USING NATURAL INTERFACES TO INTERACT WITH A VIRTUAL CONTROL DESK OF A NUCLEAR POWER PLANT

Maurício A. C. Aghina[(a)], Antônio Carlos A. Mól[(b)], Carlos Alexandre F. Jorge[(c)], Paulo Victor R. Carvalho[(d)] , Ana Paula Legey[(e)] , Gerson G. Cunha[(f)], Luiz Landau[(g)]

(a), (b),(c),(d) Instituto de Engenharia Nuclear /Comissão Nacional de Energia Nuclear, , RJ, Brazil
(b)Instituto Nacional de Ciência e Tecnologia de Reatores Nucleares Inovadores/CNPq, Brazil
(b), (e) Universidade Gama Filho, Brazil
(f), (g) Universidade Federal do Rio de Janeiro, Brazil

[(a)] mag@ien.gov.br, [(b)] mol@ien.gov.br, [(c)] calexandre@ien.gov.br , [(d)] paulov@ien.gov.br , [(e)] analegey@hotmail.com,
[(f)] gerson@ufrj.br [(g)] landau@lamce.copp.ufrj.br

ABSTRACT

This paper reports results achieved in a development for a virtual control desk, interfacing with a nuclear power plant's control system. This virtual control desk was developed aiming to combine the dynamics simulation of a nuclear power plant operation, with high fidelity control desk's visual appearance. Natural interfacing techniques where used to interact with this virtual control desk, as spoken command recognition, head tracking and body tracking. The combination of such interfacing techniques could improve user interfacing, through exploring each technique's advantages for specific tasks. For instance, spoken command recognition is used for switching among different frame views of the virtual control desk, while head tracking techniques are used for specific frame exploration to access indicators and controls. Body tracking is similar of head tracking, but the actuation of the controls of the virtual control desk are made by computational viewing of hands, instead of mouse.

Keywords: Non-conventional interfaces, Virtual environments, Virtual control desk, Nuclear plants, Speech recognition, Face tracking, Body tracking.

1. INTRODUCTION

Nuclear power plants (NPP) involve high safety requirements in their operation, thus operators must keep them into normal operational conditions, or act appropriately and fast to bring them back to normal conditions in the occurrence of any abnormal ones. Operators must run very efficient training, to prepare facing postulated incidents or accidents. These training used to be carried out through the use of full-scope control desk models, which resembled real ones with high fidelity relatively to visual appearance. This favored training in that users could see all variable indicators, actuator controls and alarm indication in the same positioning as they would do in real control desks.

Virtual reality (VR) techniques can improve user interfacing with NPP simulators, − or simulators of any other industrial plants −, since virtual control desks (VCD) resemble much the real ones in which they are based. Therefore, the VCD approach combined with the computer-based simulators bring both the online dynamics simulation capability and the high fidelity visual appearance, favoring user training and ergonomics evaluation.

Such a VCD has been developed at *Instituto de Engenharia Nuclear* (Nuclear Engineering Institute – IEN), a research and development (R&D) center belonging to *Comissão Nacional de Energia Nuclear* (Brazilian Commission of Nuclear Energy – CNEN), (Aghina et al., 2008). An existing NPP computer-based simulator was integrated with this developed VCD through computer networking, either local or through the Internet.

New interacting modes, like natural interfaces, were included for a friendlier user interfacing, to enable user interaction in front of projection screens, for example, free from computer keyboard and mouse. These new interacting modes comprise an automatic speech recognition (ASR) system, head and body tracking systems that will be discussed in the following sections.

Three systems of computational viewing devices were tested, two head tracking systems, with and without visual markers, and Kinect, from Microsoft, for body tracking. This later device can acquire 3D data of the coordinates of the human body and then interact with the VCD. This interaction can control the viewing of the VCD and controls the positioning and actuation of display cursor, by hands. We observed that a combination of these interacting modes served better user interfacing, than using one or another alone. A comparative analysis of the two head tracking systems, and the body tracking is also performed.

2. RELATED RESEARCH

A previous R&D was carried out at IEN for the development of a computer-based NPP simulator, in cooperation with the Korea Atomic Energy Research Institute (KAERI) and with the International Atomic

Energy Agency (AIEA). This cooperation resulted in a new laboratory at IEN in 2003, named *Laboratório de Interfaces Homem-Sistema* (Human-Systems Interface Laboratory – LABIHS), (Carvalho and Obadia, 2002; Santos et al., 2008).

The developed NPP simulator was coupled with a synoptic windows-based interface that communicated with the simulator through computer networking. This whole system has been used through these years at IEN for operator training and to support ergonomics evaluation. The later led to modification proposals in the form information is presented for users, so as to improve it from the ergonomics point of view, and consequently improve operational safety (Carvalho et al., 2008; Santos et al., 2008; Oliveira et al., 2007). Figure 1 shows a view of a reduced-scale model of the real control desk in that the current VCD was based. Figure 2 shows a view of the computer-based NPP simulator room, where it is possible to notice the use of multiple computer screens, to minimize the need for switching among many frame views; even so, it may still be a difficult task.

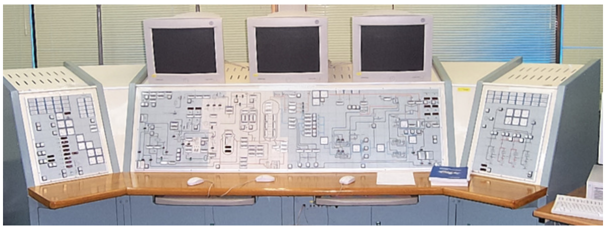

Figure 1: Reduced-scale full-scope model of the real control desk.

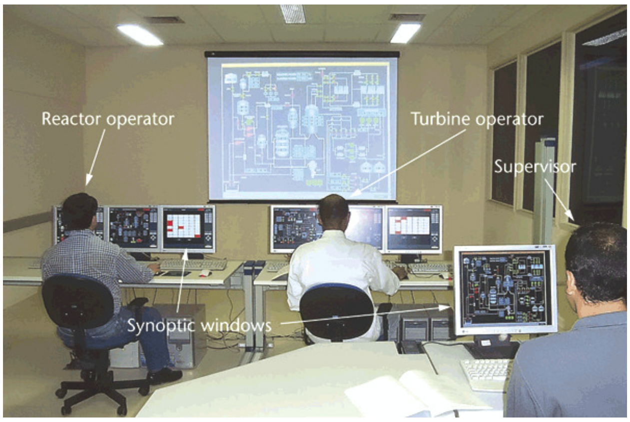

Figure 2: Computer-based NPP simulator room.

Relatively to VR approaches in designing and evaluating control desks or rooms, there are other R&D groups running similar works directed also towards the nuclear field, as can be verified by the following references (Drøivoldsmo and Louka, 2002; Nystad and Strand, 2006; Markidis and Rizwan-uddin, 2006; Hanes and Naser, 2006). The VR approaches enable the evaluation and decision making about location of variable indicators, actuator controls and alarms displays, to serve a better operator acting to run normal NPP operations and to mitigate any abnormal conditions, before the construction o real control desks. This enables also the modification of existing ones, to improve their design from the ergonomics perspective.

An overview of all the R&D developed at IEN, from the beginning with first VCD results up to the present with the most recent results is described in (Aghina et al., 2008).

3. THE VIRTUAL CONTROL DESK

The VCD was developed from the beginning to consist in an interactive interface with users, what led to the choice of OpenGL graphics pack for C/C++ languages. The VCD design was fully based on the real control desk shown in Figure 1, considering all variable indicator and actuator control types, as well as alarm indicators. These interface types were created as different classes, each on replicated as needed. Photos were used as textures for all these interfaces and for the front VCD front panel, with the schematic connections shown. Figure 3 shows a complete view of the VCD; compare it with Figure 1. Figure 4 shows a close perspective view of the VCD.

Figure 3: VCD's complete view.

Figure 4: A close perspective view of the VCD.

4. NATURAL INTERFACES

Friendlier man system interfaces were developed to improve users' interaction with the system, so they could interact in front of a projection screen, computer screen, or any other display, free from keyboard and mouse. This thus would enable more natural interaction forms, consequently improving user immersion. In a first stage, an ASR system was developed as the friendly interface, through which users could interact with the VCD, by spoken commands as: "left", "right", "up", "down", "zoom in", "zoom out", and so on. But the R&D staff soon became to try other friendly interaction modes, besides this. In particular, two head tracking techniques and body tracking were implemented and tested, and either one of them combined with the ASR system.

4.1. Speech Recognition

The ASR made use of well-known techniques for small-vocabulary isolated-word recognition system, as using Cepstral analysis (Rabiner and Schafer, 1978; Oliveira, M. V.; Moreira, D. M.; Carvalho, P. V. R.; 2007. Construção de interfaces para salas de controle avançadas de plantas industriais, *Ação Ergonômica,* Vol. 3, 8-13.
Oppenheim and Schafer, 1989) for parameter extraction, and neural networks (NN), (Haykin, 1999) for pattern recognition. The system was first implemented offline, and then upgraded to full online system, including online NN training, as detailed in the following. The speech recognition is currently performed in the following steps: (i) speech detection; (ii) end-point detection; (iii) word segmentation; (iv) parameter extraction; (v) pattern recognition. All steps are detailed.

In the first step above (i), speech sound is automatically detected through a specified threshold, adjusted experimentally, to identify speech above background noise, and start recording.

Then, in the second step (ii), once recorded with a fixed (sufficient large) time window, – also adjusted experimentally –, end-point detection is performed to isolate the spoken word itself from the background noise, following a simple approach using short-time energy (Pinto *et al.*, 1995).

In the third step (iii), the (already isolated) spoken word is segmented in an approximate range of 30 ms, where speech can be considered stationary (Rabiner and Schafer, 1978). Segmentation is performed with Hamming window (Rabiner and Schafer, 1978), with a fifty-percent superposition to compensate for attenuation in the segments' ends (Lima *et al.*, 2000; Diniz *et al.*, 1999).

Parameter extraction (iv) is performed in a simple and readily implementable form as the Cepstral coefficients obtained from the Fourier transform analysis (Lima *et al.*, 2000; Diniz *et al.*, 1999). The speech signal $s(n)$ can be considered as a convolution between an excitation signal $u(n)$ with the human vocal tract $h(n)$, as shown by Equation 1. The vocal tract is a time-varying system modeled by the variable filter $h(n)$; this is the reason because the speech signal must be segmented in a short-time range before being processed by the Fourier analysis. Then, speech is analysed in both time and frequency, leading to the spectrogram, as shown in Figure 5. Cepstral analysis performs deconvolution between excitation and vocal tract response, according to Equation 2. The later one can be used for pattern recognition. Twelve Cepstral coefficients belong to the vocal tract (Rabiner and Schafer, 1978) were extracted, discarding the zero-index one (Deller *et al.*, 1993), along fifty segments, resulting in six hundred parameters per word for pattern recognition.

$$s(n) = u(n) * h(n) \tag{1}$$

$$c_S = IDFT\{\log|DFT[s(n)]|\} = IDFT\{\log|S(k)|\} = \tag{2}$$

$$= IDFT\{\log|U(k)||H(k)|\} = IDFT\{\log|U(k)| + \log|H(k)|\} = c_U + c_H$$

where:

- IDFT: inverse discrete Fourier transform
- $S(k)$: DFT of the speech signal
- $U(k)$: DFT of the excitation
- $H(k)$: DFT of the vocal tract
- c_S: Cepstrum of the speech signal
- c_U: Cepstrum of the excitation
- c_H: Cepstrum of the vocal tract

The six hundred-dimensional data form a random vector that, with all realizations comprised by the repetitions for all the commands to be recognized, form a data set that is used for training a NN for the pattern recognition stage. In a former implementation, two parallel NN of different topologies were used for a voting system (Jorge *et al.*, 2010).

Currently, only one feed-forward NN (FFNN) trained with a more robust backpropagation (BP) -based training algorithm is used instead. The training algorithms implemented perform adaptation of learning rates (Jacobs, 1988; Cichocki and Unbehauen, 2003), what leads to faster convergence in flat regions in the search space, while slower convergence around minima. The effect of this approach is a more robust convergence in regions quite different from parabolic-like minima, as steep gutter-like ones, for example. This originated a class of backpropagation algorithms known as resilient backpropagation (RPROP), (Riedmiller and Braun, 1992), more robust for convergence.

The current implementation is based in part on the Silva and Almeida algorithm (Silva and Almeida, 1990; Cichocki and Unbehauen, 1993), according to Equation 3. Global adaptive learning rate was also implemented (Cichocki and Unbehauen, 2003), similar to the local one but using only one learning rate for the whole NN. The implemented code enables some choices for users, as: (i) online or batch training (Haykin, 1999); (ii) possible use of moment (Haykin, 1999); (iii) global or local adaptive learning rate (Cichocki and Unbehauen, 2003).

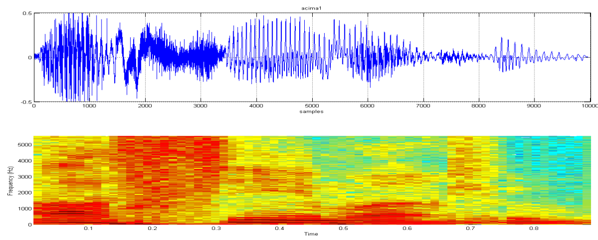

Figure 5: The time-domain spoken command “acima” (Portuguese word for “up”) in the upper part; and its corresponding spectrogram in the lower part, where it is possible to notice the time-varying nature of the speech spectrum.

4.2. Head Tracking

Two approaches for head tracking were implemented: with and without visual markers. But in both cases, the purpose is to enable a more natural interaction between users and system. Head tracking turned out to be an interesting possibility after some experimentation with the ASR system. A user had to talk commands to do anything during operation of the VCD, since moving its view up, down, left or right to zooming it in or out, or switching view among different VCD’s modules (see Figure 3, where it is possible to see three main VCD’s modules). Imagine that user had to have a close view of some detail in the left module, for example; he or she would have to switch view to that module, then adjust the zoom to a specific indicator, every action controlled by voice commands; it might be a difficult task to speak repeatedly “left”, “right”, and so on, until focusing in the desired indicator. Thus, the R&D staff became searching for other types of interaction modes. In fact, both the ASR and the head tracking systems can be combined depending upon the task to be executed. With the head tracking approach, the above mentioned task (seeing a detail information in the left VCD module) could be executed by moving user’s head to the left, and the image would then turn to that side; and to look in more detail an indicator, he or she needed to approach head towards the screen, and the projected image would zoom in. Head tracking is based on a six-degree of freedom (6 DoF) information, three for displacements relatively to the three Cartesian axes, and other three for rotation angles relatively also to the same three axes, as illustrated by Figure 6. This work makes use of only three degrees of freedom that are, according to Figure 6: (i) yaw, (ii) pitch and (iii) forward-back, the later to enable zooming in and out the VCD.

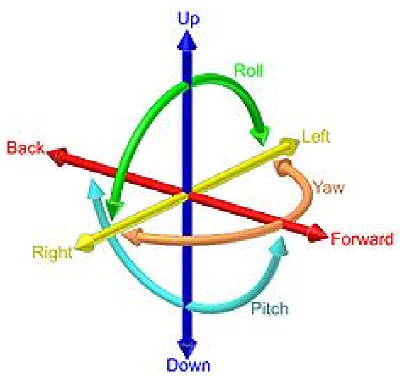

Figure 6: 6 DoF for head tracking.

4.2.1. Head Tracking with Visual Markers

This interaction mode makes use of visual markers attached to user’s head, with an infrared (IR) sensor, – Trackir5, supplied by Natural Point (www.naturalpoint.com) –, enabling head pose estimation. Three reflective markers are fixed at user’s head. The pose is estimated based on projective geometry computation, by a freeware library, – OptiTrack –, supplied by the same company. Tests showed this approach results in a good accuracy in head pose estimation. Figure 7 shows a view of this interaction mode in operation.

Figure 7: The head tracking interaction mode with visual markers.

4.2.2. Markerless Head Tracking

Another approach was also implemented and tested, markerless head tracking system. The code used performs head pose estimation based on tracking some points detected in users face, with six degrees of freedom, and is a proprietary library named FaceAPI, supplied by Seeing Machines (www.seeingmachines.com). The source code is not available, and the company does not supply any details about the tracking methodology used. Thus, it is used as an executable called by our application. It operates with either webcams or IR cameras.The disadvantage of this approach, after performing some tests, is that it does not has good accuracy as the other approach with visual markers, specially when user turns his or her head to the sides; at twenty degrees to both sides, the estimation of head angle sometimes oscillates, making the VCD image on screen to shake for both sides. The advantage is that interaction is more natural, since there is no need of using markers on user’s head.

Another disadvantage is that this system can not be used with head-up display. Since it is trained to detect faces, any other device in one’s head makes face undetectable by the code. Other details such as glasses or beard can also cause problems in face detection.

Besides this, tests showed that using the head tracking system (be it with or without markers) as the only interaction mode would also cause problems, similarly as pointed out in the first paragraph of section 4.2. In that case, executing all commands by voice would be a difficult task, but in the present one too, as explained in the following. Imagine the same situation mentioned in the first paragraph of section 4.2, repeated here for convenience: say a user had to have a close view of some detail in the left module, for example; he or she would have to switch view to that module, then adjust the zoom to a specific indicator, every action controlled by head movements. It might be a difficult task and cause discomfort for user to move head to the left sufficiently to switch to the left VCD module's view, and then approaching his or her head to the projection or computer screen for zooming in, until focusing in the desired indicator. That is the reason for using both interaction modes, as explained in Section 4.2. Figure 8 shows a screen shot of the markerless head tracking system in operation; Figure 8a shows a whole view of this system, while Figure 8b shows a detailed view of the points tracked in the user's face.

a)

b)

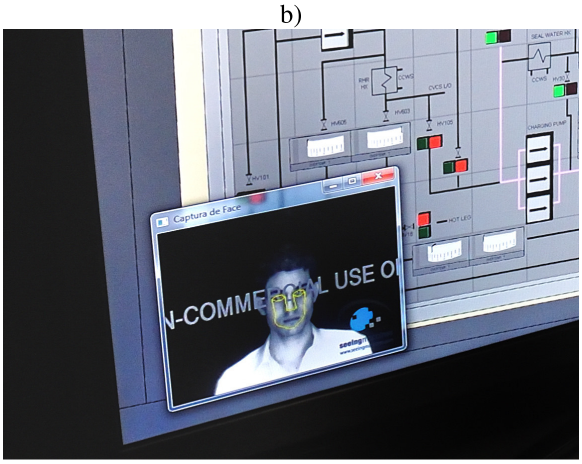

Figure 8: a) The markerless head tracking system; b) A detailed view of the points tracked in the user's face.

4.3. Body Tracking

The Kinect sensor was used for a full computational viewing of the operator body, to operate the VCD without any kind of mechanical interface with the operator.

The Kinect sensor was originally intended to be a device that recognizes the user's movements to the MicroSoft Xbox 360 (www.xbox.com), which allows him to control games through gestures and voice commands. Its main hardware components are an RGB camera, a depth sensor that consists of an infrared light emitter and a camera to get this emitting light, multi-array microphones, an engine tilt, and three-axis accelerometer.

The Kinect depth sensor (wikipedia.org/wiki/Kinect) consists of an infrared light source, an emitter projecting a pattern of dots that are read back by a monochrome infrared camera, which is called structured light (wikipedia.org/wiki/Structured light). The sensor detects segments reflected by the pattern dots and converts the images into a depth map, so that the image has, besides the x, y coordinates, the z axis distance of the objects in the scene image to the Kinect sensor.

In this project, we used the open source OpenNI library that contains a middleware (NITE), which, through its algorithm, can recognize the shape of the user's body and provides the x, y, z body coordinates, such as: head, shoulders, chest, hands, etc. so that through them you can make an interaction with VCD. It also presents the image using the Kinect OpenGL graphical library format, which is the same used in the VCD.

To capture the coordinates three softwares were used:

1) OpenNI

The OpenNI is a software design framework (OpenNI.Org) focused on interoperability of natural interaction devices.

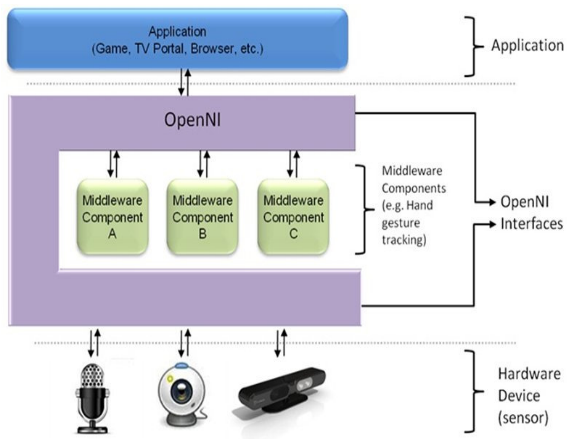

Figure 9: OpenNI Arquiteture

2) Kinect Driver sensor for Windows

The idea of OpenNI is to support various types of natural interaction devices. This specific Microsoft Kinect Sensor driver for Windows must be installed (github.com/avin2/SensorKinect).

3) NITE

This software is a middleware made to operate together with OpenNI. It was developed by Prime Sense (www.primesense.com/nite), which is the company that makes the Kinect hardware. It analyzes the image with depth information generated by Kinect and generates the x, y, z point coordinates of the operator's body.

We used an application of NITE, "players", which generates 14 coordinates x, y, z of the skeleton of the operator, as shown in Figure 10. The advantage of using this application is that the generated image is made using the OpenGL visualization library, which is the same used in VCD, facilitating the integration of the two programs. The final program displays two windows and the skeleton of the VCD, as can be seen in figure 11.

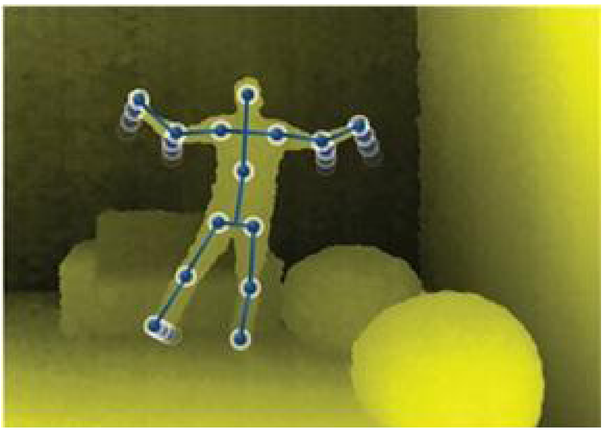

Figure 10: Joints of the skeleton generated by the "players"

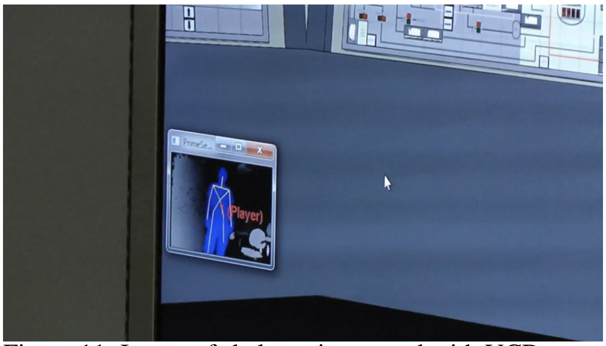

Figure 11: Image of skeleton integrated with VCD

The desired controls for natural interaction with VCD are:

- VCD rotation on its vertical axis.
- Zoom to Observe Details of the VCD.
- Control Actuation of VCD.
- Show Predetermined Regions of Interest of the VCD.

4.3.1. VCD rotation on its axis vertical

The NITE model treats the head as if it were a unique skeleton joint as is impossible to make head tracking, with a unique coordinate.

To face the problem of getting the head tracking for VCD rotation we used the body rotation angle of the shoulders position. The arc tangent between the coordinates of the shoulders was used, where:

$\theta = atg((\Delta \text{ shoulders } z) / (\Delta \text{ shoulders } x))$ (3)

This approach is practical because it is independent of the distance from the sensor to the operator.

4.3.2. Zoom to observe details of the VCD

The Zoom function was used with the z coordinate of the chest, making the visualization of VCD turn larger or smaller, with the approach or departure of the user's body.

4.3.3. Control actuation of VCD

The control actuation of the VCD is made by the buttons of the desk, that are the way to pass the user information to program operation of the simulator, such as: turn pumps on or off, deploy control rods in a deliberate shutdown of the reactor, etc.

With Kinect, the cursor movement is done by moving the right hand. With OpenNI NITE middleware, using the application "Players", it is possible to extract the centroid x, y position of the right hand in the space depth map.

With the mouse Windows API, it is possible to match the hand's position in 640X480 points resolution (depth map) to place the cursor on any size of viewed VCD screen.

For the actuation of the control buttons of the VCD, we used a methodology to recognize the opening and closing of the right hand to click the cursor.

NITE provides the position of the centroid of the hand. The space around the centroid of the hand was analyzed in the depth map. This area constitutes a square of approximately 20 cm sides that corresponds to a 50x50 dot matrix in the dept map.

With access to this matrix, we can analyze the difference of the Z coordinate of each point, relative to the Z coordinate of the centroid point of the hand, as seen in Figures 12 and 13. If the absolute value of each difference is less than 20 cm, the result is included in a summation. As the area of the open hand on the matrix is greater than that of the closed hand, the summation of the differences is bigger. This process is repeated for each generation of VCD frame. If the summation is 40% larger or smaller than the previous one, it is determined whether the hand is open or closed and thus the click is triggered or not.

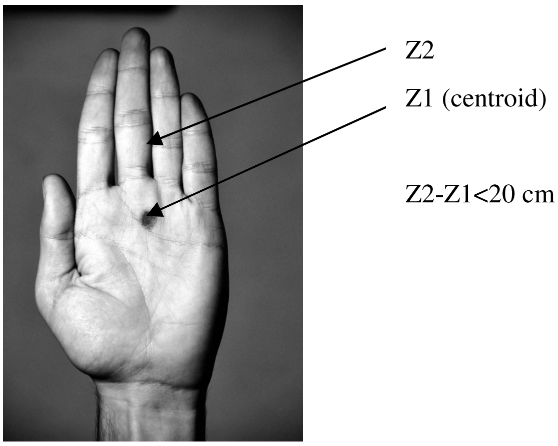

Figure 12: Open hand Centroid

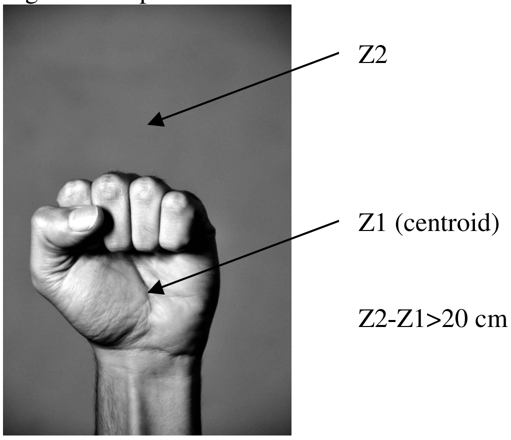

Figure 13: close hand Centroid

4.3.4. Predetermined Regions of Interest of the VCD

The VCD has five predetermined regions of interest to facilitate rapid observation of the VCD by the operator. To operate Kinect we used the X axis of the left hand movement. The image of the VCD is calculated continuously, generating multiple frames per second. In each VCD frame the current position is analyzed and compared to the previous left hand position, calculating the centroids differences in relation to the x-axis. A negative difference decreases the region of interest and a positive increases it, making a sequential selection of the regions of interest.

5. RESULTS

5.1. Comparative Analysis between marker-based and markerless head tracking systemsis Length of the Paper

This section shows a comparative analysis between the two head tracking methods used in this work, with and without markers. The user's head angles for the left and right sides were measured for both methods, approximately around three positions: (i) frontal, with user looking directly to the screen; (ii) rotated to the left, approximately at −20 degrees; and (iii) rotated to the right, approximately at +20 degrees. Table 1 shows results for the marker-based method (using TrackIR5), while Table 22 shows similar results for the markerless method (using FaceAPI). Each line in both Tables is in fact an average among ten measurements. One can verify the higher standard deviation (Std. Dev.) for the later method, what means a higher pose estimation instability for angles around 20 degrees, to the left or to the right sides. Thus it follows some advantages and disadvantages of both methods in Table 3.

Table 1: Marker-based tracking performance

	FRONTAL		ROTATED TO THE LEFT		ROTATED TO THE RIGHT	
	AVERAGE (DEGREES)	STD. DEV. (DEGREES)	AVERAGE (DEGREES)	STD. DEV. (DEGREES)	AVERAGE (DEGREES)	STD. DEV. (DEGREES)
1	0,508	0,200	-26,72	0,190	24,369	0,097
2	1,306	0,177	-22,15	0,134	27,679	0,139
3	0,256	0,237	-20,22	0,151	27,696	0,125
4	-0,758	0,214	-25,20	0,156	28,001	0,129
5	0,566	0,291	-21,48	0,078	26,597	0,151
6	-1,179	0,263	-17,06	0,115	26,966	0,172
7	-0,153	0,183	-18,28	0,135	28,669	0.127
8	0,015	0,155	-19,02	0.087	24,890	0,159
9	-0,744	0,140	-22,31	0,083	26,795	0,115
10	0,657	0,211	-22,03	0,170	20,284	0,180

Table 2: Markerless tracking performance

	FRONTAL		ROTATED TO THE LEFT		ROTATED TO THE RIGHT	
	AVERAGE (DEGREES)	STD. DEV. (DEGREES)	AVERAGE (DEGREES)	STD. DEV. (DEGREES)	AVERAGE (DEGREES)	STD. DEV. (DEGREES)
1	-0,566	0,231	-16,24	0,576	14,218	1,927
2	-0,432	0,153	-15,05	1,396	14,307	1,509
3	-0,625	0,122	-16,90	1,286	18,249	1,889
4	-0,512	0,161	-16,20	0,963	15,518	1,462
5	-0,072	0,119	-14,59	1,589	13,266	4,203
6	-0,619	0,155	-13,22	0,586	16,255	0,607
7	-0,336	0,406	-20,97	0,343	17,923	1,18
8	-1,937	0,174	-17,44	0,531	13,549	0,353
9	1,117	0,3425	-15,28	1,475	12,742	0,552
10	-1,218	0,164	-16,38	0,861	14,417	0,648

Table 3: Comparative analysis between the marker-based and the markerless tracking methods relatively to some usage parameters

	MARKER-BASED METHOD	MARKERLESS METHOD
CAMERA TYPE	Proprietary IR camera	General purpose camera (webcam)
CAMERA CHARACTERISTICS	Operates with low illumination	Higher tracking errors for low sensitivity cameras and low frame rates (requires minimum of 30 fps)
POSE ESTIMATION	Robust pose estimation even for high rotation angles	Poorer pose estimation for high rotation angles
TRACKING ERRORS	Losses tracking if there are reflexive objects in scene	Do not loose tracking when new face appears in scene

INITIALIZATION AND OPERATION	Immediate tracking from the beginning	Requires from 5 to 6 s from initial to current tracking
SOFTWARE AND LIMITATIONS	Free software, can be used for commercial purpose by $150.00 (including camera)	Free software demo for non-commercial applications, full software with paid license
COUPLING WITH HEAD-UP DISPLAY	Can be coupled with head-up display	Can not be coupled with head-up display

5.2. Interaction based on the ASR system

Interaction based on the ASR system was analyzed by some figures of merit, as explained in the sequel. As already mentioned, the current ASR system's implementation makes use of FFNN. First figure of merit is the word error rate (WER), defined as:

$$\mathrm{WER} = (S + I + D)/N \tag{4}$$

where:

- N: total number of words to be recognized
- S: number of replaced words
- I: number of inserted words
- D: number of deleted words

The three later are errors. WER should be ideally zeroed. Another figure of merit is the word recognition rate (WRR), defined as:

$$\mathrm{WRR} = (1 - \mathrm{WER}) \tag{5}$$

which should be ideally one. We show these later results, in the sequel.

Another important figure of merit is the real time factor (RTF), defined as:

$$\mathrm{RTF} = TP/TA \tag{6}$$

where:

- TA: the input word duration
- TP: corresponding processing duration.

This should ideally be as low as possible.

Experiments were carried out by using cross-validation (Haykin, 1999), where the data set was split into a number of subsets, some of them comprising the training subset, and the remaining, the test subset. Learning was repeated, each time using a different subset as the test one. Table 4 shows the scheme adopted. The whole set was split into four subsets, three of them comprising the training subset at each learning repetition. Bold characters indicate the test subsets. The last two columns show the figures of merit's results for each individual experiment run.

Table 4: Cross-validation scheme adopted

	Subsets				WRR (%)	RTF
Experiment 1	**25 %**	25 %	25 %	25 %	94.5	0.026
Experiment 2	25 %	**25 %**	25 %	25 %	98.2	0.023
Experiment 3	25 %	25 %	**25 %**	25 %	97.1	0.023
Experiment 4	25 %	25 %	25 %	**25 %**	93.5	0.025

The resulting average WRR among the four learning repetitions was: WRR = 95.8 %. The resulting average RTF was: RTF = 0.024.

5.3. Body Tracking

Tests were made taking the screen cursor in a lower region of the screen and placing it on a upper button and triggering it by closing the hand. This procedure was done with two users to test diversity operation between users.

Thirty samples were conducted of two users in two sets of tests, separated by a space of time, so that the sample would not be users addicted (Table 5).

Table 5: Press Button Tax error

30 presses (unit 1)				30 presses (unit 2)		
	Right presses	Wrong presses	error (%)	Right presses	Wrong presses	error (%)
user 1	26	4	13,3	25	5	16,6
user 2	24	6	20	27	3	10

6. CONCLUSIONS

The developed VCD seems to be a very good alternative relatively to the synoptic windows-based approach, by aggregating both the high fidelity visual appearance with the corresponding real control desk, and the computer-based PWR NPP simulation system's functionalities.

Relatively to the interaction modes, it seems that a combination of the ASR with either one of the head tracking systems would result in a more natural users' interaction with the VCD. Each interaction mode is used for the tasks it serves best, with ASR performing better for switching among different VCD views, – what one could call macro movements –, and with head tracking performing better for small movements within a particular VCD module view, – what one could call micro movements.

The intention to use the Kinect Sensor, was integrate all the above mentioned interfaces in one device. Making use of the body to control the VCD. Changing regions of interest using the left hand, their microphones for voice commands and the right hand to control the screen cursor and your click, making the user interface with the VCD more natural as possible. With the use of the various possibilities of interfaces, the intention of this work was offer to the user several options for their interaction with the VCD, and so enable him to make a choice for their preferred use. MCV developed, with its low cost of construction, its similarity to the original control board and its natural interfaces, which help the user to have interaction more friendly, show the originality of this work in the nuclear field.

7. REFERENCES

Aghina, M. A. C.; Mól, A. C. A.; Jorge, C. A. F.; Pereira, C. M. N. A.; Varela, T. F. B.; Cunha G. G.; Landau, L.; 2008. Virtual control desks for nuclear power plant simulation: improving operator training, *IEEE Computer Graphics and Applications*, Vol. 28, No. 4, 6-9.

Aghina, M. A. C.; Mól, A. C. A.; Jorge, C. A. F.; Espírito Santo, A. C.; Freitas, V. G. G.; Lapa, C. M. F.; Landau, L.; Cunha, G. G.; 2011. Virtual control desks for nuclear power plants, In: Tsekov, P.; *Nuclear Power – Control, Reliability and Human Factors*, InTech: Rijeka, Croatia, 393-406.

Carvalho, P. V. R.; Obadia, I. J.; 2002. Projeto e implementação do laboratório de Interfaces Homem Sistema do Instituto de Engenharia Nuclear, *Revista Brasileira de Pesquisa e Desenvolvimento*, Vol. 4, No. 2, 226-231.

Carvalho, P. V. R.; Santos, I. J. A. L.; Gomes, J. O.; Borges, M. R. S.; Guerlain, S.; 2008. Human factors approach for evaluation and redesign of human-system interfaces of a nuclear power plant simulator, *Displays*, Vol. 29, No. 3, 273-284.

Cichocki, A.; Unbehauen, R.; 1993. *Neural Networks for Optimization and Signal Processing*, John Wiley & Sons.

Deller Jr., J. R.; Proakis, J. G.; Hansen, J. H.; 1993. *Discrete-Time Processing of Speech Signals*, MacMillan, New York.

Diniz, S.; Thomé, A. G.; Santos, S. C. B.; Silva, D. G.; 1999. Automatic speech recognition: a comparative evaluation between neural networks and hidden markov models, *International Conference on Computation Intelligence for Modeling, Control and Automation (CIMCA 99)*, Vienna, Austria, February 17 – 19.

Drøivoldsmo, A.; Louka, M. N.; 2002. Virtual reality tools for testing control room concepts, In: Liptak, B. G.; *Instrument Engineer's Handbook: Process Software and Digital Networks*, third ed., CRC Press.

Foley, J. D.; Wallace, V. L.; Chan, P.; 1998. *Human Computer Interaction*, Prentice-Hall.

Hanes, L. F.; Naser, J.; 2006. Use of 2.5D and 3D technology to evaluate control room upgrades, *The American Nuclear Society Winter Meeting & Nuclear Technology Expo*, Albuquerque, NM, 12 to 16 November.

Haykin, S.; 1999. *Neural Networks – A comprehensive foundation*. Upper Saddle River: Prentice-Hall.

IAEA TECDOC 995; 1998. Selection, specification, design and use of various nuclear power plant training simulators.

ICRP Publication 60; 1991. ICRP Publication 60 – Recommendations of the International Commission on Radiological Protection.

Ishii, H.; 2008. The Non-conventional User Interface and Its Evolution, *Communications of the ACM (CACM) – Special Issue "Organic User Interfaces"*, Vol. 51, No. 6, 32-36.

Jacobs, R. A.; 1988. Increased rates of convergence through learning rate adaptation, *Neural Networks*, Vol. 1, 295-307.

Jorge, C. A. F.; Mól, A. C. A.; Pereira, C. M. N. A.; Aghina, M. A. C.; Nomiya, D. V.; 2010a. Human-system interface based on speech recognition: application to a virtual nuclear power plant control desk, *Progress in Nuclear Energy*, Vol. 52, No. 4, 379-386.

Jorge, C. A. F.; Mól, A. C. A; Couto, P. M.; Pereira, C. M. N. A.; 2010b. "Nuclear plants and emergency virtual simulations based on a low-cost engine reuse", In: P. V. Tsvetkov (Ed.); *Nuclear Power*, InTech: Rijeka, Croatia, 367-388.

Kim, Y. H.; Park, W. M.; 2004. Use of simulation technology for prediction of radiation dose in nuclear power plant, *Lecture Notes in Computer Science*, Vol. 3314, 413-418.

Lima, A. A.; Francisco, M. S.; Lima Netto, S.; Resende Jr., F. G. V.; 2000. Análise Comparativa de Sistemas de Reconhecimento de Voz, *Simpósio Brasileiro de Telecomunicações*, Gramado, Rio Grande do Sul, Brazil, p. 001-004.

Markidis, S.; Rizwan-uddin; 2006. A virtual control room with an embedded, interactive nuclear reactor simulator, *The American Nuclear Society Winter Meeting & Nuclear Technology Expo*, Albuquerque, NM, 12 to 16 November.

MSDN, Microsoft Developer Network; http://msdn.microsoft.com (Most recent access in April 2012).

NaturalPoint a; TRACKIR, by Natural Point, Inc.: http://www.naturalpoint.com/trackir/ (Most recent access in April 2012).

NaturalPoint b; OptiTrack, by Natural Point, Inc.: http://www.naturalpoint.com/optitrack/ (Most recent access in April 2012).

Nitendo; Wii system, by Nitendo of America, Inc.: http://www.nitendo.com/wii/ (Most recent access in April 2012).

Nystad, E.; Strand, S.; 2006. Using virtual reality technology to include field operators in simulation and training, *27th Annual Canadian Nuclear Society Conference and 30th CNS/CNA Student Conference*, Toronto, Canada, 11 to 14 June.

Oliveira, M. V.; Moreira, D. M.; Carvalho, P. V. R.; 2007. Construção de interfaces para salas de controle avançadas de plantas industriais, *Ação Ergonômica*, Vol. 3, 8-13.

Oppenheim, A. V.; Schafer, R. W.; 1989. *Discrete-Time Signal Processing*. Englewood Cliffs: Prentice-Hall.

Pinto, R. G.; Pinto, H. L.; Calôba, L. P.; 1995. Using neural networks for automatic speaker recognition: a practical approach, *38th IEEE Midwest Symposium on Circuits and Systems*, Rio de Janeiro.

Rabiner, L. R.; Schafer, R. W.; 1978. *Digital Processing of Speech Signals*. London: Prentice-Hall.

Riedmiller, M.; Braun, H.; 1992. RPROP – a fast adaptive learning algorithm, *Seventh International Symposium on Computer and Information Sciences (*ISCIS VII*)*, Antalya, Turkey.

Santos, I. J. A. L.; Teixeira, D. V.; Ferraz, F. T.; Carvalho, P. V. R.; 2008. The use of a simulator to include human factors issues in the interface design of a nuclear power plant control room, *Journal of Loss Prevention in the Process Industries*, Vol. 21, No. 3, 227-238.

SeeingMachines; FaceAPI, by Seeing Machines: http://www.seeingmachines.com/product/faceapi/ (Most recent access in April 2012).

Silva, F. M.; Almeida, L. B.; 1990. Speeding up backpropagation, In: Eckmiller, R. (Ed.); *Advances of Neural Computers*. Elsevier Science Publishers, p. 151-158.