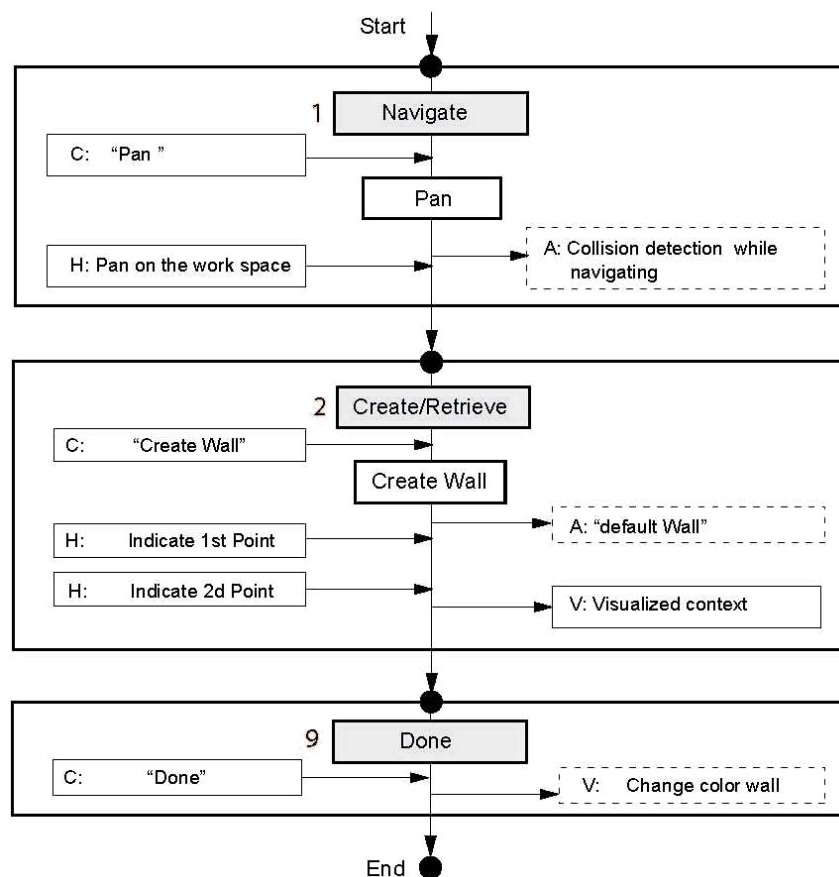


Fig.35 Design Task Network “Create a Wall”

In order to analyze more in deep how the Design Task Network functions on a design process a more detail sequence of steps which also include information such as channels of communication is being described on a the diagrammatic way (Fig.36), The purpose is also to establish a language for reading this type of interaction also based on [9] and to explain the way the different elements of a Multimodal Interface take place, the sequence of events and the process of communication between user and the computational systems.



Interaction Mechanism

V: Visual Output
A: Auditory Output
T: Tactile/ force feedback Output
E: Eye motion Input
C: Voice Command Input
H: Hand motion/ Gesture Input

[] Node
 [] Operation
 [] Primary interaction mode
 [] Subsidiary interaction mode

↓ Workflow Vector
 ← Input
 ← Output
 ● Enter/Exit Node

Fig.36 "Create a Wall" workflow

In each of the Interaction Nodes a number of alternatives are provided by the system so when the user jumps into the next node another set of choices are open and like this the number of levels can vary depending on the type of tool selected and the specificity of a tasks. For example in the Locate node the user chooses "Move" then he has to decide, to chose in which direction "x, y, z", all this information can be introduced to the system either by voice recognition or with direct manipulation of the geometries. In this type of interaction the user has to be aware that any action can be detected so the concentration is crucial for a fluid communication with the system,

from the computational part the feed back can be constant or set according to user preferences, for example “Collision detection “reports or “reading of coordinates” for location.

In this type of Interface there is a Primary Interaction mode which makes possible the direct design routine and a Subsidiary Interaction mode which supports the design and usually takes place with the Channels that are not used or are not necessary for the Primary interaction mode to be processed. These two ways of interacting can be programmed by the user and respond to individual preferences in order to make the design more human centered.

A set of hand standard gestures (Fig.37) extracted from the various sources and real applications [26, 27, and 28] enable the physical communication by allowing the user to have control of the tools and navigation provided by the Multimodal Interface.

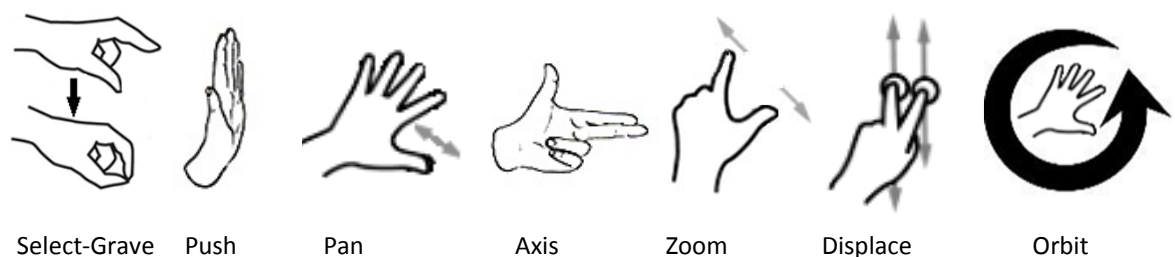


Fig.37 Operate virtual objects with Hand gestures [26,27,28 and@Apple]

Understanding and recording the way the user travels from node to node can be useful on the development of a new kind of Interface now that multiple ways of interaction can speed up the process of traveling between nodes, nevertheless confusion when processing the multiple input data may rise. A Design Task Network that and organizes and maps this information in an readable way can be useful for study and analyzing this interaction, later the information gather can be valuable for purposes of evaluation of prototypes when developing a Multimodal interface for the future.

4.3 Design process visualization

A way to organize and visualize this type of interaction is also been proposed on this document, it consist on 3 concentric rings, each one dealing with different information and organized one inside of another allows the user to have a visual feed back for awareness of where he is located on a design process.

The first ring contains the general Intensions of the user defined previously as the 9 main nodes 9 for CAAD interaction. The middle ring or 'Palette of Tools' is where the user specifies what he wants to do based on which tool he selects, in this ring is where the interface has already one information from the previous ring which can be used for providing interpretation for the next steps; the most inside ring or 'Count of Actions', is where the actions that deal with the generation or modification of geometrical entities take place by the means of real time Interaction using any of the 3 channels for interaction (Fig.38).

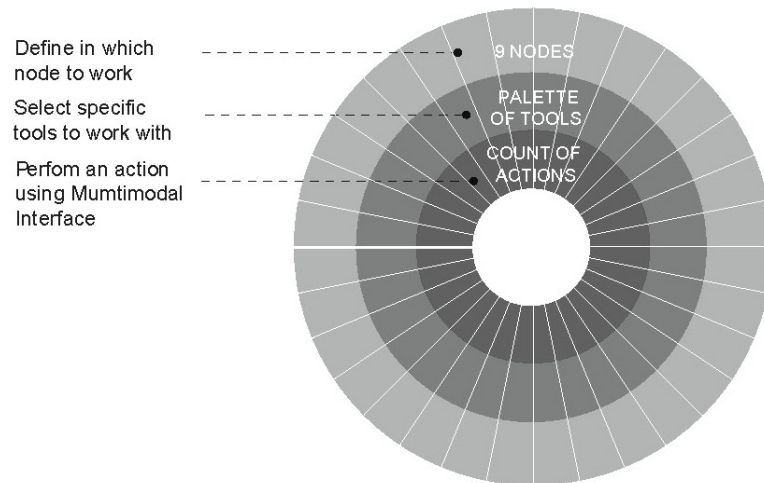


Fig.38 Concentric design task network

When moving within the rings each crossing means transferring information to the system for the next step, this because before and after crossing the channels that is being used can be associated for recalling one of the 9 Nodes , recalling specific tools, or performing a real time modification on a 3d geometry.

This workflow can be visualized with a path that travels from outside to inside of the Rings in an iterative process depending the type and the complexity of a given task. The path (poly line) connects dots located on each of the stages (Fig.39). Each dot is the representation of an event input or retrieval, the lines that connect the dots would represent a linear workflow, also each intersection of the path and the boundaries can be represented with color code dots from the different Channels of Interaction.

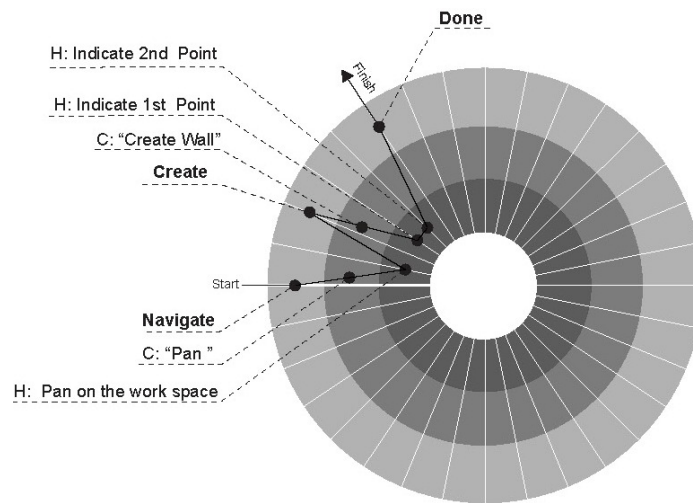


Fig.39 Concentric design task network workflow

One Path traveling condition is that the user has to go through each of the rings when traveling to the center (Fig.40). And from any or the in inside rings the user can jump to the most outside ring. Like this more rings could be added as each task routine become more complex or has more elements in play, but for this document only 3 rings or stages will be consider.

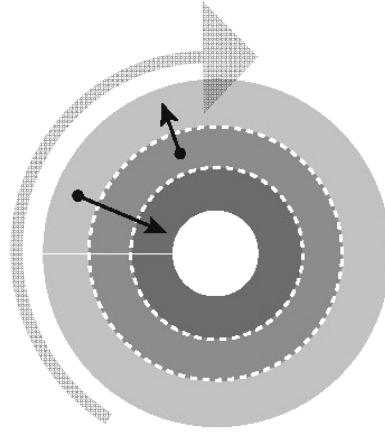


Fig.40 Path traveling definition

A complete revolution which means the 360 degrees of a circular path trajectory will represent a complete process, nevertheless this interpretation can not be truly assume on this type of process now that for an design point of view any given process doesn't have an ending but only a stop point.

The workflow path can show variations as we can see in the three diagrams from (Fig.41) sometimes the user could perform a low number of actions for example place a window, in a) the three Operation Modes could have the same number of dots, in b) the user could also realize more selection of tools from a same Node Root, in c) the user explores constantly the work by NAVIGATING after each PERFORM ACTION.

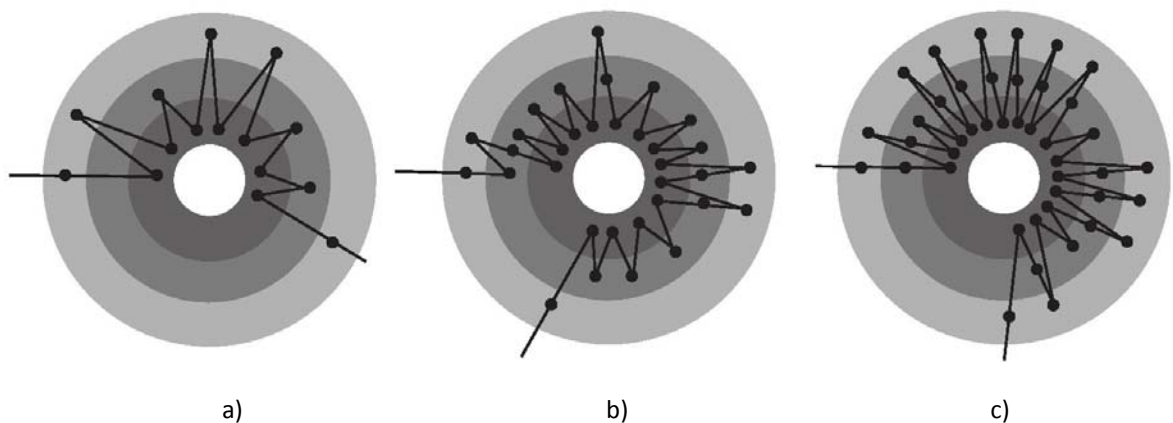


Fig41. Path traveling variations

Multi-modal Interaction allows users to use a mode to perform a task and another to perform a parallel task or support the one in process. For Example the user can set navigation with Voice Command for input and used the Hands for placing a Building element when doing so the recorded data can be taken as a starting point to establish performance indicators when working this type of Multimodal Interface, and later maybe even helpful to find semantic interpretations. [14]

Like in actual software application certain modes of Interaction of any of the different channels can be previously assigned for each of the 3 Stages of the Network, for example when the user has some kind of preference for specific channels. On the other hand design processes can be recorded by the system which means that the system could establish a certain criteria based on behavioral information of the user, and apply these preferences to facilitate the communication. All this information could also be represented as tables or in a linear organization of events to be later used for HCI analysis.

4.4 Interface mockup

4.4.1 “Create a Wall” workflow

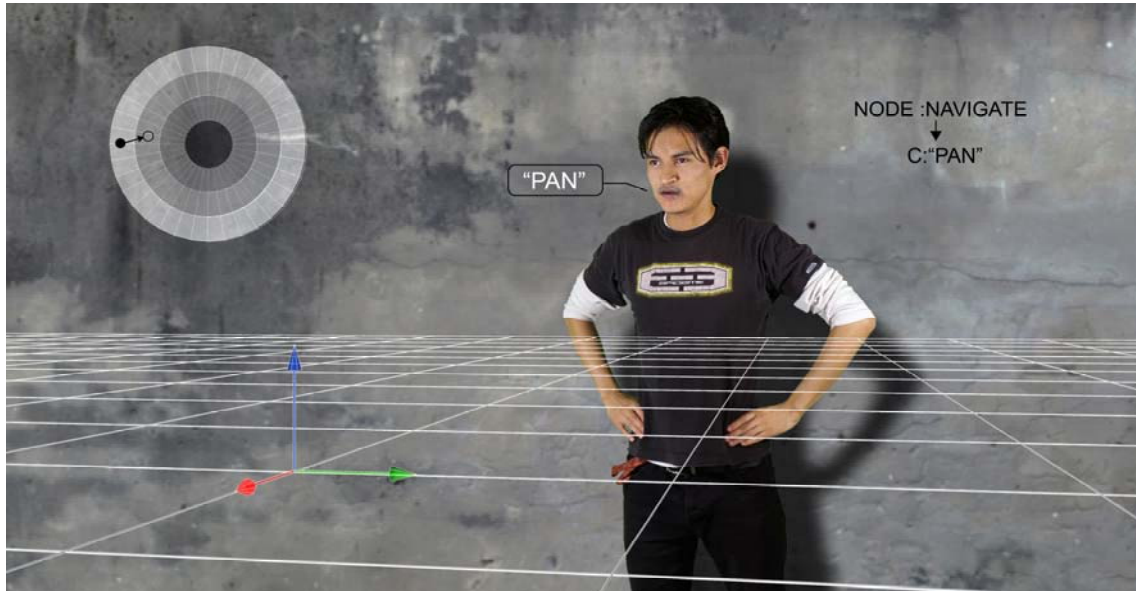


Fig.42 Interface Mockup - Set “Navigation”

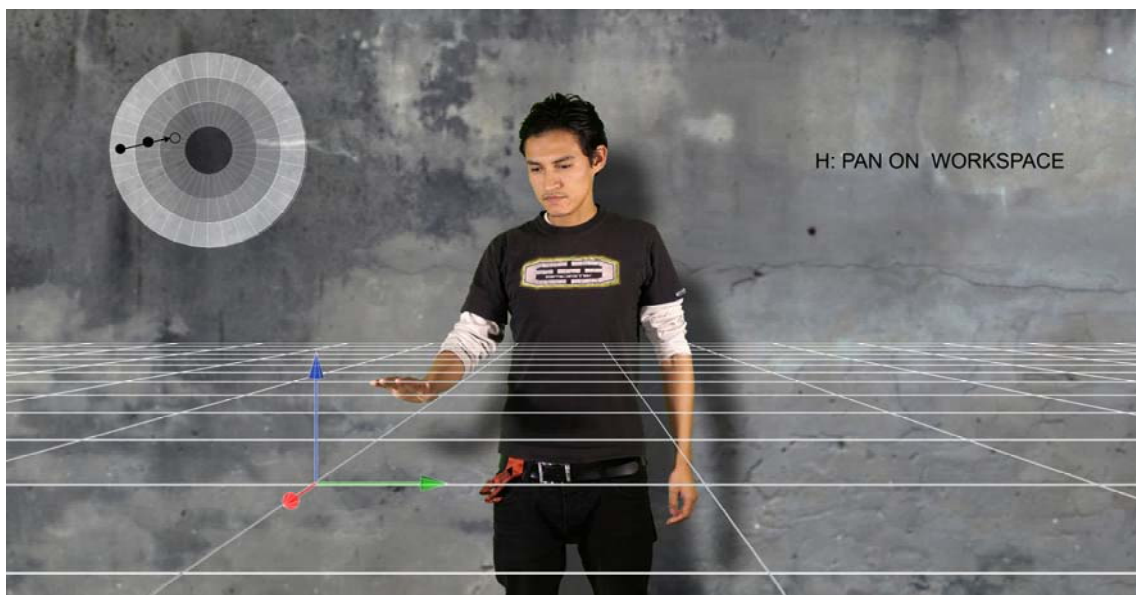


Fig.43 Interface Mockup - “Pan” on workspace



Fig.44 Interface Mockup - Set " Create wall"

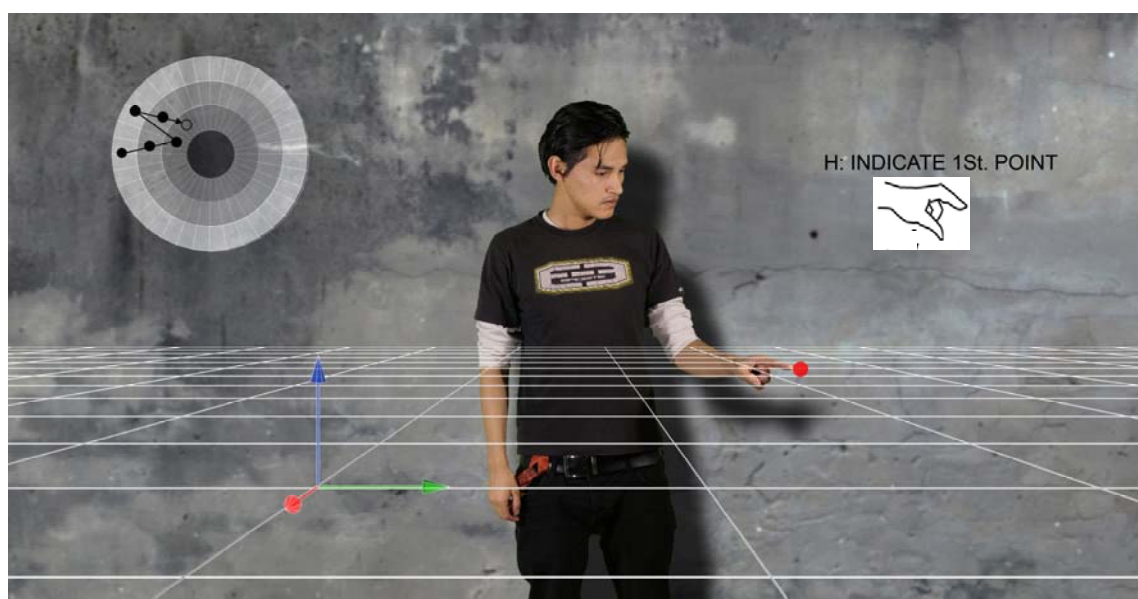


Fig.45 Interface Mockup – Indicate 1st. point

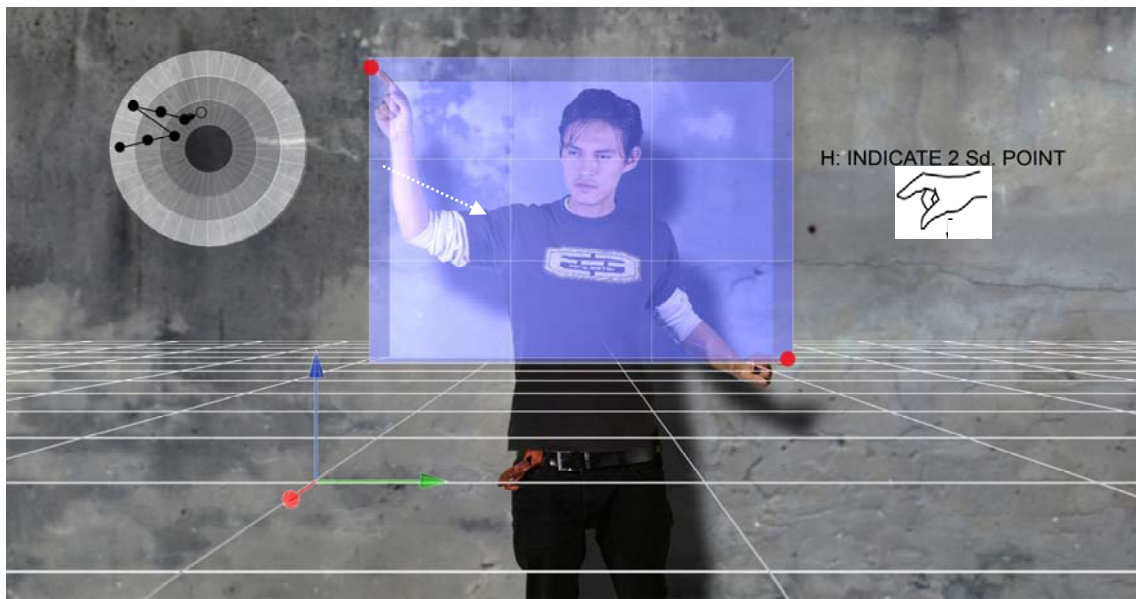


Fig.46 Interface Mockup – Indicate 2nd. point.

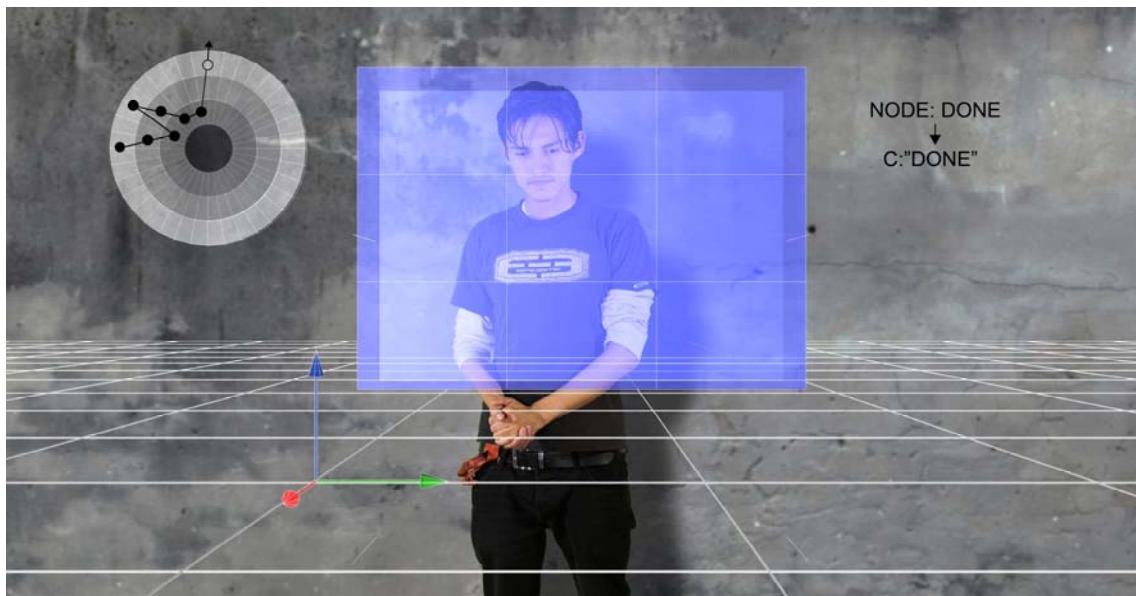


Fig.47 Interface Mockup – Finalizing the task.

4.5 Examples

4.5.1 Insert a Window on a Wall

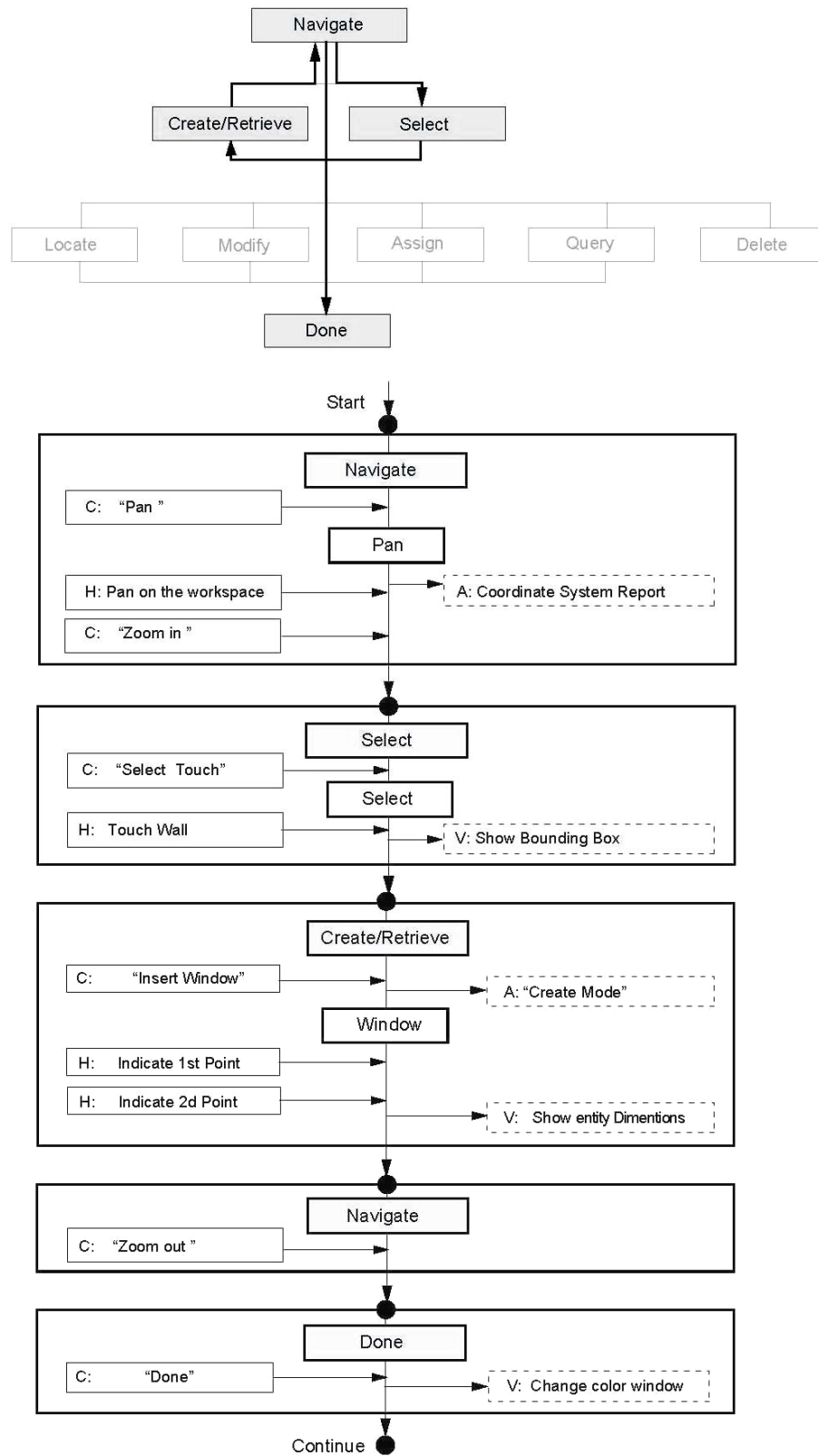


Fig.48 "Insert a Window on a Wall" workflow.

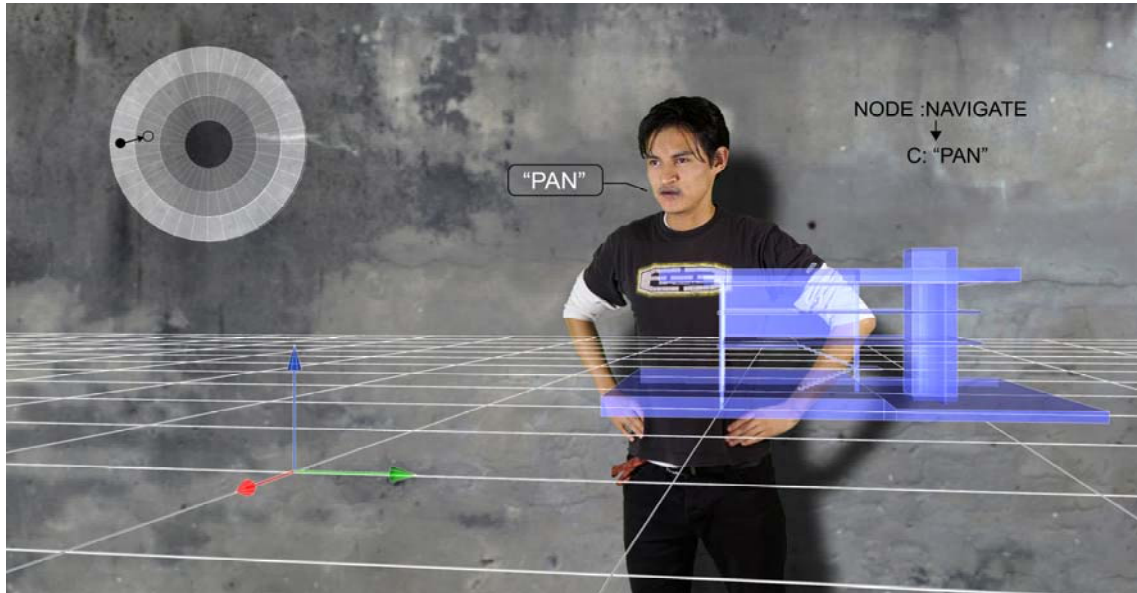


Fig.49 Interface Mockup - Set "Navigation"

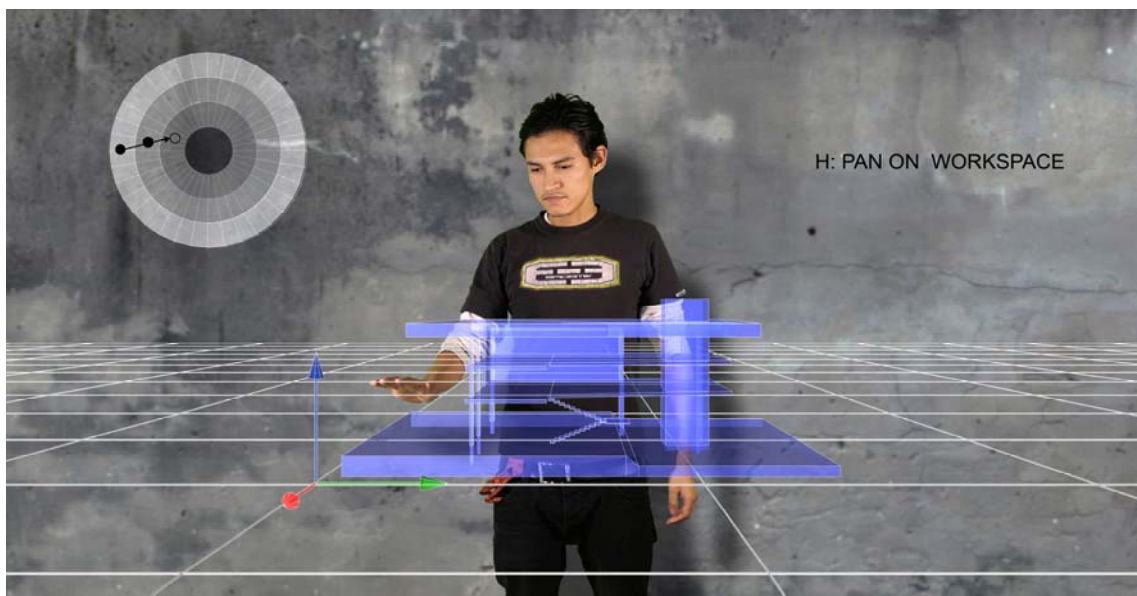


Fig.50 Interface Mockup - "Pan" on workspace

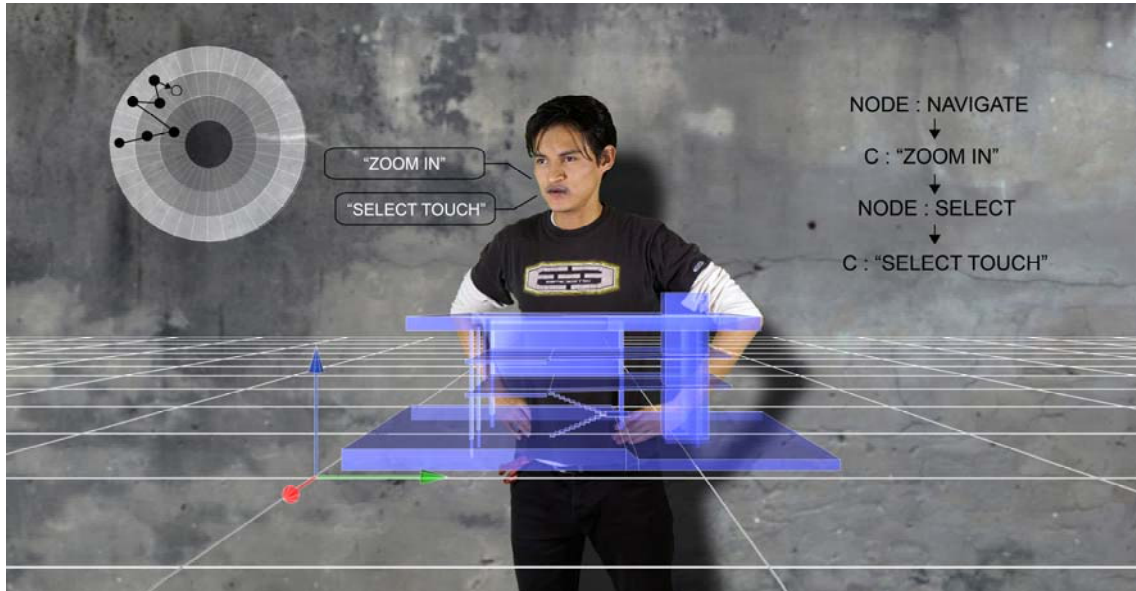


Fig.51 Interface Mockup - "Zoom In" - "Select Touch"

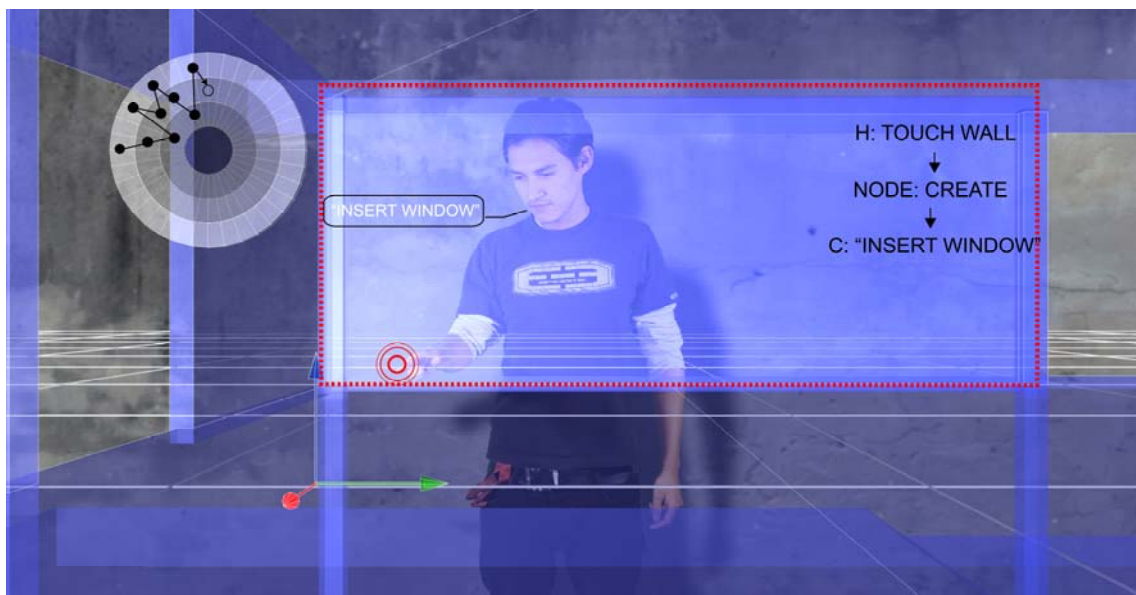


Fig.52 Interface Mockup - "Insert window"

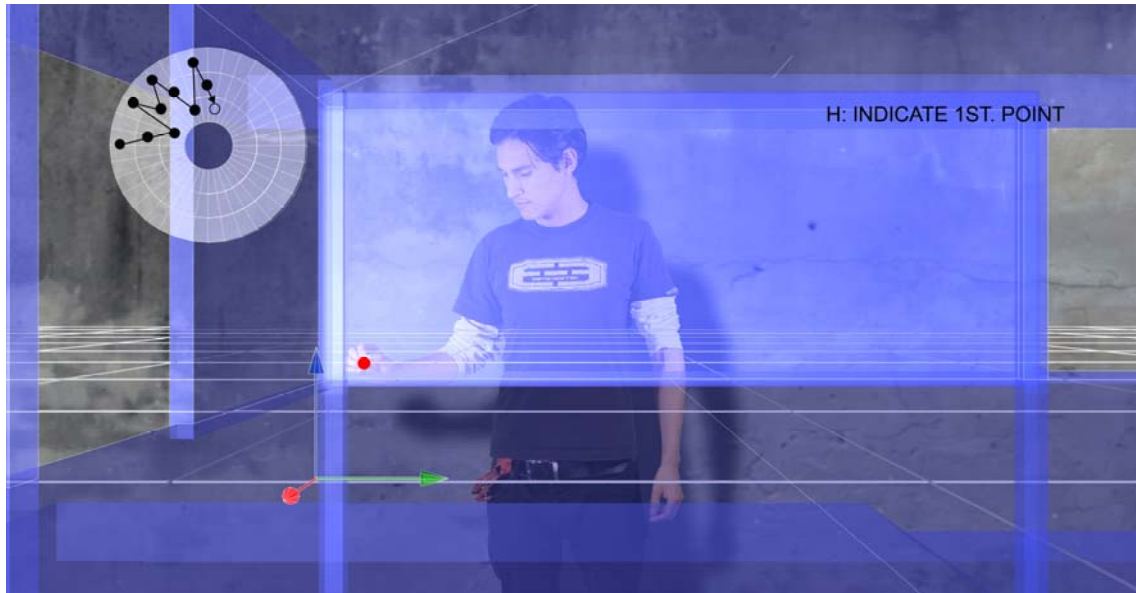


Fig.53 Interface Mockup - Indicate 1st point.



Fig.54 Interface Mockup - Indicate 2nd . point.

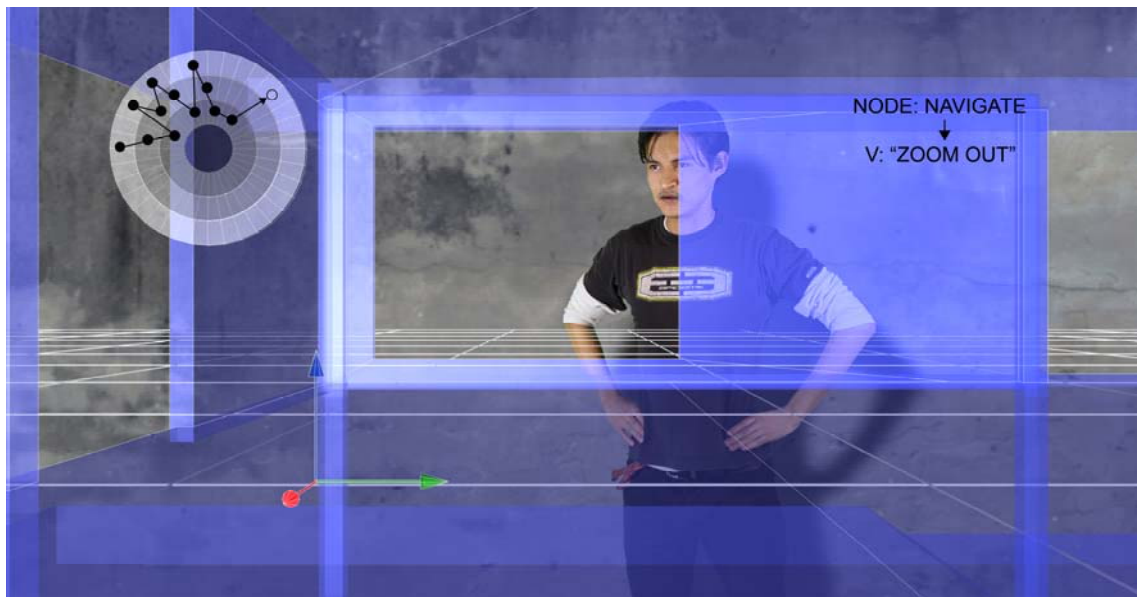


Fig.55 Interface Mockup - "Zoom Out".

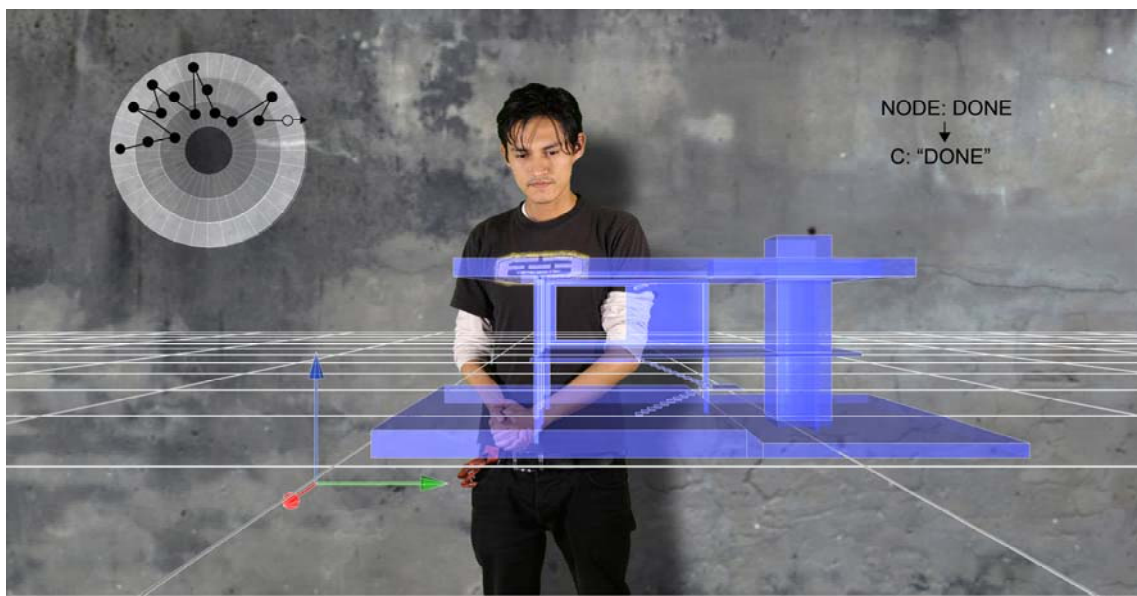


Fig.56 Interface Mockup - Finalizing task.

4.5.2 “Select a window and relocate it”

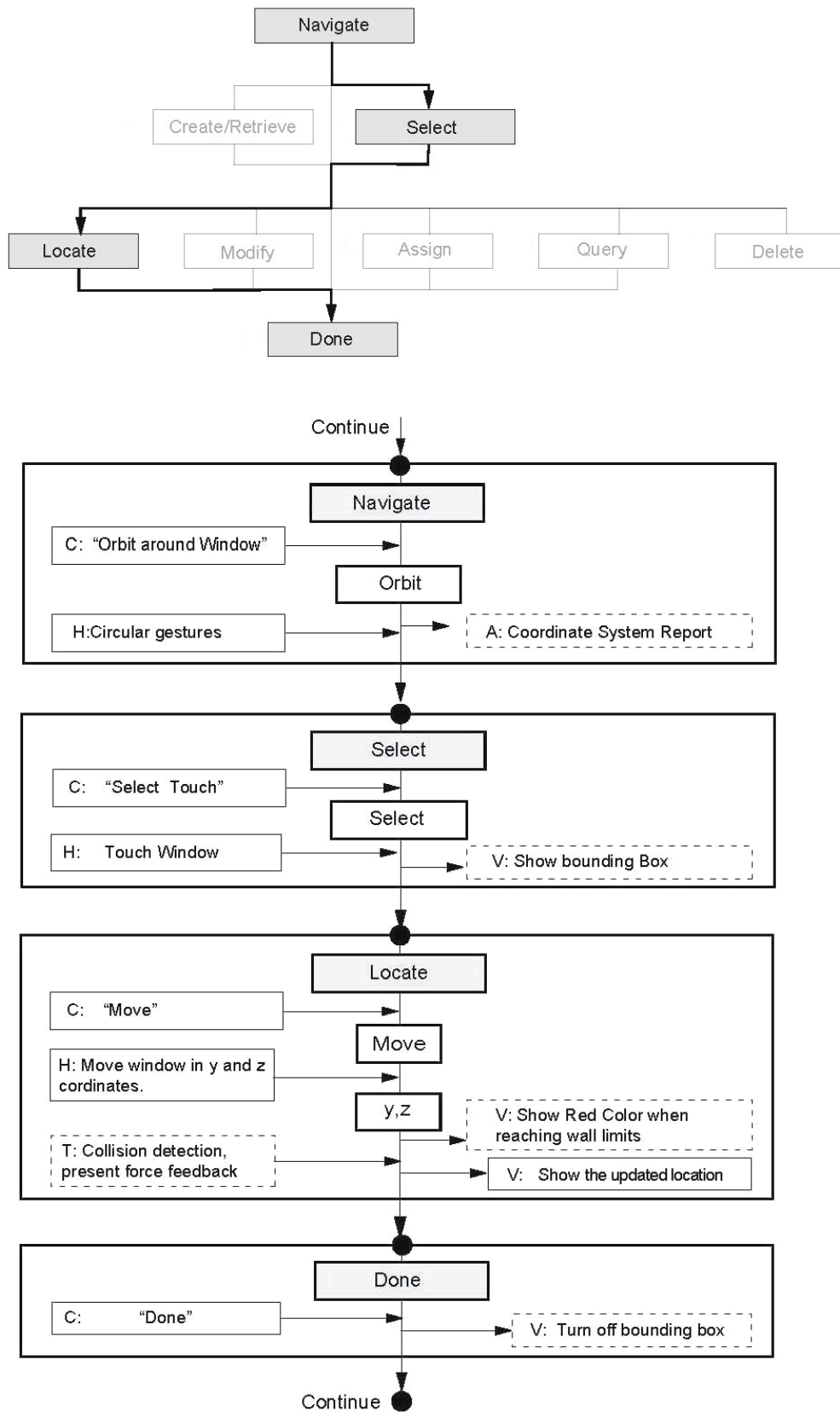


Fig.57 “Select a window and relocate it” workflow.

4.5.3 “Select a wall and Assign an ID”

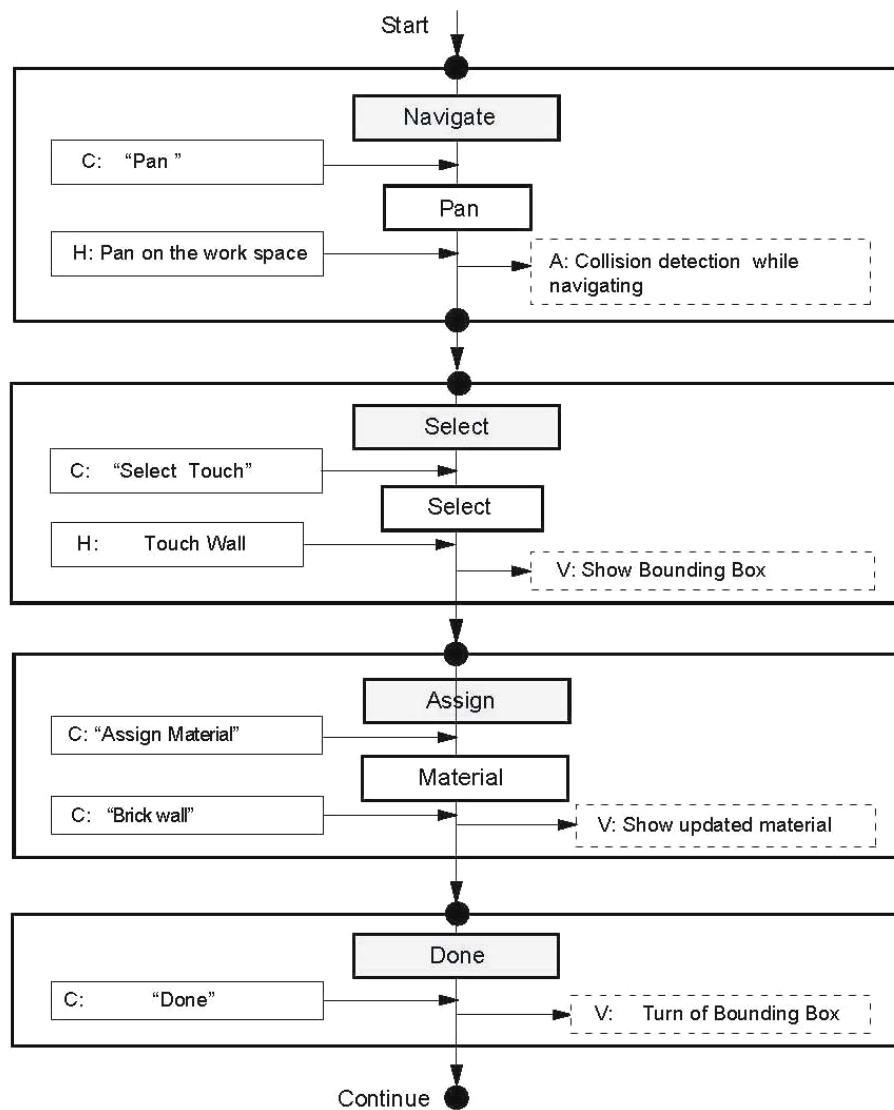
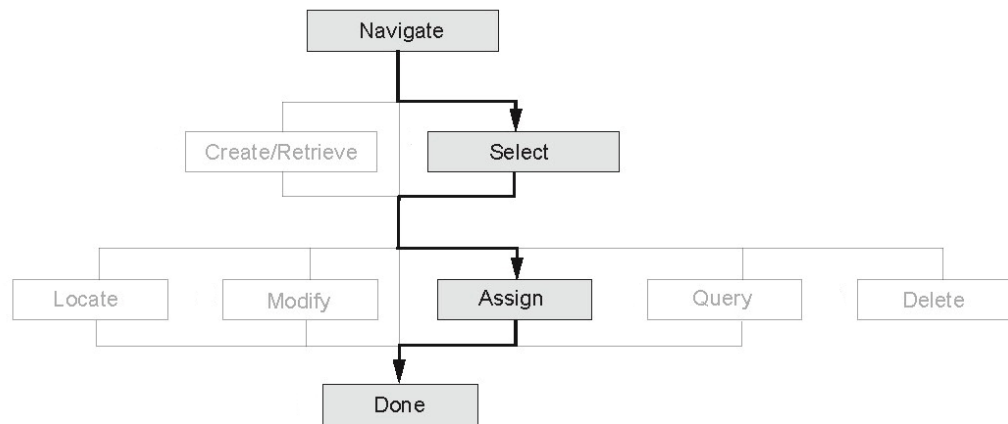


Fig.58 “Select a wall and Assign an ID” workflow.

4.6 Visualization information for Multimodal interaction

The network system depicts information by using a reduced number of nine elements which cover a larger set of subsidiary elements or tools needed for the 3d design process, In addition a “Concentric Design Visualization” is been proposed to support the user by providing real-time visualization and awareness of his location in relation to the process of interaction with the Multimodal interface system.

The purpose is to visualize the workflow by; showing the links between the steps of decision making of the tool elements contained in each of the nine nodes of interaction (Fig.51). The user is provided with a condensed graphical output which represents the reading of the Multimodal Design process; this can be helpful for navigating and making bookmarks on specific steps during the design process. This type of visual feedback can also be able susceptible of modification by real time manipulation of the visual elements it contains as it happens in some existing touchpad applications.

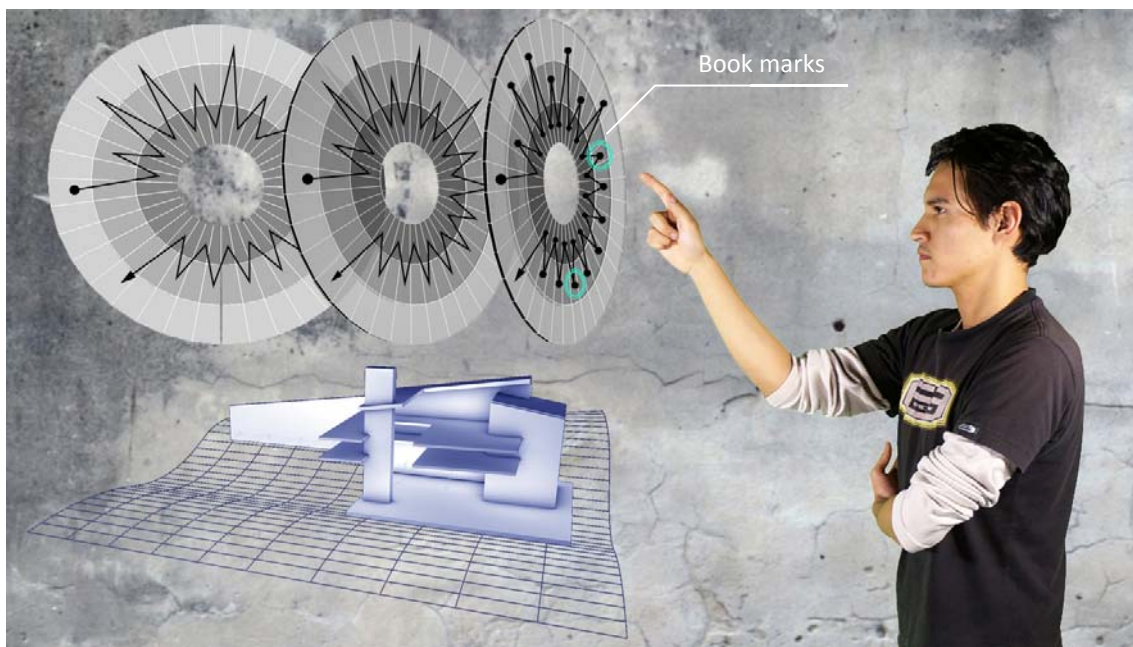


Fig.59 Visualization Information of the interface

Color coded - Multimodal Modal Selection

The graph uses a color code that determines the type of mode is been used in each of the elements of the path. Every time the user switches mode the color changes and it's been recorded. This reading shows how many times a certain mode its being used , with this information the system could be calibrated in order to establish a balance, so that the user doesn't make an excessive used of a mode and avoiding others (Fig.52).

“Voice commands” for example can be used to input numerical data but when moving a 3d element the reading of the 3d coordinate system could be a tedious work which can be easy substituted by performing Hand gesture manipulation.

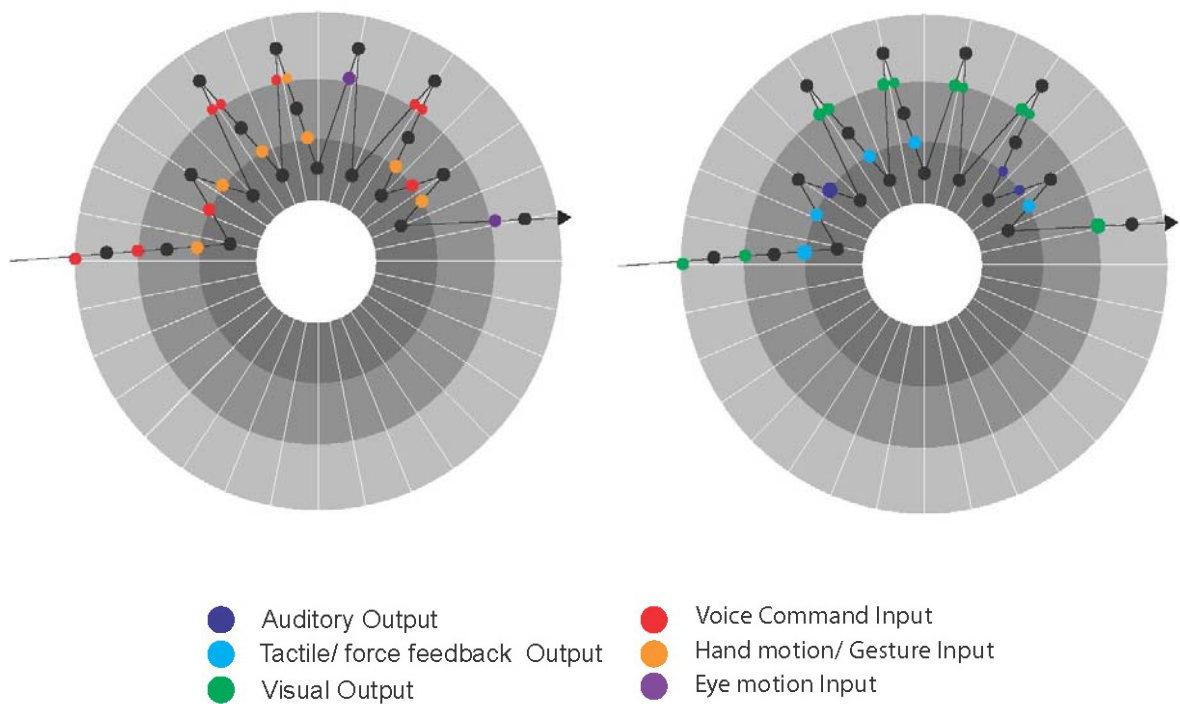


Fig.60 Multimodal color-coded Interpretation

Time lapse visualization.

It is related to behavioral analysis, in this type of visualization a revolution of 360 degrees on the circular trajectory can be set as a reference of unit of time, a minute or an hour, so the diagram can show how much time the user needs to perform a given task when operating with the interface, making decisions and to performing different kind of operations (Fig.53).

Another type of data that can be collected is the count of dots and its location in relation to the circular trajectory, this represents a numerical value that can be compared in between different users for different task. For example how many times the user has to take decisions of design in relation to the number of times the user has to take decisions about which tool to use.

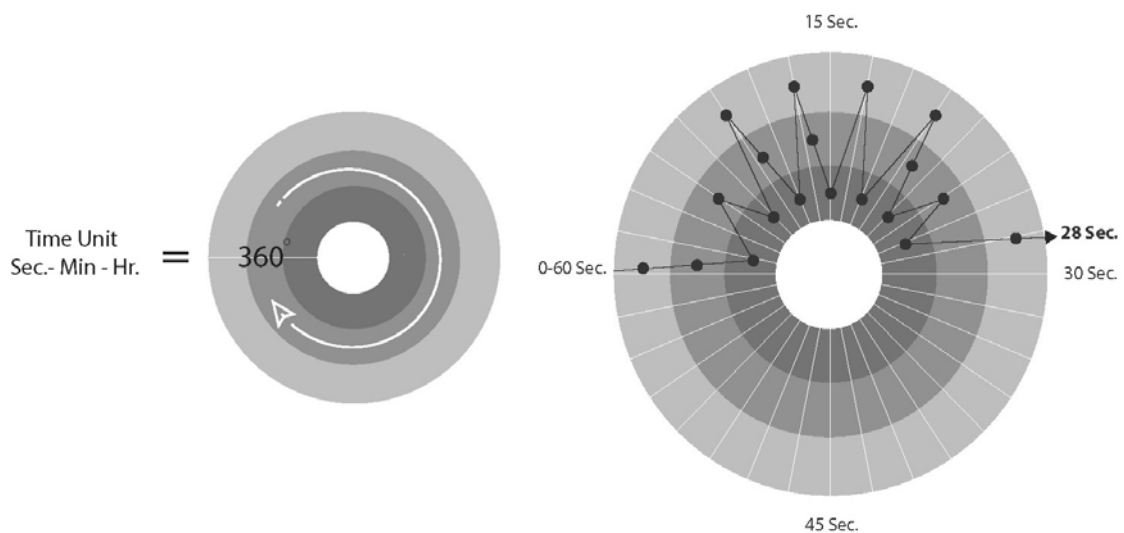


Fig.61 Time Consuming Interpretation

5. Conclusions

Exploring possible alternatives and modifications during design has become an iterative process where computational means can be used as a tool for inputting and retrieving information. With this the way users interact with digital information evolves as new devices are been developed and introduced into many professions. For Architects and designers understanding 3d information either real or digital is crucial when developing ideas of a project. The type of interfaces that allows communication between real and digital requires special attention in order to provide a more intuitive and fluid language for the interaction to happen.

By analyzing multimodal interaction with CAAD interfaces is evident that some channels of interaction may be recalled and used more frequently than others. This implies that users could input information with voice commands for the most part of a design , but at any time they could also used a subsidiary channel either to realize a different task of modeling or to support the ones in process. At this point the design and modeling would change from a linear process into a branching one. The visualization of this type of Interface would show multiple changes happening at once and so the time required to perform one entire operation could be reduced and so the design process in general could be faster.

On the other hand when using multiple Channels for interaction the interface becomes more human centered now that it tries to mimic the interaction that happens on the real world, person to person. Like sculptors architects would be able to manipulate and produce 3d models in a more intuitive way perceiving what is digital in real time and having multi-sensorial feedback from different information parameters. Working with such an Interface would demand a previous training and knowledge about geometry and form generation as well as knowledge of Architectural language terminology.

A tool like this could be use not only as a mean for production and representation but also as an extension for running simulations so that different environmental conditions could be represented and manipulated freely by the users in order to explore different formal and spatial solutions with a broad way of understanding the changes that take place.

Another conclusion that merges out of the work is that when dealing with multiple channels of interaction, the system will be able to recognize design routines and interaction paths. When reading and recording the system could learn and anticipate in order assisting the user. The software application would evolve to reach a certain level of intelligence and became more responsible to the user intentions. This type of support provided by the system should in any case showed possibilities but never take decisions other wise the human design would be compromised.

As software developers and computer scientist improve the tools that allow humans to interact with the digital world, a new understanding of the design merges where anything can be possible in a fictional world. This type of technologies taken as an extension of the reality offer multiple possibilities when dealing with geometric paradigms and are also promising to field of Architecture by recreating a world where the design process can flourish without technical restrictions.

References

- [1] Aliakseyeu, D., Martens J.B. & Rauterberg M. "A computer support tool for the early stages of architectural design." *Interactive Computer-Aided Desing*.2006 Vol.18: pp.528–555.

- [2] Tamura, H. "Steady steps and giant leap toward practical Mixed Reality Systems and applications." *Proceedings Int'l Status Conf. on Virtual and Augmented Reality, Leipzig, Germany. 2002*.

- [3] Petric, J., Conti G. & Ucelli G. "Design within virtual worlds. Digital design research and practice." *Proceedings of the 10th International Conference on Computer Aided Architectural Design, Futures*.2003: pp.213-223.

- [4] Myers, B.A. "A Brief History of Human Computer Interaction Technology." *ACM interactions. March.1998* Vol.5: pp.44-54.

- [5] Jeng, T. "Advanced Ubiquitous Media for Interactive Space." *Proceedings Computer Aided Architectural Design, Futures*.2005: pp.341-350.

- [6] Jaimes, A & Sebe, N. (2005), 'Multimodal Human Computer Interaction: A Survey'IEEE International Workshop on Human Computer Interaction in conjunction with ICCV , Beijing, China.2005.

- [7] Sun, Y., Chen, F., Shi,Y. & Chung,V. "A novel method for multi-sensory data fusion in multimodal human computer interaction" *Proceedings of the Australasian Computer-Human Interaction Conference, OZCHI, Sydney, Australia. 2006*.

- [8] Krevelen, D.W.F. & Poelman, R. "A survey of augmented reality technologies, applications and limitations." *The International Journal of Virtual Reality, Delft, The Netherlands.2010*: pp.1-20.

- [9] Chu, C.C., Tushar H.D. & Gadh R. "Multi-sensory user interface for a virtual-reality-based computer aided design system." *Computer-Aided Design,NO 10.1997* Vol.29: pp. 709-725.

- [10] Kou, X.Y. & Tan, S.T. "Design by talking with computers." *Computer-Aided Design & Applications, 5(1-4).2008*: pp.266-277.

- [11] Xue,S., Kou,X.Y. & Tan,S.T. "Natural Voice-Enabled CAD: Modeling via natural discourse." *Computer-Aided Design and Applications 6(1).2009*: pp.125-136.

- [12] Chan, C.S. "Evaluating cognition in a workspace virtually." *CAADRIA.2007*: pp.451-458.

- [13] Lu, K.T. "Navigating 3d Information Space with 6 Degree of Freedom Devices." *Proceeding of the 10th International Conference on Computer Aided Architectural Design Research in Asia, CAADRIA 04, Seoul, Korea.2004*: pp. 93-101.

- [14] Sánchez, E.& Hernández, M.. "Hand Gesture Recognition for Human-Machine Interaction." *Journal of WSCG, Plzen, Czech Republic.2003* Vol.12.

- [15] Malima, A., Özgür E.& Çetin M. "A fast algorithm for vision-based hand gesture Recognition for robot control." *IEEE 14th Signal Processing and Communications Applications, Antalya.2006*.

- [16] Lockton, R. "Hand Gesture Recognition Using Computer Vision." Master's thesis: Oxford University.

- [17] Mistry, P. & Maes, P. 'SixthSense - A Wearable Gestural Interface" In the Proceedings of SIGGRAPH Asia 2009, Sketch. Yokohama, Japan. 2009.
- [18] Hsiang, S.H. 'Using Hand Movement System to Operate 3D Objects in Virtual Environment" The Quest for New Paradigms , 23rd eCAADe Conference Proceedings, Lisbon, Portugal, September. 2005, pp. 779-786.
- [19] Sagisaka, Y. & Tokuda, K. "The Seventh ISCA Tutorial and Research Workshop on Speech Synthesis." *ISCA Archive, Kyoto, Japan, September. 2010*: Available from: <http://www.isca-speech.org/archive/ssw7>; accessed 20 January 2012.
- [20] Sandor, C., & Kuroki Tsuyoshi & Uchiyama Shinji & Yamamoto H. "Exploring Visuo-Haptic Mixed Reality." *Technical Report PRMU2006-199 IEICE, Tokyo, Japan. 2007*.
- [21] Lingley, A.R., Ali M. Liao Y. Mirjalili R. Klonner M. Sopanen M. Suihkonen S. Shen T. Otis B.P. Lipsanen H. & Parviz B.A. "A single-pixel wireless contact lens display." *Journal of Micromechanics and Microengineering. 2011* Vol. 21.
- [22] Wang, X., Gu, N., Marchand, D., Gong, Y., Jeong, K.M. 'Perceptions Of Augmented Reality In The Design Sector" 7 International Conference on Construction Applications of Virtual Reality. 2007.
- [23] Kato, H. "Virtual object manipulation on a table-top AR environment" Proceedings IEEE and ACM International Symposium on Augmented Reality , ISAR. 2000.
- [24] Panayiotis, Z. & Chee Siang, A. "Editorial: HCI issues in computer games." *Interact. Comput.* Vol. 19: 135--139.

[25] Dix, A., Finlay J. Abowd G.& Beale R. "Multi-Sensory Systems." *Human-Computer Interaction, Prentice Hall.1993.*

[26] Drewes, H. "Eye Gaze Tracking for Human Computer Interaction." Master's thesis: Ludwig-Maximilians-Universitaet Munchen,2010.

[27] Brunner, S. "Using microsoft kinect sensor to perform commands on Virtual objects.2012." Master's thesis: Universitaet Bern, 2012.

[28] Chen, C.T. & Chang T.W. "1:1 Spatially Augmented Reality Design Environment." *Innovations in Design & Decision Support Systems in Architecture and Urban Planning. DDSS. Dordrecht: Springer.2006: pp.487-499.*

[29] Azuma, R.T. "A Survey of Augmented Reality." *Presence: Teleoperators and Virtual Environments , August.1997 Vol.6(4): pp.355 - 385.*

[30] Babak, A.P. "Augmented Reality in a Contact Lens." *IEEE Spectrum.2009:* Available from:<http://spectrum.ieee.org/biomedical/bionics/augmented-reality-in-a-contact-lens/0>; accessed 15 February 2012.

[31] Khemlani, L. "Revit Architecture 2012." *AEC bytes:* Available from: <http://www.aecbytes.com/review/2011/RevitArch2012.html>; accessed 10 January 2012.

[32] Carpenter, T. "Apple See-Through Augmented Reality HMD Glasses." : Availablefrom:<http://gamesalfresco.com/2009/12/29/apple/see/through/augmented/reality/hmdglasss>; accessed 15 January 2012.

[33] Siegel, C.& Zhao, T. "Haptic Glove." : Available from: [http://courses.cit.cornell.edu/ee476/FinalProjects/s2008/crs54_tz36/crs54_tz36/two co umn.html](http://courses.cit.cornell.edu/ee476/FinalProjects/s2008/crs54_tz36/crs54_tz36/two%20co%20umn.html); accessed 26 January 2012.

[34] Fagel, S. & Clemens, C. "An articulation model for audiovisual speech synthesis; determination, adjustment and evaluation." *Speech Communication*. 2004 Vol.44: pp.141 - 154.

[35] Net Industries, Virtual Reality. "The Origin Of Virtual Reality, Components Of Virtual Reality, Applications Of Virtual Reality, On The Horizon Of Virtual Reality." : Available from: <http://science.jrank.org/pages/7200/Virtual-Reality.html>; accessed 26 January 2012.

Web Sources

Archicad	http://www.graphisoft.com ; accessed 22 January 2012.
Andol	http://www.andol.info ; accessed 23 January 2012.
Android	http://www.android.com/developers/ ; accessed 15 January 2012.
AMOLED	http://www.oled-info.com ; accessed 15 March 2012.
Apple	http://support.apple.com/ ; accessed 04 January 2013.
Blender	http://www.blender.org ; accessed 22 January 2012.
Cinema4d	http://www.maxon.net ; accessed 22 January 2012.

Cyber Touch <http://www.cyberglovesystems.com;>
accessed 25 January 2012.

Cyber glove systems <http://www.cyberglovesystems.com;>
accessed 25 January 2012.

Dataglove <http://www.metamotion.com;>
accessed 25 January 2012.

EyeSight's <http://www.eyesight-tech.com;>
accessed 23 January 2012.

EyeTracking [http://mirametrix.com/products/;](http://mirametrix.com/products/)
accessed 25 January 2012.

Eyetechnology Digital systems <http://www.eyetechnology.com;>
accessed 25 January 2012.

eSpeak <http://espeak.sourceforge.net/index.html;>
accessed 27 January 2012.

Flexible Eye Tracking <http://www.tobii.com;>
accessed 26 January 2012.

HandVu <http://www.movesinstitute.org;>
accessed 23 January 2012.

Maya <http://usa.autodesk.com/maya;>
accessed 22 January 2012.

Tobii <http://www.tobii.com;>
accessed 26 January 2012.

Model talker <https://www.modeltalker.com;>
accessed 25 January 2012.

Microsoft [http://www.microsoft.com/en-us/tellme/;](http://www.microsoft.com/en-us/tellme/)
accessed 20 January 2012.

Microsoft Research <http://research.microsoft.com;>
accessed 27 January 2012.

P5 Glove <http://www.vrealities.com/P5.html;>
accessed 25 January 2012.

Rhino3d	http://www.rhino3d.com; accessed 22 January 2012.
Revit	http://usa.autodesk.com/revit/; accessed 22 January 2012.
Shapehand.	http://www.shapehand.com; accessed 25 January 2012.
Sketchup	http://sketchup.google.com; accessed 22 January 2012.
Speak4CAD™	http://www.automation.com; accessed 26 January 2012.
Speech Synthesizer	http://download.cnet.com; accessed 27 January 2012.
SixthSense	http://www.pranavmistry.com/projects/sixthsense/; accessed 23 January 2012.
Eye Writer	http://www.eyewriter.org/developer/; accessed 27 January 2012.
VisionTrak	http://polhemus.com; accessed 26 January 2012.
VoiceOver	http://www.apple.com/accessibility/voiceover/; accessed 26 January 2012.
Vox Forge	http://voxforge.org; accessed 24 January 2012.
3ds Max	http://usa.autodesk.com/3ds-max; accessed 22 January 2012.
3ITU Gaze	http://www.gazegroup.org/downloads/23-gazetracker; accessed 25 January 2012.

Nomenclature

CAAD	:	Computer Architectural Aided Design
BIM	:	Building Information Modeling
HCI	:	Human Computer Interface
TUI	:	Tangible User Interface
HMD	:	Head Mounted Device
ASR	:	Automatic Speech Recognition
VR	:	Virtual Reality
AR	:	Augmented Reality
MR	:	Mixed Reality
MMHCI:		Multimodal HCI
GUI	:	Graphical User Interface.
AEC	:	Architecture, Engineering and Construction Industry
UCI	:	User Computer Interaction
NUI	:	Natural User Interaction

List of Figures

- Fig.1 Desktop computing; www.nismat.or/ptcor/ergo; accessed: November 14, 2012.
- Fig.2 Immersive reality; www.christiedigital.co.uk; accessed: November 14, 2012.
- Fig.3 Human-Computer Interaction Channels (16)
- Fig.4 Skin color detection; www.andol.info/hci/2010.htm; accessed: November 14, 2012.
- Fig.5 Emgu Contour representation [27].
- Fig.6 Eye Sight Technologies; www.eyesight-tech.com; accessed: August 15, 2012.
- Fig.7 Using palm for dialing numbers; www.pranavmistry.com/projects/sixthsense/; accessed: August 15, 2012.
- Fig.8 Using Infrared Devices; www.ncl.ac.uk/press.office/press.release/item/power-in-the-palm-of-your-hands#.UNYt82_OhtA/; accessed: December 20, 2012.

- Fig.9 Cyber Glove -Hand motion Capture; <http://www.cyberglovesystems.com/>;
accessed: December 20, 2012.
- Fig.10 Pupil tracking; <http://www.digitaltrends.com/computing/death-of-the-mouse-how-eye-tracking-technology-could-save-the-pc/> ; accessed:
December 20, 2012.
- Fig.11 HMD for eye tracking Tobii glasses; <http://www.tobii.com>; accessed:
December 20, 2012.
- Fig.12 Portable device for eye tracking; <http://www.tobii.com>; accessed: December
20, 2012.
- Fig.13 Flow diagram of speech synthesis;
http://en.wikipedia.org/wiki/Speech_synthesis; accessed: August 15, 2012.
- Fig.14 Cyber Touch Technologies; <http://www.cyberglovesystems.com/> ; accessed:
December 20, 2012.
- Fig.15 P5 Glove; <http://www.vrealities.com/P5.html>; accessed: December 20, 2012.
- Fig.16 CyberGlove [48]
- Fig.17 OLED Technologies [51]
- Fig.18 OLED Technologies [51]
- Fig.19 Virtual Reality; <http://electronics.howstuffworks.com>;
accessed: November 14,2012.
- Fig.20 Augmented Reality
- Fig.21 Reality-Virtuality continuum [69][8]
- Fig.22 Augmented Reality Markers;
<http://www.creativeapplications.net/processing/augmented-reality-with-processing-tutorial-processing/>; accessed: November 15,2012.
- Fig.23 Wii User Interaction; <http://www.thegamingvault.com/2011/08/opinion-nintendo-are-about-to-head-back-into-the-n64gamecube-wilderness/>;
accessed: October 10, 2012.
- Fig.24 Wii Game Controllers; <http://www.actionbutton.net/?p=590>;
accessed: November 10, 2012.
- Fig.25 Holodeck, Star Trek , Roddenberry (1987);
http://en.memory_alpha.org/wiki/Holodeck; accessed: December 22, 2012

Fig.26 2001: A Space Odyssey, Kubrick (1968);http://drafthouse.com/movies/2001_a_space_odyssey/san_antonio; accessed: November 16,2012

Fig.27 HIC, Iron Man, Favreau, 2011;
<http://www.comicbookmovie.com/fansites/MarvelFreshman/news/?a=17900>
 ; accessed: November 13, 2012

Fig.28 Beam Properties Chart. [56] - Archicad

Fig.29 2D Graphical Representation [56] - Archicad

Fig.30 3D Building Model [56] - Archicad

Fig.31 3D BIM Window [56] - Archicad

Fig.32 Navigation Project Map [56] Archicad

Fig.33 Design Routine "Make a wall"

Fig.34 Nine nodes Design task Network based on [12]

Fig.35 Design Task Network "Create Wall"

Fig.36 "Create Wall" workflow

Fig.37 Operate virtual objects with Hand gestures [65,66,67]

Fig.38 Concentric Design task network

Fig.39 Concentric Design task network workflow

Fig.40 Path traveling definition

Fig.41 Path traveling variations

Fig.42 Interface Mockup - Set "Navigation" node

Fig.43 Interface Mockup - "Pan" on workspace

Fig.44 Interface Mockup - Set "Create wall"

Fig.45 Interface Mockup – Set 1st point

Fig.46 Interface Mockup – Set 2nd point

Fig.47 Interface Mockup - Finalizing task

Fig.48 "Insert a window on a wall "workflow

Fig.49 Interface Mockup - Set "Navigation" node

Fig.50 Interface Mockup - "Pan" on workspace

Fig.51 Interface Mockup - "Zoom In"-"Select Touch"

Fig.52 Interface Mockup - "Insert window".

Fig.53 Interface Mockup - Indicate 1st point.

Fig.54 Interface Mockup - Indicate 2nd point

Fig.55 Interface Mockup - “Zoom Out”

Fig.56 Interface Mockup - Finalizing task

Fig.57 “Select a window and relocate it “workflow

Fig.58 “Select a wall and assign an ID “workflow

Fig.59 Visualization information of the interface

Fig.60 Multimodal color-coded Interpretation

Fig.61 Time consuming Interpretation

List of Tables

Table 1. Multi – Sensorial Interaction Channel

Table 2. Software Characteristics