SPATIAL WORKSPACE COLLABORATION: A SHAREDVIEW VIDEO SUPPORT SYSTEM FOR REMOTE COLLABORATION CAPABILITY

Hideaki Kuzuoka

Department of Mechano-Informatics, University of Tokyo,
7-3-1 Hongo, Bunkyo-ku
Tokyo 113, Japan
+81-3-3812-2111 (ext 6369), kuzuoka%ihl.t.u-tokyo.ac.jp@relay.cs.net

ABSTRACT
Collaboration in three-dimensional space: "spatial workspace collaboration" is introduced and an approach supporting its use via a video mediated communication system is described. Verbal expression analysis is primarily focused on. Based on experiment results, movability of a focal point, sharing focal points, movability of a shared workspace, and the ability to confirm viewing intentions and movements were determined to be system requirements necessary to support spatial workspace collaboration. A newly developed SharedView system having the capability to support spatial workspace collaboration is also introduced, tested, and some experimental results described.

KEYWORDS: Remote collaboration, CSCW, spatial workspace collaboration, focal point, verbal analysis, video mediated communication.

INTRODUCTION
Most of the current studies on Computer Supported Cooperative Work (CSCW) have considered desk-top or office-work type collaboration methods such as conferencing systems [9, 11] and shared drawings or writing tools [2, 10, 13]. However, as communication networks develop, it is becoming possible to collaborate among geographically distributed laboratories or within an integrated industrial framework (e.g., among design sections, manufacturing sections, etc.). In the present approach, collaboration between a machine designer and a manufacturer could occur, for example, in situations where the manufacturer needs to show the designer how a machine operates in order to explain its manufacturability. In this case, different types of communication problems arise since the workspace consists of a three-dimensional (3-D) space. In such an environment, objects which must be explained may be dispersed in 3-D space, as well as having 3-D motions and directions. Here, collaboration in this kind of environment is classified as **"spatial workspace collaboration"** so as to distinguish it from other types of workspace collaboration.

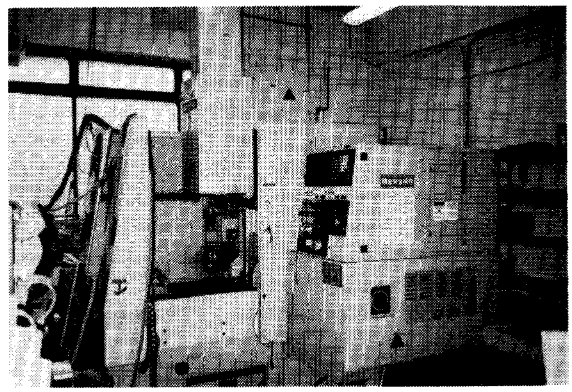

Fig. 1: Machining Center (MC).

OBSERVATION OF MACHINERY OPERATION INSTRUCTION
To clarify spatial workspace collaboration activities, face-to-face instruction sessions on the Machining Center (MC) (Fig. 1) were examined. The MC is a numerically controlled machine that performs complex machining functions such as milling, drilling, etc., and can be either manually operated by switches or automatically operated by a computer.

During an instruction session, an instructor taught a subject (operator) to manually operate the machine to cut a work-piece. All subjects in the present study were either graduates or senior undergraduates in the Department of Mechanical Engineering, University of

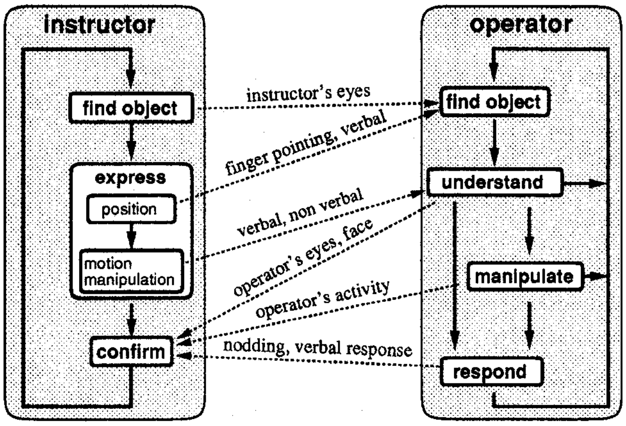

Fig. 2: Communication patterns.

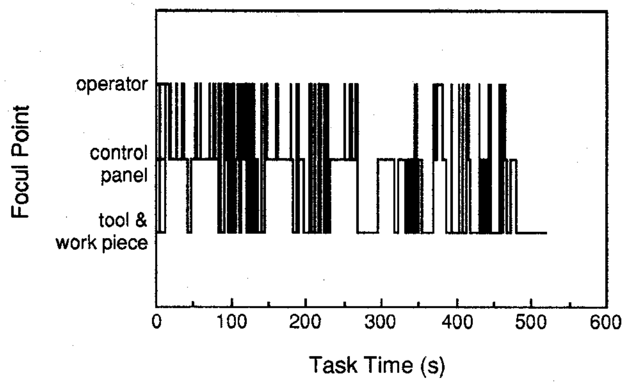

Fig. 3: Changes in the instructor's focal point.

Tokyo. Instruction sessions were video taped for later analysis. It is well-known that both verbal and gestural expressions play an important role in communications [1, 3, 12]; hence the times and locations of these expressions were recorded when used. Since many expressions were also used to direct attention to objects in the spatial workspace, the locations at which the subjects were looking were also recorded. The study's primary objective is to support communication in a 3-D environment, thus analysis of 3-D activities and expressions were subsequently focused on.

Communication Patterns

Following an analysis of the video taped sessions, it was found that structures of communication can be in represented as shown in Fig. 2, with the instructor's actions being categorized as follows:

find object– Directing attention to the object to be explained.

express– Expressing an idea using verbal/gestural means.

confirm– Confirmation that the operator understood the instructor's expressions, i.e. by confirming the viewed object by the operator, or by observation of proper MC operation.

Corresponding to the instructor's actions, the operator acted as follows:

find object– Found the object indicated by the instructor.

understand– Watched and listened to the instructor's expressions so as to understand them.

manipulate– Operated the machine according to the instructor's explanation.

respond– A response by the operator that indicated comprehension; either by affirmatively nodding or saying "OK" for example

In Figure 2, the broken-arrow lines show the direction of the flow of expressions corresponding to each subject's action. It is realized that some actions cannot be included in these categories, however, since only the characteristic activities for a spatial workspace are considered here, other actions are neglected. In addition, it is known that communication patterns cannot be exactly represented as shown in Fig. 2, although this representation was helpful to assist in developing a basic understanding that was useful in designing the presented communication support system.

Changes in Focal Point

The change of focal points in communication patterns is repeated almost every few seconds; and in every "loop" pattern the instructor looks at the objects of attention, the operator's manipulation and facial response, and the MC's motion. Figure 3 shows a typical example of how frequently an instructor changes his focal point, where the y-axis indicates the object at which the instructor is referring to. It was determined that in some instances the instructor changed his focal point almost twice in a second. Hence, to support spatial workspace collaboration, a system must have movable 3-D focal points. Since the system should be able to show locations that the instructor/operator want to see/show, it is also important that the system's focal points accommodate their viewing intention. This is one of the biggest differences of the presented collaboration system as compared to desk-top ones.

Expressions for Spatial Workspace Collaboration

The instructor's expressions that were used in the previously described focal point communication loop were classified as *position*, *motion/manipulation*, and *confirmation*. Many other expressions, e.g., counting numbers and showing the length of an object, were also used, but were neglected since only 3-D expressions were analyzed here.

position: The operator sees the object which the instructor directed attention towards. Verbal expressions

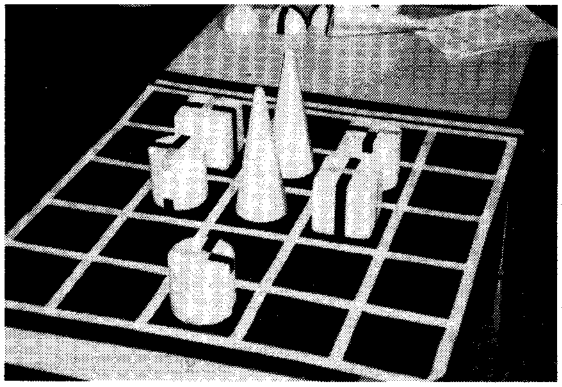

Fig. 4: Typical model task workspace.

such as "this button" and gestures, such as finger pointing were used. These expressions represent the **express** action which occurs immediately following the instructor's **find object** action.

motion/manipulation: Expressions for describing the motion of a tool, the manner in which to push buttons, the way to turn switches, etc., being verbally expressed as "this direction," "turn like this," etc. These expressions were usually given soon after the *position* expressions during the **express** action.

confirmation: Expressions which confirmed that the operator understood what was being explained, or by motions indicating they were looking at the right location. These expressions were seen during the **confirm** action.

In order for the instructor to give instructions smoothly, a communication system should incorporate features which support the ability to smoothly exchange these expressions.

MODEL TASK EXPERIMENT

Model Task

Use of the MC for experiments requires significant preparation time and not many qualified instructors were available, thus a so-called "model task" was created to enable experiments to be easily and routinely conducted. The task had to include the expressions of *position*, *motion/manipulation*, and *confirmation*, in addition to having 3-D object movement capability. The task selected used an area divided into 5×5 squares similar to a chess board, with four different 3-D objects being utilized. Initially, several of the objects were placed onto some squares (Fig. 4) and during the task the instructor directed the operator to move objects to another square or to rotate them in some direction. Three types of activities were performed and each one consisted of a specific number of required manipulations (A, 5; B, 6; C, 9).

Experiment

The following four cases of instruction format were evaluated.

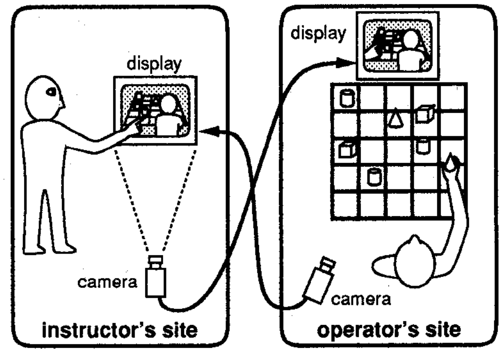

Fig. 5: Case 3 of the model task experiment.

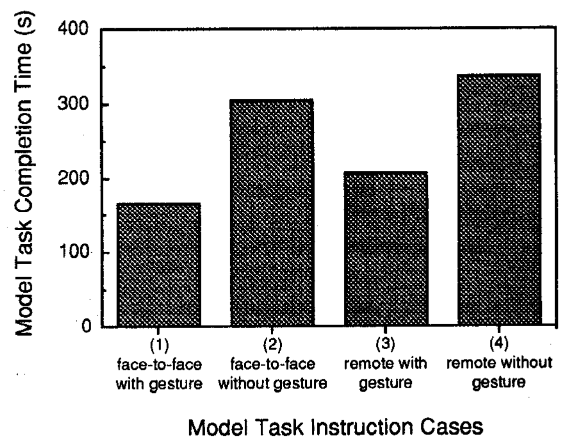

Fig. 6: Model task completion time for each case.

1. **case 1 (face-to-face, with gesture)** Instruction was given face-to-face and the instructor could use any gestures or words, but could not move objects themselves (10 experiments).
2. **case 2 (face-to-face, without gesture)** Instruction was given face-to-face and the instructor could use only words and no gestures (7 experiments).
3. **case 3 (remote, with gesture)** Instruction was given remotely using a VideoDraw-like configuration (Fig. 5) [13, 4]. The instructor's gestures were superimposed on the image received from the operator's site and then sent back to the operator's site (11 experiments).
4. **case 4 (remote, without gesture)** Instruction was given remotely as in case 3, but the instructor was restricted from using gestures (7 experiments).

Expression Analysis Evaluations

Figure 6 shows the completion times of the model task for each case, with gestures clearly increasing the communication efficiency. However, it should be noted that the required time does not necessarily indicate communication variations between different communication cases. When people communicate with each other, more than one sensory channel [8] and expression channel are complementarily used, i.e., if one channel is modified, or re-

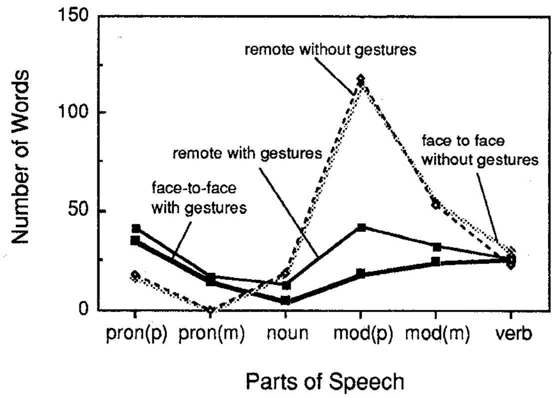

Fig. 7: Parts of speech used in Model Tasks.

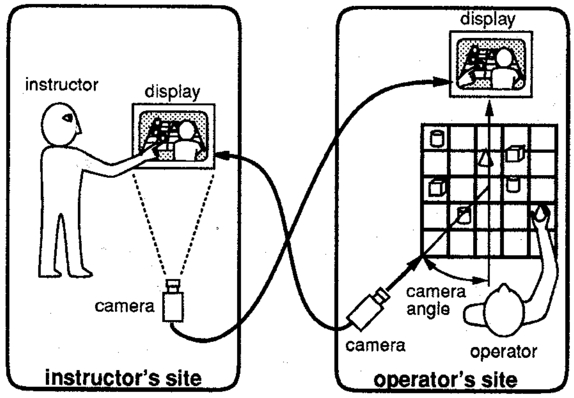

Fig. 8: Experiment using different camera angles.

stricted in usage, our communication strategy is quickly altered to compensate for the lack of information via the use of other channels. As a result, it is beneficial to analyze changes in expressions to understand the effect of modifications in a communication system. Although standard deviation was calculated for the results, their value is not significant enough to be discussed since the number of experiments was limited.

Verbal Analysis

Figure 7 shows the average number of words used in each case for a particular grammatical part of speech. The definitions applicable to the x-axis are respectively "pron" representing pronouns and "mod" representing modifiers, where modifiers include adjectives, adverbs, and spatial words. Spatial words are expressions to describe spatial position and direction, e.g., "up", "down", "right", "left", "inside", "far", and "center" [14]. "Noun" includes nouns which are not considered as modifiers such as "button", "cone", and "cube". Modifiers were classified into two types according to the expressions they were used for, i.e., modifiers to express *position* (p), or modifiers to express *motion/manipulation* (m) expressions (mainly rotation directions). Since nouns were seldom used for expressing motion, "noun" represents words used to express an object's name. Since interest was directed only at words that were used to express *position* and *motion/manipulation*, other words were not counted. Furthermore, since the increase and decrease in the number of words were the primary objective, normalized word counts per part of speech were not utilized.

In Figure 7, the two communication cases that did not use gestures showed almost the same results, pattern, while the other two cases that used gestures are markedly different. The difference in gestures vs. no gestures indicates that the gestural expressions significantly decreased the required number of verbal expressions, especially declarative expressions such as modifiers. For example, when gestures were used, the operator said, "Place this cube here, and turn 90 degrees like this," whereas when gestures were restricted, the operator said, "Place the cube to the right one step and down one step. Then, horizontally and clockwise, turn it 90 degrees." It is also interesting to note that the use of pronouns decreased considerably without gestures. Although this verbal analysis may require more thorough examination, it is safe to assume that oral channels can be used equally for each case, and that the number of declarative words increase and the number of pronouns decrease when communication is not smooth. Since the number of subjects was limited, the statistical significance cannot be strongly argued here. However, using time results and verbal analysis together, the effectiveness of gestures is clearly evident.

Translation of Directional Expression

Experiment observations indicated that a difference in the directional orientation between the workspace of the instructor and the operator was one of the primary causes in making communication difficult. The effect of different workspace viewing angles was examined by changing the camera angle of the workspace (Fig. 8). Six pairs of subjects were divided into two subclasses, with each subclass subsequently performed activities A, B, and C under certain prescribed circumstances. One subclass performed activities A and C with a camera angle of 45° and activity B at 90°. The other subclass performed A and B at 90° and C at 45°. The data obtained from a 45° angle was classified as category (I), whereas the data using 90° was classified as category (II). This enabled a comparison for the effects of camera angle on performance.

Figures 9 and 10 show that as the camera angle of workspace increases, communication becomes more difficult, i.e., orientational differences vary the visual cognition between two subjects. When the instructor expressed position or rotation instructions to the operator, directional expressions had to be translated so that the operator could understand them. For example, the following conversation was recorded:

Instructor: "180 degrees to the right. This direction."

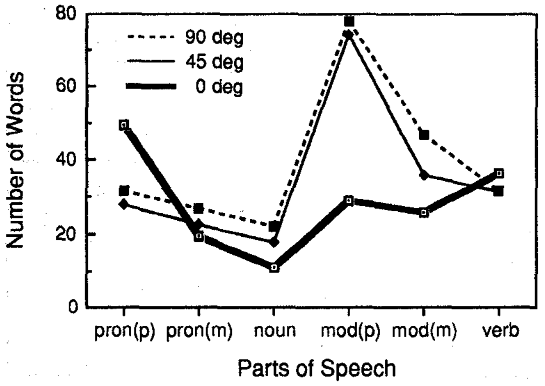

Fig. 9: Relationship between parts of speech and number of words at different camera angles.

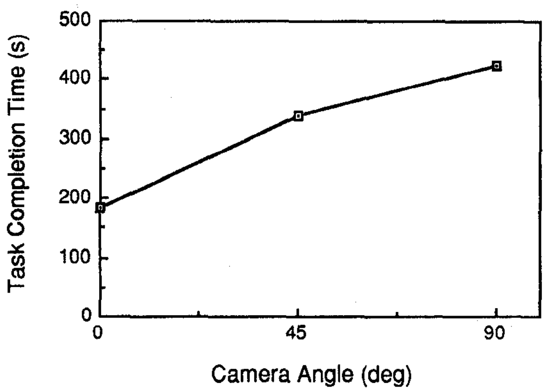

Fig. 10: Relation of task completion time and camera angle.

Operator: "That's left, for me."
Instructor: "180 degrees to the right."
Operator: "Like this?"
Instructor: "The other way, the other way."

Another problem which occurred involved the instructor's superimposed gestures becoming more ambiguous for the operator to follow since the screen images were viewed from a different perspective than their own. This caused an increase in the number of total words (Fig. 9) and task completion time (Fig. 10). If the workspace is in 3-D, and the camera location is fixed, directional expression translation frequently occurred; thus the ability to change the focal point is required in this situation.

Communication System Requirements

Based on the results of these experiments, the following communication system requirements are required to effectively support spatial workspace collaboration.

- Variability of focal point to optimally accommodate viewing intentions.
- Ability to share a focal point; thereby minimizing differences in directional expressions.

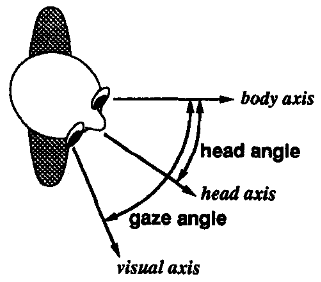

Fig. 11: Diagram showing body, head, and visual axes.

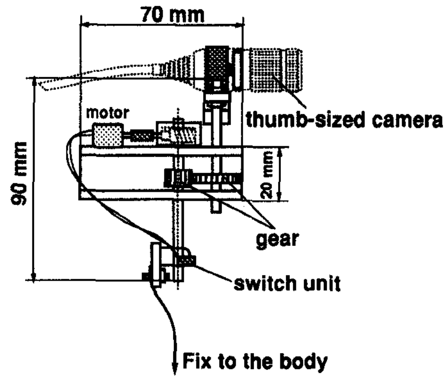

Fig. 12: The SharedCamera's operating mechanism.

- Capability to use superimposed gestures.
- Since the focal point should be variable, the operator's display showing applicable instructions (shared workspace [13]) should also be variable.
- Possess the ability to confirm an operator's comprehension and the object's actual manipulation.

SHAREDVIEW: VIDEO COMMUNICATION SUPPORT SYSTEM

A system, named SharedView, was developed to satisfy the requirements to support spatial workspace collaboration, and consists of the following two devices:

SharedCamera

In order for the instructor and the operator to both share the same focal points and to increase the ability to change them, a small camera about the size of a person's thumb was mounted on the head of the operator. However, the head's axis does not always correspond to the eye's visual axis, e.g., when looking sideways a person's head turns although not as much as their eyes (Fig. 11); thus if the camera is rigidly fixed to the head it will not always be directed at the proper location. This necessitates turning the camera so that it faces in the right direction. It was noticed in the experiments that the gaze angle (angle between body axis and visual axis) and head angle (angle between body axis and head axis) were almost proportional, therefore, if the head angle could be

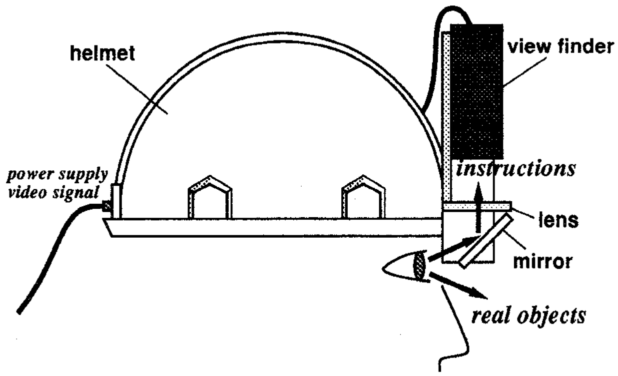

Fig. 13: Head Mounted Display (HMD).

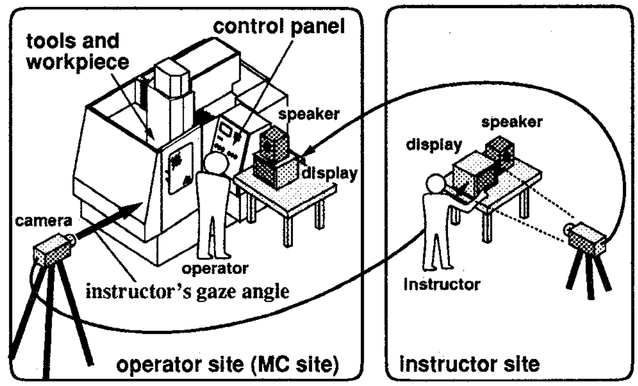

Fig. 14: Case 2 of MC instruction.

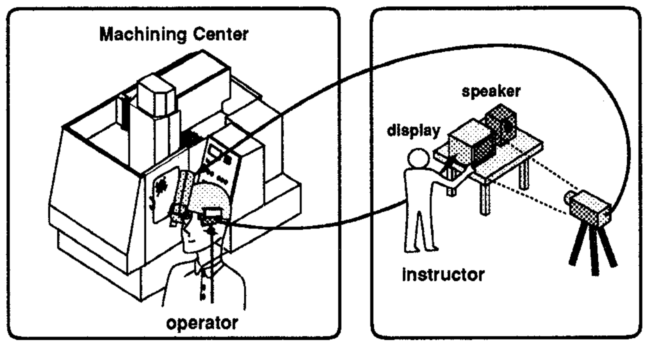

Fig. 15: Case 3 of MC instruction.

detected then the gaze angle could be estimated. For this purpose, the angle occurring between the head and body was detected using a wire with one end fixed to the body and the other end leading to the SharedCamera's switch unit (Fig. 12). The switch unit was designed to turn on when the head turned, and to turn off when the camera turns a particular number of degrees. By utilizing this mechanism the camera was directed at almost the same location in which the subject was actually looking. One concern was that the resultant image from the Shared-Camera might be shaky and difficult to see. However, since a person's head does not move so frequently, head "jiggle" was a negligible effect.

Head Mounted Display (HMD)

The initial experiments showed that superimposing gestures on a shared workspace image was an effective method for instruction. Since the operator's focal point changes frequently, the display is required to be movable. If the operator had to instead look at a stationary display in order to observe the instruction, the Shared-Camera would not be directed towards a specific location in the shared workspace. For this reason, a small display monitor was mounted on the head so that the operator's head need not be turned, i.e., allowing the operator's pupils to move upwards to view the display in order to see the remotely given instruction (Fig. 13). On the other hand, when the operator wanted to see real objects, the pupils moved downwards. The displayed image was approximately equivalent to viewing an 8-inch display.

This head mounted camera was formerly called "Shared-View" [7], although here it is renamed as "SharedCamera" and the integrated SharedCamera and HMD setup called "SharedView."

SYSTEM EXPERIMENTAL USE

The SharedView system was used to perform remote instruction of the MC. To enable comparison the same task was examined using the following three cases:

1. **case 1 (face-to-face, with gesture)** Instruction was given face-to-face using gestures and words, although no actual control of the MC was allowed (2 experiments).
2. **case 2 (remote, fixed camera and display)** Instruction was remotely given with gestures using a camera and display that were set at a fixed location as shown in Fig. 14 (9 experiments).
3. **case 3 (remote, SharedView)** Instruction was remotely given with gestures using the SharedView as shown in Figs. 15 and 16 (9 experiments).

The MC instructions were given in such a way that the operator only needed to change directions toward the MC by approximately 90°, i.e., to face the control panel, the tools, or a work-piece. For expression analysis, expressions were classified into three types as:

position: Expressions indicating the locations of buttons, switches, etc.

manipulation: Expressions indicating the manner in which switches are turned, buttons are pushed, etc. The operator primarily looked toward the control panel when these expressions were used.

direction: When the instructor was required to spatially describe the MC, expressions indicating 3-D orientation (e.g., direction of the tool's x, y, z axis) and tool movements had to be used (Fig. 17). At this

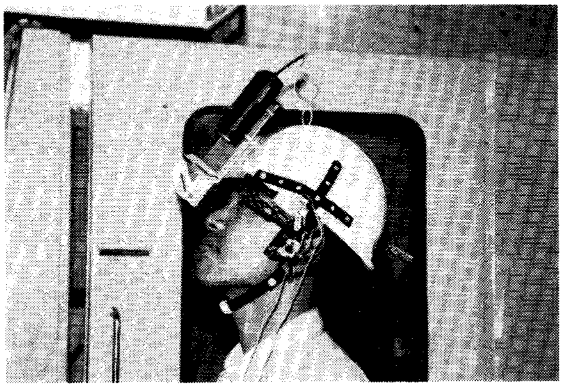

Fig. 16: Wearing SharedView for remote MC instruction.

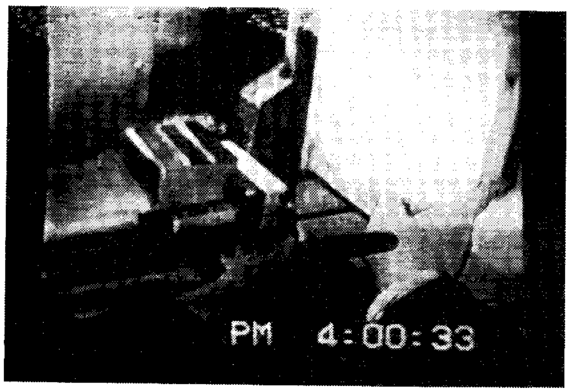

Fig. 17: Instruction location as displayed on the HMD. Both the instructor's and operator's gestures can be seen.

time the operator looked toward the actual direction of the tools or a work-piece.

In the model task expression classifications, a *direction* expression was included in the *motion/manipulation* expressions, although in this experiment it was separated to more clearly show the effect of the SharedView. In Figure 18 the definitions applicable to the x-axis are respectively "p" representing pronouns, "n" representing nouns, "mod" representing modifiers, and "v" representing verbs. Each of them were classified into three types according to the expressions they were used for, i.e., *position* (p), *manipulation* (m), or *direction* (d). Only two instructors (A, B) were available; hence A gave instruction to 5 operators while B gave instructions to 4 operators (case 2 and 3). In case 1, A gave instruction to both subjects. Since the experiment's objective was to compare case 2 with 3, only two case 1 experiments were performed. Results indicated that when using SharedView, modifiers for *direction* decreased (Fig. 18), as did the time required to explain the MC's spatial orientation (Fig. 19). This occurred because in case 2, the direction

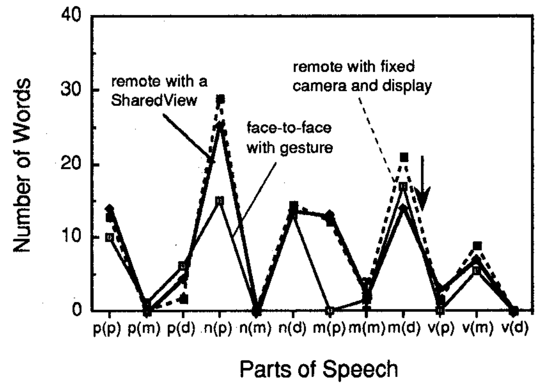

Fig. 18: Effect of SharedView and HMD for remote MC instruction.

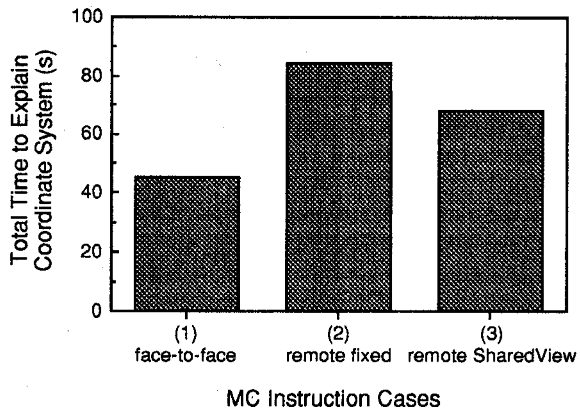

Fig. 19: Elapsed time to explain the MC's coordinate system.

of the viewing angles between the instructor and the operator differed by $\sim 90°$ while explaining the MC's internal spatial orientation. A typical example of an actual misunderstanding of a direction expression observed in case 2 is as follows:

Instructor: "Left to right. This direction."
Operator: "Like this? Like this?" (He was showing forward and back directions.)
Instructor: "No, opposite. Right to left."
Operator: "Right to left?"
Instructor: "Yes."

These results showed the effect of SharedView's ability for focal points, although further experiments are required to determine statistical significance.

CONCLUSIONS

In this paper, a new collaboration type, i.e., "spatial workspace collaboration" was introduced. As an example of spatial workspace collaboration using video media, remote instruction experiments were undertaken to clarify actual visual interactions. For evaluation of communicationability, verbal expression analysis was utilized as well as time evaluation. However, it cannot be said that use of only verbal analysis is the best indicator. In fact, it is believed that it is necessary to analyze communication using more than one method in order to evaluate its effectiveness and to clarify complex human communication patterns. The importance of having a movable focal point was focused on here because it has not been investigated in previous systems. Using the presented SharedView system, an instructor can easily show using gestures what they want an operator to see. Also of importance is the need for the instructor to confirm where the operator is looking.

When communication is informal, it occurs at any time, in any place, and for any purpose. To support this kind of communication, movability is considered to be an important feature in a communication system. Here, movability is defined as how easily the system can be setup before actual interactions occur. The movability of focal points in a shared workspace were subsequently investigated, with it being found that these factors affect the overall interaction time efficiency. Therefore, not only the actual interaction stage, but also **setup stage** should be taken into account when a communication system is designed. Although experiments in this paper were remote instructions, the concept introduced here can be applied to collaboration inside-the-office. Heath and Luff addressed problems associated with a static visual system [5]. Generally, objects that people want to see/show may be dispersed anywhere in three-dimensional space. However, people may not always be at a location where a video camera can easily function, nor will they may not always be in a situation where they can conveniently use a video camera. Therefore, it is concluded that for video mediated communication systems to be more widely used, spatial movability of the system must be included.

For good communication system design, movability and an ability to share an image appears to be a common goal. However, these goals should be realized based a qualitative understanding of human activities, i.e., the system should be designed to accommodate the intended tasks of the user. SharedView is an example of such a system.

ACKNOWLEDGMENT

Much gratitude is extended to Dr. Takemochi Ishii, Keio University, and Dr. Michitaka Hirose, University of Tokyo, for all their beneficial advice and suggestions. I am also thankful to Gen Suzuki and Hiroshi Ishii, NTT, for their kind assistance.

REFERENCES

[1] Sara A. Bly, A Use of Drawing Surface in Different Collaboration Settings. In *CSCW88*, Portland, 1988, pp. 250–256.

[2] Sara A. Bly and Scott L. Minneman, Commune: A Shared Drawing Surface. In *SIGOIS Bulletin*, Massachusetts, 1990, pp. 184–192.

[3] Alphonse Chapanis, Interactive Human Communication. In Irene Greif, editor, *Computer-Supported Cooperative Work: A Book of Readings*, Morgan Kaufmann Publishers, 1988, pp. 127–140.

[4] Jack F. Gerrissen and John Daamen, Inclushion of a 'Sharing' Feature in Telecommunication Services. In *13th International Symposium HFT '90 HUMAN FACTORS IN TELECOMMUNICATIONS*, Torino, Italy, 1990.

[5] Christian Heath and Paul Luff, Disembodied Conduct: Communication Through Video in a Multimedia Office Environment. In *Proc. of the Conference on Computer Human Interaction (CHI) '91*, 1991, pp. 99–103.

[6] Hiroshi Ishii and Naomi Miyake, Toward an Open Shared Workspace: Computer and Video Fusion Approach of TeamWorkStation. *Communications of the ACM*, 34(12), 1991, pp. 37–50.

[7] Takemochi Ishii, Michitaka Hirose, Hideaki Kuzuoka, Tsutomu Takahara, and Takeshi Myoi, Collaboration System for Manufacturing System in the 21st Century. In *Proc. of the International Conference on Manufacturing Systems and Environment*, 1990, pp. 295–300.

[8] Robert Kraut, Carmen Egido, and Jolene Galegher, Patterns of Contact and Communication in Scientific Research Collaboration. In *CSCW88*, Portland, 1988, pp. 1–12.

[9] Marilyn Mantei, Capturing the Capture Lab Concepts: A Case Study in the Design of Computer Supported Meeting Environments. In *CSCW88*, Portland, 1988, pp. 257–270.

[10] Susanna Opper, A Groupware Toolbox. *BYTE*, Dec 1988, pp. 275–282.

[11] Mark Stefik et al, Beyond the Chalkboard: Computer Support for Collaboration and Problem Solving in Meetings. In Irene Greif, editor, *Computer-Supported Cooperative Work: A Book of Readings*, Morgan Kaufmann Publishers, 1988, pp. 335–366.

[12] John C. Tang and L. J. Leifer, A Framework for Understanding the Workspace Activity of Design Teams. In *CSCW88*, Portland, 1988, pp. 244–249.

[13] John C. Tang and S. L. Minneman, VideoDraw: A Video Interface for Collaborative Drawing. In *Proc. of the Conference on Computer Human Interaction (CHI) '90*, Seattle, 1990, pp. 313–320.

[14] Atsushi Yamada et al, The Analysis of the Spatial Descriptions in Natural Language and the Reconstruction of the Scene. *Information Processing Society of Japan*, 31(5), 1990, pp. 660–672. (in Japanese).