Elucidation of Perceptual and Behavioral Human Characteristics in Virtual Environments

-Toward the Elucidation of a Sensation of Presence-

Susumu Tachi

Faculty of Engineering,
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113 JAPAN
tachi@star.t.u-tokyo.ac.jp
http://www.star.rcast.u-tokyo.ac.jp/

Abstract

The American Heritage Dictionary defines "virtual" as "existing in essence or effect though not in actual fact or form," and this definition also applies to virtual reality (VR) per se. For the VR users, the "virtual reality" exists "in effect or in essence" although "not in actual fact or form." The Japanese national project, "Fundamental Study on Virtual Reality -Generation of Virtual Space and Human Interfaces for Virtual Environments-," tries to clarify how the man made environment can be made "virtual" for human beings in the above sense. The project has been led by the author and proceeded by four groups. This paper reports the goal and scope of the project per se and also reports some of the interim results of the first group of the project, i.e., the elucidation of perceptual and behavioral human characteristics in virtual environments.

Key words: Virtual Reality, Priority Scientific Research Area, Sensation of Presence, Reality, Sensation of Existence

1. Introduction

Virtual reality is a technology which enables a human being to "virtually" exist in a computer generated environment or a remote environment where his/her surrogate exists and can gain various experiences, can communicate with others and can conduct tasks while interacting with the virtual environment. This technology is gradually being applied to several areas in our society today.

However, fundamental research on human behavior in virtual environments and the effect of these virtual environments on human beings is only at a preliminary stage.

The Japanese national project of "Fundamental Study on Virtual Reality -Generation of Virtual Space and Human Interfaces for Virtual Environments-," was launched in 1995 supported by the Ministry of Education, Science, Sports and Culture (Monbusho) as a Priority Scientific Research Area. The project is led by the author (Prof. S. Tachi of the University of Tokyo), and more than 200 professors and researchers of the universities in Japan joined the project. In this study, fundamental research on virtual reality has been conducted in the following groups:

1)Elucidation of perceptual and behavioral human characteristics in virtual environments (Group Leader: S. Tachi).

2)Optimal design of interface devices between human and virtual environments (Group Leader: M. Sato).

3)Establishment of virtual world construction and/or representation methods (Group Leader: M. Hirose).

4)Assessment of the influence of virtual reality technology on human society and/or human health and welfare (Group Leader: T. Ifukube).

This paper reports some of the interim results conducted in the first group. The results of other three groups are reported in other papers written by other group leaders in this Proceeding of ICAT'97.

2. Elucidation of perceptual and behavioral human characteristics in virtually environments

Virtual space constructed for virtual reality must have three basic characteristics, i.e., 1) three dimensional space structure which faithfully represents the original spatial structure, 2) three dimensional constructed virtual space which follows precisely to head and/or torso movement and human real time interaction with the constructed space, and 3) self projection, i.e., coherent sensation of proprioception with vision, audition and taction.
The above three conditions are satisfied in a tele-existence system. By using a tele-existence system, a human operator has a feeling of being inside a remote robot with three dimensional visual and auditory sensation controlled

precisely by the motion of the operator, and a manipulator interacting with an environment, which is seen at the place where his/her real arm is felt by his/her propioception. A tele-existence manipulation system was constructed and evaluated quantitatively by comparing tasks of tracking a randomly moving target under several operational conditions. The effects of various characteristics, e.g., binocular vision and the effect of natural arrangement of the head and the arm, were analyzed by comparing quantitatively the results under these operational conditions. Human tracking transfer function was measured and was used for comparison. The results revealed the significant dominance of the binocular vision with natural arrangement of the head and the arm [1], which demonstrated the importance of the above three conditions of virtual reality (the three conditions of VR).

The concept of tele-existence in a remote environment using a surrogate robot is easily extended to the tele-existence in virtual environment using a surrogate virtual human, which satisfies the three conditions of VR.

The aim of the study is to elucidate the human perceptual and behavioral mechanisms in three dimensional space using computer generated virtual environments, especially to elucidate how a sensation of presence or a sensation of existence emerges.

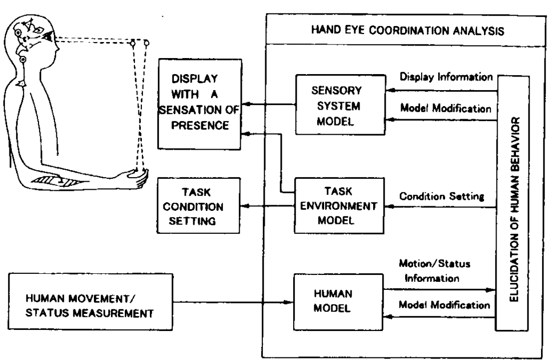

Fig. 1 Schematic Diagram on an Approach to the Elucidation of Human Perceptual and Behavioral Characteristics

Figure 1 shows an example of the approach. A human subject is asked to interact with a computer generated virtual environment, of which spatial parameters are controlled by the computer, while the human movement and/or internal status are measured in real time and is used to evaluate quantitatively the sensation of presence by comparing his/her behavior in the real environment as a control.

A human perceptual and behavioral model will be proposed and tested by inputting the same virtual environment and observing the output behavior of the model. By modifying the model to show the same behavior of the real human, the perceptual and behavioral model will be obtained.

3. Results

The following sections show the interim results of the members of our research group (Elucidation of Perceptual and Behavioral Human Characteristic in Virtual Environment). The results will be updated on the web [2]. In order to clarify the originality and credibility of the research, each section has a name of each principal investigator who is responsible for each work. Questions and comments should be sent directly to each principal investigator at the e-mail address which is cited at each section title.

The results are not integrated as in the form of chapter 2, for they are interim. For further results, please check the web [2].

3.1 S. Tachi (tachi@star.t.u-tokyo.ac.jp)

A method is proposed to represent haptic space with edges, vertices and surfaces with mechanical impedance information. The system is intended to provide not only force feedback but also shape information of the contact object during large space interaction with surfaces of various shapes and mechanical impedance. A user wears a light weight passive seven-degree of freedom exoskeleton goniometer that measures the position and orientation of a fingertip. A six-degree of freedom manipulator, which we call "Active Environment Display (AED)," moves to anticipate and if necessary counter the user's action to display contact with virtual surfaces, edges and vertices. The robot manipulator is a pantograph with range of motion of 60x60x60 cm for translation and 360, 180, and 135 degrees for yaw, roll, and pitch, respectively. The wrist of the manipulator carries a device with a complex surface geometry as well as convex and concave edges and flat surfaces. It is thus possible to simulate contact with continuous surfaces and edges by moving the Shape Approximation Device (SAD). Contiguous surfaces are displayed by reorienting the corresponding facet of SAD. A User wears a head mounted display which immerses him/her in the virtual environment. Thus the user can feel exactly what (s)he sees at the position it is observed with the haptic shape just as it is seen. Experiments are conducted to present several shape objects with several kinds of mechanical impedance. It has been shown that an object shape with vertices, edges and surface can be represented. Concave edges as well as convex edges are represented successfully using the test hardware [3]. Effect of touching the same shape exactly at the same place as is seen is now being evaluated.

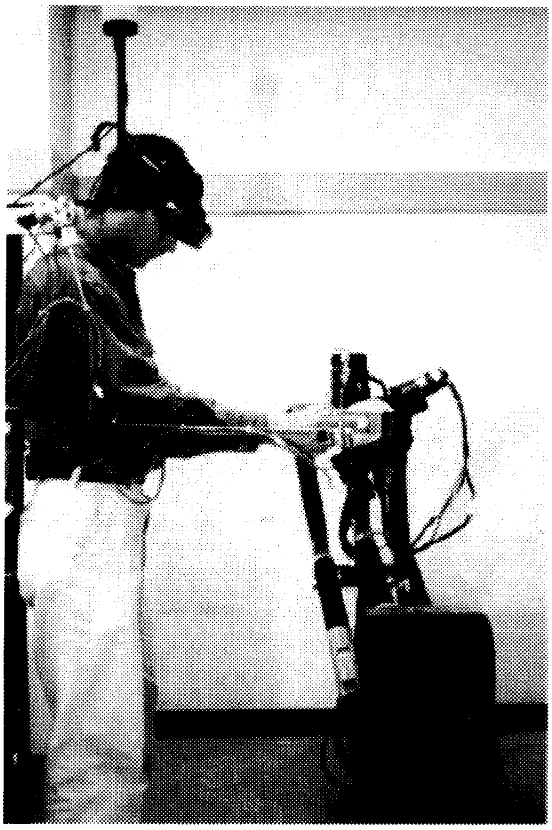

Fig. 2 General View of the Haptic Graphic Device which Generates Shape of a Virtual Object

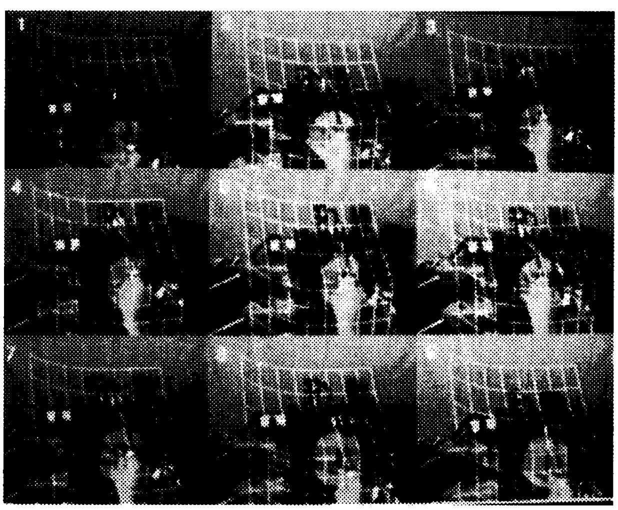

Fig. 3 A Construction of two surfaces with an edge.

3.2 A. Mikami (mikami@smtp.pri.kyoto-u.ac.jp)

In the primate visual cortex, visual information is processed in two streams, a dorsal stream (spatial vision) and a ventral stream (form vision). Integration of both of streams is necessary to create images with reality. The middle portion of the superior temporal sulcus (STS) is a candidate area to do this job. The STS is know to be closely related with the inferior temporal cortex (IT) on the ventral stream, and anatomically connected with area 7a, LIP and MST in the dorsal stream. Neuronal activity from the STS of a monkey during the performance of the visual discrimination and memory task using visual stimuli with components of both form and motion was recorded. About a half of neurons analyzed were selective for both form and direction of motion. The data suggested that the STS plays a role in the integration of visual form and motion information [4].

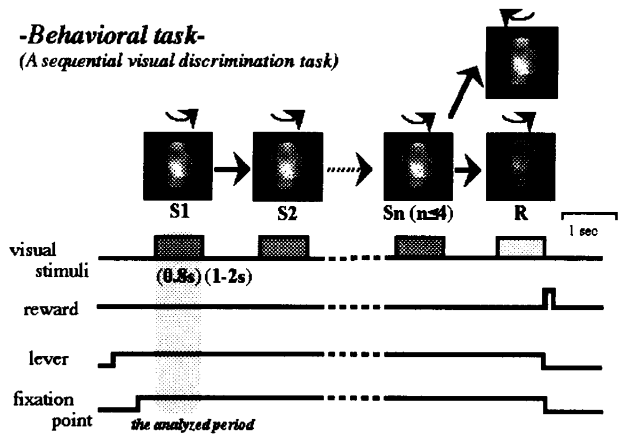

Fig. 4 In this task, visual stimuli with components of both form and motion were used. While the monkey pressed a lever to fixate a small fixation point (FP) in the center of the screen, a visual stimulus was presented for 0.8 s and was repeated 1-4 times. A delay period of 1-2 s was inserted between stimulus-presentations. The monkey was rewarded for releasing the lever, when the presented stimulus differed in either form or motion from the first stimulus. Therefore while performing this task, the monkey was actively attending to both the form and motion components.

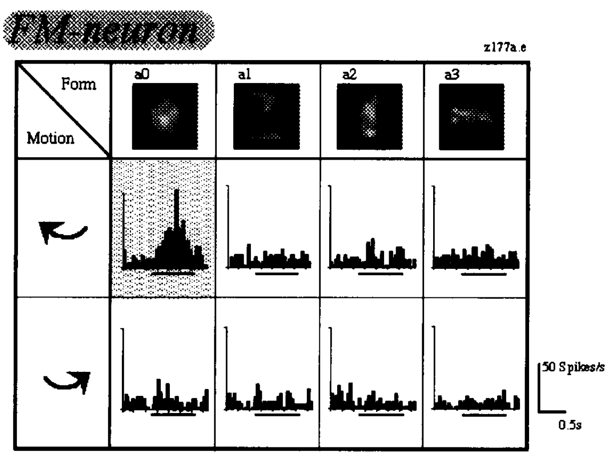

Fig. 5 Each vertical column represents 4 different forms. The upper row represents clockwise motion. The lower row represents counterclockwise motion. Each horizontal bar represents the S1 period. This neuron was activated by the form (a0) with a particular motion (clockwise).

3.3 M. Ohmi (ohmi@orchid.mattolab.kanazawa-it.ac.jp)

Body sway was induced with observing real-world images taken from a moving car on curvilinear path. Amount of body sway increased linearly as a function of centrifugal acceleration of the car. It shows that sense of presence can be quantified by amount of body sway. It was also found by factor analysis that increase of sense of presence was accompanied with physiological fatigue [5].

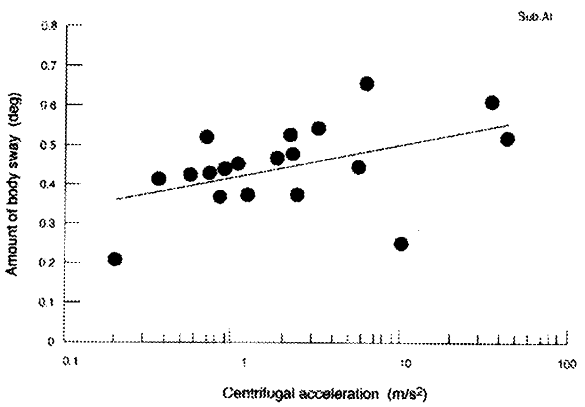

Fig. 6 Relationship between centrifugal acceleration and body sway induced with observing real-world images taken from a moving car on curvilinear path. Since amount of body sway increased linearly as a function of centrifugal acceleration of the car, it shows that sense of self motion, namely sense of presence, can be quantified by amount of induced body sway.

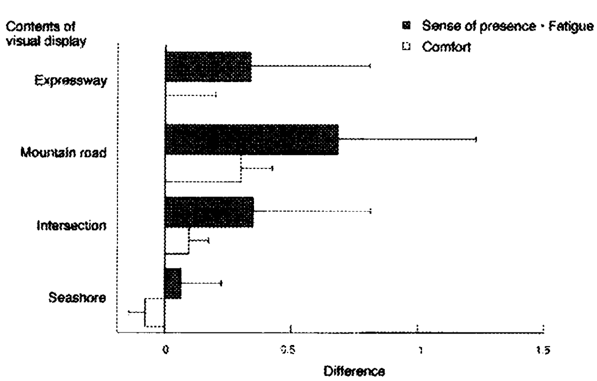

Fig. 7 Difference of evaluated values of psychological effects with viewing 2D- and 3D-display measured by the semantic differential method. Although sense of presence increases with 3D-display, it is accompanied with physiological fatigue.

3.4 M. Taira (masato@med.nihon-u.ac.jp)

Recent neuro-physiological and clinical neuro-psychological studies have suggested that the parietal association cortex may play an important role in perception of 3D features of the object. For 3D visual information process in this area, the binocular disparity cue seems to be the most important. In this study, whether the monkey could discriminate the 3D features of the object by integrating the binocular disparity cues was examined. The monkey was trained to perform a delayed-matching-to-sample(DMTS) task using computer generated 3D figures with binocular disparity or in the form of random-dot stereogram (RDS). The success rate of the DMTS task with RDS figures became about 85% after enough training. This suggests that the monkey can discriminate the 3D features of the object by binocular disparity cue. Furthermore, we reversibly inactivated the parietal cortex of the monkey by microinjection of muscimol or lidocaine. The success rate of the DMTS tasks in which the monkey had to discriminate difference of surface curvature or axis orientation with binocular disparity cues became significantly lower after injection. This suggests that the parietal cortex plays a crucial role in the perception of 3D features [6].

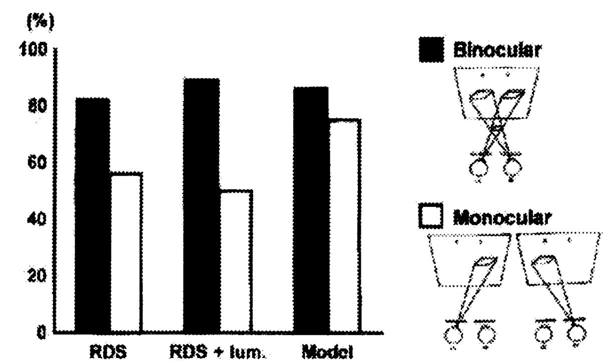

Fig. 8 Comparison of Success Rate between Binocular and Monocular Presentation.

3.5 S. Shimojo (sshimojo@real.c.u-tokyo.ac.jp)

Retinal image motion, which is the critical input to the visual system, consists of non-selective mixture of motions caused by object structure, object motion, and self motion in 3-D space. Human visual system conducts ambiguity solving to decompose/interpret retinal image motion into object structure, object motion, and self motion. We showed that the moving observer perceived different types of object structure/motion depending on the direction of self motion, even though the 2-D stimulus motion was not yoked with self motion, and that its perception depended on types of optic flow (motion field). Specifically, the observer perceived different relative motion when the optic flow contained velocity discontinuity ((a) and (b) in Fig.9). On the other hand, she/he perceived different depth order (depth flipping) when the optic flow consisted of smooth velocity gradient ((d) and (e) in Fig.9) and caused perception of smooth single surface. Though the observer knew the fact that the motion display has constant velocity independently of self motion, she/he experienced 'real perception' of different types of object structure/motion depending on self-motion direction [7].

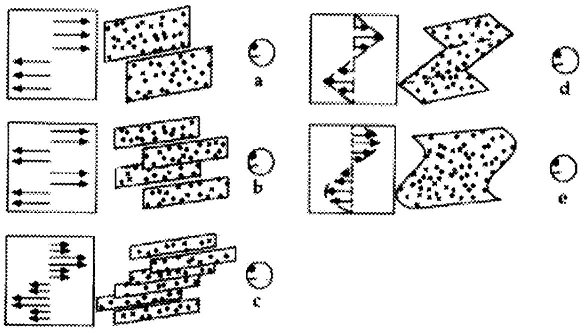

Fig. 9 Different Types of Object Structure/Motion depending on the Direction of Self Motion

3.6 C. Nishimura (nishimuc@med.toho-u.ac.jp)

Audio-visual activities in the human brain was studied relating to the virtual reality by measuring magneto-encephalography (MEG). The state-of-the-art MEG system (whole-cortex type) enables noninvasive measurement of human brain functions with high temporal and spatial resolutions (as short as 1ms and 5mm, respectively). By using the system, we study the following neural processes; 1.Apparent motion perception. 2.Auditory discrimination. 3.Association of vision and audition. 4.Implicit motor imagery [8].

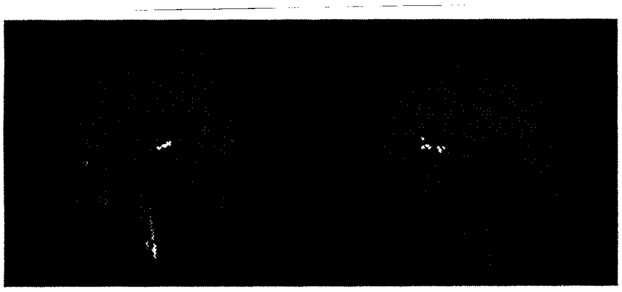

Fig. 10 Equivalent Current Dipoles in the Posterior Hippocampal Region estimated in the 300-400 msec Range during Auditory Discrimination.

At present, we obtained the following results:1.In the apparent motion perception which was produced by two consecutive photic stimuli, a neuronal interference process in the occipital area was observed at 120ms from the onset of the second stimulus.2.In the auditory discrimination task, several dipoles were estimated in the 280-400ms latency range and their activities were prominent in the posterior hippocampal region.3.In the audio-visual association process, MEG analysis for dipoles andP300m activities revealed that the two modalities were separately processed in the 80-130ms latency range, while at 350ms and later the association might be already accomplished. 4.In the implicit motor imagery process, activities in the inferior parietal lobe followed by those in the premotor area (those activities were partially overlapped) were observed as well as those in the visual cotex prior to them.

These results show how human cognition is associated with neuronal processes in the brain, and especially how virtuality is based on the neuronal activity in both temporal and spatial aspects.

3.7 H. Nishijo (nishijo@ms.toyama-mpu.ac.jp)

Figure 11 shows the experimental diagram. Rats were prepared for chronic recording by forming receptacles of dental cement to accept artificial earbars (A). Electrodes were implanted in the ventral tegmental area for ICSS[9].

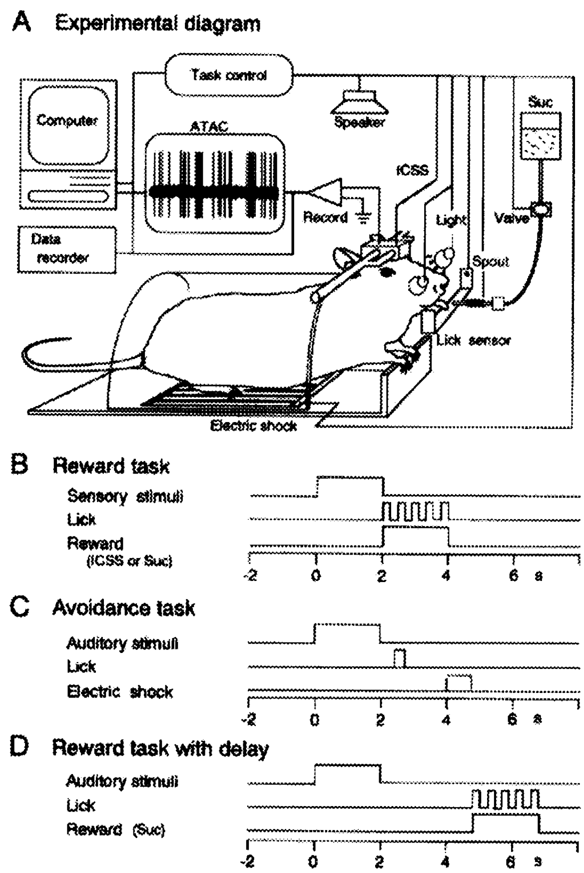

Fig. 11 Experimental Diagram.

The rat was trained to lick when the spout was automatically placed close to its mouth. Licking was signaled by a photoelectric sensor triggered by the tongue. Electric shock was delivered from the grid on a floor of a restraining cage. B-D: Time chart of reward task(B), avoidance task(C), and reward task with delay(D). In the reward associative task, one of conditioned sensory stimuli (tone, light, or configural stimuli) associated with and without rewarding stimulus (ICSS or sucrose solution) was presented for 2.0 s prior to placing the spout close to a rat's mouth(B). In the avoidance task, a rat could avoid electric shock if the spout was licked within 2.0 s after the conditioned auditory stimulus(C). In the reward task with delay, a conditioned auditory stimulus was presented as in the reward task, but a rat had to wait for a delay period (2.4 s) before licking a spout. Suc is 0.3 M sucrose solution.

3.8 Y. Iwamura (iwamura@med.toho-u.ac.jp)

Figure 12 shows the digital images of monkey retrieving food pellets with a rake. When a food pellet was dispensed beyond its reach, the monkey wielded the tool (A) and pulled the food closer to retrieve it with the other hand (B) [10].

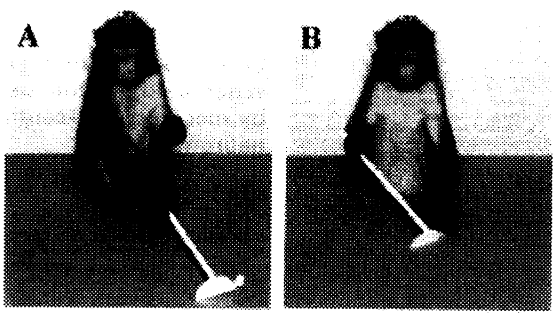

Fig. 12 Monkey retrieving Food Pellets with a Rake

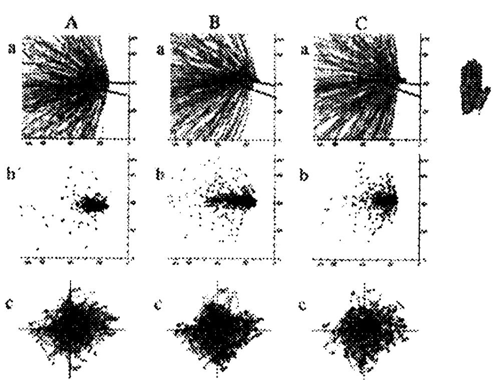

Fig. 13 Visual Receptive Field of a 'Distal Type' Bimodal Neurone.

Fig.13 shows visual receptive field of a 'distal type' bimodal neurone. (A) Before tool use, (B) immediately after tool use, (C) 3 min after retrieving food without the tool, but holding it in the head. (a) Trajectories (shaded lines) of a scanning object projected to a horizontal plane. The scanning object was visible to the monkey; it was moved toward or away from the monkey's hand equally in all directions at a speed of about 0.5 ms-1. (b) The locations of the scanning object in the horizontal plane caused the neurone to fire; each dot represents one spike discharged at an instantaneous frequency > 3.0 Hz (the spontaneous firing level of this neurone). (c) Eye movements (shaded line) and gaze positions at which neuronal discharges occurred (dots). Abscissa and ordinate correspond to horizontal and vertical gaze angle, respectively. Shaded area in the inset depicts tactile receptive field of this neurone.

3.9 H. Tanaka (tanaka@tmd.ac.jp)

Human 3D spatial cognition is the most crucial factor to evaluate VR environment. In order to evaluate this property an eye-movement to clarify a difference between a real environment and a virtual one(VR) was observed. An psycho-physical experiment using two real objects (cube and a pyramid) was conducted. A task for a subject is to determine the kind of object which is presented in front of him. A latency and trajectory of eye-movement until the subject realized the object shape were measured by Eye Mark Recorder(EMR 600 by NAC Co.). Characteristics for the trajectory of eye-movement were as follows:(1) Scanning areas were not spread uniformly over the object. (2) Trajectories of eye-movement were rather concentrated on specific areas of the object such as edges and corners [11].

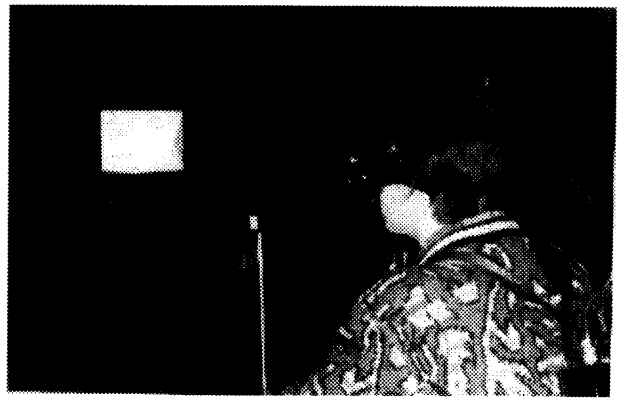

Fig. 14 General View of the Experiment.

3.10 K. Matsunaga (matsnaga@is.kyushu-u.ac.jp)

A new stereoscopic video camera and monitor system with high central resolution has been studied. On the teleoperator system, a human operator needs three dimensional visual information of the working site of the slave robot to manipulate it. It is common to use the stereoscopic video camera and display system to create the images with binocular disparity in order for the human operator to get depth cues to synthesize three dimensional visual images of the objects and the environment. The standard video system or even the high definition video system does not give sufficient spatial resolution for the fine control on the teleoperation when rather wide angle lenses are used on the video cameras for large spaces. The resolution of the human visual system is high at the central area around the fovea and low at the peripheral area. The newly developed stereoscopic video system (the Q stereoscopic video system), which has higher resolution at the central area, was developed using four video cameras and four video monitors. The stereoscopic video camera system (the Q stereoscopic video system) is constructed with two cameras with narrow-angle lenses which are combined as the stereoscopic camera system, and two cameras with wide-angle lenses which are combined with the narrow-angle camera to have the same optical center axis using half mirrors. The Q stereoscopic video monitor system is composed of four monitors, in which two sets of the monitors are larger than the other two sets of monitors, are combined as a stereoscopic monitor system to display the images from the Q stereoscopic video camera system. With the Q stereoscopic video monitor system, the image of the working site from the narrow-angle stereoscopic video cameras are inserted in the images from the wide-angle video cameras of the Q video camera system using three mirrors respectively. With this system, human operators are able to see the stereoscopic images on a large size monitor and the image on the small-size monitor inserted on the larger size image. When the field of view of the large size monitor has a horizontal angle of 40 degrees and the small size display has a horizontal angle of 6 degrees, the spatial resolution is around 40 times higher for the same area on the small size video monitor system than the larger size video monitor systems of the same video standard, for example the NTSC format [2].

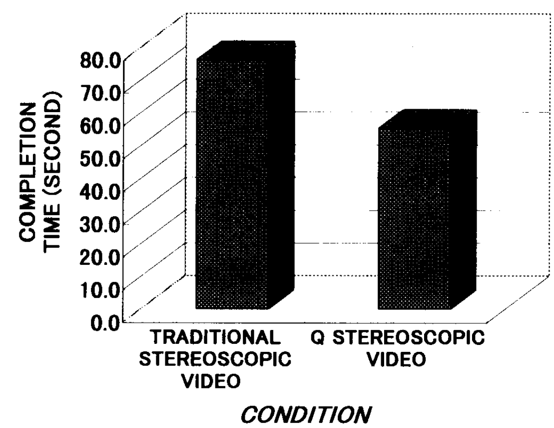

Fig. 15 Comparison Result.

The Q stereo video system has a wide view angle and a central high resolution. Therefore, it is expected that operators will be able to get spatial orientation of the objects in the working space by the large size monitor and fine depth information by the small size monitor. Completion times of the pick and place task of the remote controlled slave robot under the stereoscopic video system with four cameras and four monitors are shorter than that of the conventional stereo video system.

Figure 15 shows completion times under the traditional stereoscopic video system and the Q stereoscopic video system. Completion time under the Q stereoscopic video system is shorter than that under the traditional stereoscopic video system.

3.11 K. Matsunami (kmm@cc.gifu-u.ac.jp)

Gravitational effects on vestibulo-autonomic system during parabolic flight were stuidied. Figure 16 shows one of the results of cardiovascular changes obtained during hypergravity during parabolic flight. Blood pressure and heart rate increased, while pulse pressure slightly decreased. Accordingly, CV generally decreased. However, CV slightly increased at the beginning, and at the end of +Gz hypergravity. In the later stage of parabolic flight experiments, HMD was used to induce optokinetic nystagmus [12].

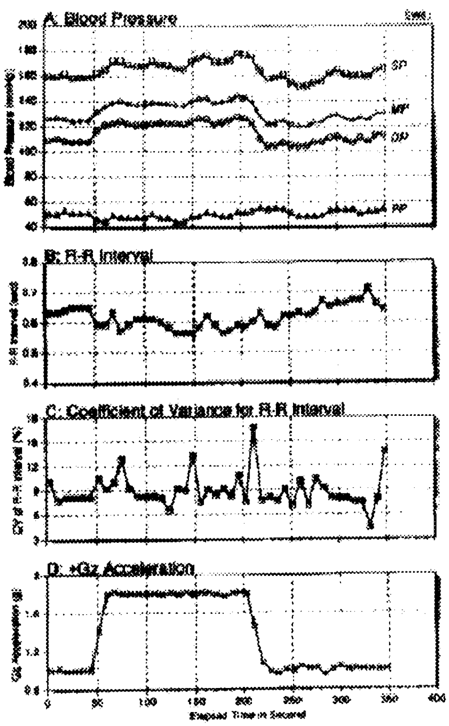

Fig. 16 An example of cardiovascular responses. A value of each time point is the mean values for 8 sec: A; blood pressure, B; R-R interval, C; coefficient variation of R-R interval, and D; exposed +Gz force. Systolic (SBP), diastolic (DBP) and mean blood pressure (MBP) increased, while pulse pressure (PP) slightly decreased. An interval of R-R wave in ECG shortens as shown in B. CV revealed no remarkable change.

3.12 T. Sato (lsato@hongo.ecc.u-tokyo.ac.jp)

Transient and sustained mechanisms underlying stereoscopic depth perception have been studied. Stereoscopic depth can be perceived from temporary asynchronous stereograms if stimulus-onset asynchrony (SOA) is shorter than about 100ms. This phenomenon has been attributed to an overlapping of transient responses at the level of disparity evaluation (Fig.17A). This hypothesis was examined by using two types of asynchronous random-dot stereograms (RDS). Subjects identified near/far depth of a central target within RDS. In Experiment1, one of the stereo pair (reference) was presented for 240ms while the other (probe) was shown for 10ms at various temporal positions overlapping the reference (Fig.17B). In Experiment2, the two patterns had equal duration (50 to 150ms) and were presented successively with no inter-stimulus interval (ISI) (i.e. duration = SOA). Several contrast levels were used for both experiments. The results from Experiment1 revealed transient contrast responses at lower contrasts (Fig.17C). Detection decayed very quickly and reached chance level around 50ms SOA. But at high contrasts, detection was almost perfect up to 200ms SOA whenever there was temporal overlap between the two stimuli. This implies either the involvement of a sustained mechanism or a large transient response at higher contrasts. Results from Experiment2 supported the latter hypothesis. Detection declined around 80ms SOA regardless of contrast level, suggesting that transient responses have a limited duration even at higher contrasts. These results indicate that the human stereo system involves at least two separate contrast mechanisms -- one marked by transient characteristics and higher contrast sensitivity, the other by sustained characteristics and lower contrast sensitivity. The sustained mechanisms seems to operate only when there is temporal overlap between two inputs [13].

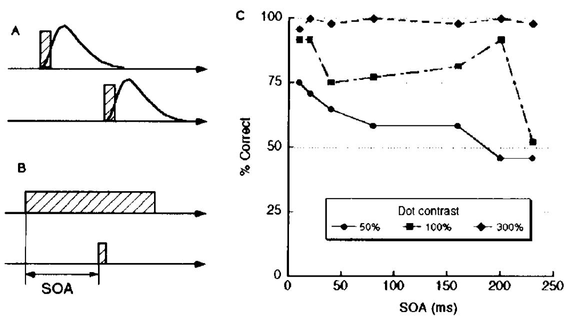

Fig. 17 A. Arrows indicate time. Shaded rectangles indicate physical stimulus to left (top) and right eyes, and solid lines indicates hypothesized visual system's responses to the stimulus. B. The stimulus used in Experiment 1. Arrows indicate time, and shaded rectangles indicate physical stimulus to left (top) and right eyes. C. The results of Experiment 1. At lower contrasts, the correct detection decayed quickly as SOA increased.

4. Conclusion

Interim results of Group 1 of the Priority Scientific Research Are of "Fundamental Study on Virtual Reality -Generation of Virtual Space and Human Interfaces for Virtual Environments-" were reported. The first group aims at the elucidation of perceptual and behavioral human characteristics in virtual environments. This is just a first step toward the long range target of the elucidation of a sensation of presence. The group consists of 12 principal investigators with about 50 researchers of various disciplines, i.e., measurement & control engineering, experimental psychology, psycho-physics, physiological engineering , electro-physiology and neuro-science. Further results will be presented on the web [2].

References

1. S.Tachi and K.Yasuda: "Evaluation experiments of a teleexistence manipulation system." Presence, vol.3, no.1, pp.35-44 (1994).
2. http://www.star.rcast.u-tokyo.ac.jp/VRjuten/research.html
3. S.Tachi, M.Maeda, R.Hirata and H.Hoshino: "A construction method of virtual haptic space," Proceedings of the 4th International Conference on Artificial Reality and Tele-existence (ICAT'94), pp.131-138, Tokyo, Japan, July (1994).
4. Y.Tanaka and A.Mikami: "The integration of visual form and motion information in the primate superior temporal sulcus," Neurosci.Res. Sup. 20 :S202 (1996).
5. M.Ohmi: "Egocentric perception through interaction among many sensory systems," Cognitive Brain Research, vol.5, pp.87-96 (1996).
6. H.Sakata, M.Taira, M.Kusunoki, A.Murata and Y.Tanaka: "The parietal association cortex in depth perception and visual control of hand action,". TINS vol.20, pp.350-357 (1997).
7. M.Kitazaki and S.Shimojo: "The 'generic-view principle' for three-dimensional-motion perception: optics and inverse optics of a single straight bar," Perception, vol.25, pp.797-814 (1996).
8. Y.Kikuchi, H.Endo, S.Yoshizawa, K.Kita, C.Nishimura, M.Tanaka, T.Kumagai and T.Takeda: "Human cortico-hippocampal activity related to auditory discrimination revealed by neuromagnetic field," Neuro Report, Vol.8, No.7, pp.1657-1661 (1997).
9. T.Oyoshi, H.Nishijo, T.Asakura, Y.Takamura and T.Ono: "Emotional and behavioral correlates of mediodorsal thalamic neurons during associative learning in rats, ".J. Neurosci., vol.16, pp.5812-5829 (1996).
10. A.Iriki, M.Tanaka and Y.Iwamura: "Coding of modified body schema during tool use by macaque postcentral neurones," Cognitive Neuroscience & Neuropsychology (NeuroReport), vol.7, no.14, pp.2325-2330 (1996).
11. H.Tanaka, H.Nakamura, T.Nariai and K.Hirakawa : "Brain surgery simulation system using VR technique and improvement of presence," MMVR97 (1997).
12. H.Satake, T.Konishi,. and K.Matsunami: "Effects of tilting upon the cardiovascular responses and opto-kinetic nystagmus during parabolic flight," ISTS Abstract, pp. 1-3-1-1-3-5 (1996).
13. T.Sato.and J.Yanagi: "Transient and sustained mechanisms underlying stereoscopic depth perception," Investigative Ophthalmology & Visual Science, vol.38(4),pp.954-954 (1997).