Behaviors and Communications in Working Support through First Person Vision Communication

Yuichi NAKAMURA, Takahiro KOIZUMI, Kanako OBATA, Kazuaki KONDO
Academic Center for Computing and Media Studies, Kyoto University
Sakyo, Kyoto, Japan
Email: yuichi@media.kyoto-u.ac.jp

Yasuhiko WATANABE
Faculty of Science and Technology
Ryukoku University
Seta, Otsu, Japan

***Abstract*—This paper introduces a novel idea of analyzing logs of "working support through first person vision (FPV) communication", a working style in which a worker with a head-mounted camera works under the guidance of an experienced mentor monitoring the FPV at a distance. The logs contain various types of multimodal interactions concerning intentions, instructions, and explanations as well as usual working information. We first propose frequent pattern extraction and investigation of the resultant characteristics, and subsequently show experimentally that some characteristics are tightly linked to the smoothness of the work environment and failures.**

I. Introduction

Figure 1 shows a brief overview of working support through first person vision (FPV). A "worker" (the acting person) wears a camera that captures FPV, a microphone, and other sensors such as location sensors. Captured FPV and sensor data are sent to a "helper" (the assisting person) who monitors the work. The worker and the helper can converse. For example, the monitoring person can teach, guide, explain, or query something, and the acting person can report details, and ask questions. FPV, voice, and other sensory data are recorded as a working log. A worker can use both hands free from holding a camera and aiming at a target while FPV is transmitted to a helper. One important advantage of this framework is that the helper can see what the worker see referring to the worker's attention, which cannot be realized with fixed surveillance cameras. This working support method is widely applicable in situations such as on-the-job training, factory work, and support for people suffering from cognitive impairment.

Practical applications and their usefulness have been reported. Remote skill acquisition in surgery and telepresence in accident emergency rescue [1] are promising applications. We can, in addition, easily foresee applications wherein a single or multiple individuals perform critical tasks in dangerous work places with the assistance of experts. A wearable device that with a camera mounted on a helmet and associated mobile phone communication is commercially available [2].

Analysis of communications that in this style of working support is useful both for practical applications and for human communication analyses: how can people maintain good communications in this working style and how is it possible to prevent possible errors or accidents in the work place. Examining and evaluating the quality of communications between a worker and a helper can guide the users on using the system safely and effectively.

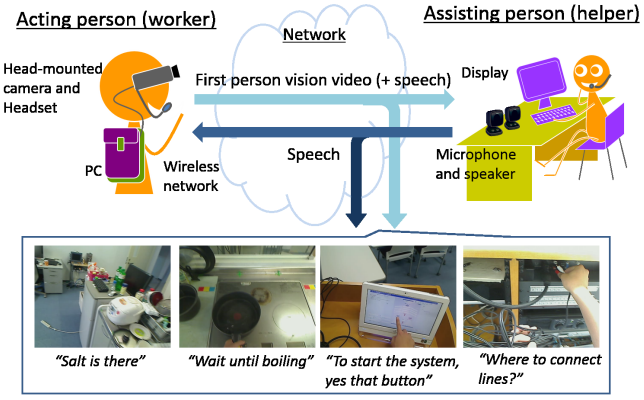


Fig. 1. Overview of working log with FPV communication

For this purpose, we performed quantitative analysis of FPV communications. Experimental results showed that certain characteristics of frequent patterns are tightly related user behaviors and possible failures.

The rest of this paper is organized as follows. Previous work related to working support through FPV communication are presented in Section II. The problem is described in Section III, and the framework and brief characteristics of FPV communications are presented in Section IV. The features and their intermodal cohesiveness are discussed in Section V. In Section VI and VII, experimental results are presented along with the statistics of features and cohesiveness; the performance assessment of the system in detecting important situations is also presented.

II. Related Works

Various possibilities of mobile video communication have been explored from the early days of wearable computer devices [3] [4] [5] [6] [7] [8]. Kraut et al. performed a case study of bicycle maintenance assistance using video communication [9], reporting efficiency and behavior changes due to the existence of a supporting expert and video communication. Efficiency was doubled by expert supports, whereas video communication did not significantly affect the work efficiency and accuracy. Behavior of the workers and the experts, however, significantly changed by the addition of video communication, with typically drastic increases in proactive behaviors of experts.

Fussel et al. reported a similar experiment, but they added a side-by-side condition in which an expert remains beside a worker [10]. They observed significant differences in using visual information: for the side-by-side case and video communication case, deictic expressions often appear that make conversation simpler, though the worker and helper had differences in this tendency. Efficiency was significantly better in the side-by-side case, with no significant differences regardless of video communication or the helper's experiences.

Billinghurst investigated the effect of using asymmetric devices between a worker and a helper [11]. The subjective impression of easiness for collaboration using audio only, video conferencing, and augmented reality (AR) with a head mounted display (HMD) were analyzed both in symmetric and asymmetric cases. Results suggest that asymmetric devices for a worker and a helper is sufficient if their roles are different. Advanced devices such as AR with HMD do not always improve performance and evaluation.

Kraut et al. devised the task of a collaborative online jigsaw puzzle, and analyzed the behaviors of a worker and a helper [12]. The shared video space caused differences in acknowledgment of understanding and behaviors as well as deictic expressions. Gergle et al. analyzed behaviors in object reference and placement situations by providing utterance and behavior codes to communication behaviors [13]. They concluded that significance differences with and without video were caused by the principle of least collaborative effort and not by the principle of least effort [14].

Previous works assessed some aspects of working support with mobile video communications and primarily argued advantages and disadvantages of varying styles of working support. In contrast, our research focuses on a more detailed analysis of measuring the communication behaviors and on whether different behaviors cause different results. A crucial aspect is quantitative analysis based on low-level features that have the potential to be detected automatically.

Multimodality is one of the most important aspects in FPV. To analyze intermodal relationships, Norris proposed a useful concept of "modal density" [15] [16], which refers to actions in multiple modalities that cooperatively work to form a high level action. The complexity (modal complexity) and intensity (modal intensity) of these actions are also discussed as important aspects of human communications and behavior. Those notions can be used to explain communication patterns wherein different modalities are mutually related. Norris, however, did not provided quantitative methods for analyzing those communication characteristics.

Considering the above, the works by Kraut et al. [12] [13], were good experiments that typical actions used different patterns in varying communication styles; in addition, they quantified the occurrence probability of certain actions. The features used in those works are at a relatively high level, and analyzing such actions with raw recorded data appears difficult. Therefore, a new method which based on low-level features and their temporal characteristics was required.

III. Problem statement

Based on the above review, this research aims to conduct the following analysis on FPV communication.

(1) Classification and quantification of the communication behavior of workers and helpers
(2) Analyses of the relationship between communication behavior and smoothness of work or failures.

Concerning the former, nonverbal communication behaviors must be considered and analyzed them in conjunction with verbal communication. For the latter, ignorance, misunderstanding, lack of care, and disobedience often cause failures or accidents [17], with insufficient communication clearly increasing this tendency. However, it is not always the case and failures are not frequently observed.

Based on this idea, we employed the following strategy.

(a) Analyze multimodal communication in cases of with and without failure. For this purpose, we extracted frequent patterns and examined their qualitative and quantitative characteristics.
(b) Use low-level features that have potential to be detected automatically; do not use deep semantic analyses.

This strategy potentially enables us to deal with large amounts of data automatically and provides a basis for real-time processing.

IV. Description of communication and behavior

A. Modality and Features

We first need to consider modalities and features. The followings are typical features in FPV communications.

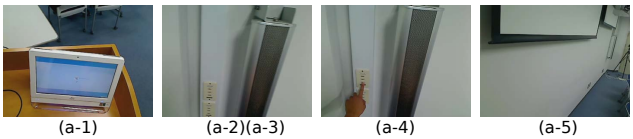
Speech (utterance): exophoric or deixis, name of physical object, appearance/spatial description, request/query/explanation of actions.
Visible actions: hand motions, location movements, gestures.
Observational actions: staring/gazing still, looking around.

B. Typical scenes and interaction patterns

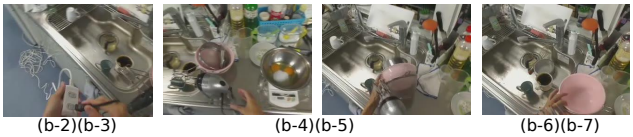
Figure 2 shows a typical situation with working support through FPV communication: a worker is performing unfamiliar tasks with assistance from an expert.

Story (a) in Figure 2 is an ideal case in which the task was performed smoothly and where enough information was exchanged through FPV communications. The first two utterances referred a screen and a switch. The motion of looking around implies the worker acknowledged the task of pull the screen down, and attempting to locate the switch. The third and fourth utterances are directed at the switch with its color and spatial characteristics. Next, (a-5) reported the same event through both speech and video. Thus, the behaviors of "looking around", "pointing", "gazing still" correspond to the utterances of specifying a task, requesting position, and identifying and observing an object, respectively. Almost ideal communications were given in correct order and with the appropriate timing.

Story (b) and story (c) in Figure 2 are cases where problems occur. In story (b), the worker pushed a blender down, one of the causes of which was inappropriate interaction. In (b-2), the



helper (a-1): “Let down the screen.”
worker (a-2): “Where is the switch?” (looking around)
helper (a-3): “The red button on your left.”
worker (a-4): “This one?...” (pointing the button and push it)
worker (a-5): “OK. It’s going down.” (watching the screen)



helper (b-1) “Connect it to the blender”
worker (b-2) “Here?” (gazing at power outlet)
helper (b-3) “Yes”
worker (b-4) “Oops!” (having pushed the blender down)
helper (b-5) “What’s the matter?”
worker (b-6) “Disaster!” (looking at the fallen blender)
helper (b-7) “What happened?”

helper (c-1) “Slice it into two”
worker (c-2) “Hmm, baked in slanting” (gazing from the above)
helper (c-3) “Let’s see... yes”
worker (c-4) “Too badly shaped” (with wrong slicing method)
helper (c-5) “Oh! slice it by turning the cake”

Fig. 2. Typical scenes in working support through FPV communication (Story (a)–(c))

helper was unaware that the worker was not paying attention to the blender, i.e., the worker did not look at the blender nor did he notice it. After the blender fell, an appropriate amount of information was not shared, as shown in (b-5)–(b-7). The worker did not make optimum use of the communication channels. For story (c), advice was not given at an appropriate time because the helper’s attention was heavily focused on the worker’s utterances (c-2) and (c-3).

C. Frequent patterns and cohesiveness

We use term “cohesiveness” to represent the state where multiple features in communication occur adjacently. Cohesiveness is for story (a)–(c). All the features in story (a) are “coherent”, i.e., they are consistent and in the right order, while some features in stories (b) and (c) are incoherent.

To analyze these communications, we make the following assumptions:

(i) If both the worker and the helper share attention on the same target and communication is sufficient, both cohesiveness and coherency is deemed satisfactory. This implies that when a task is smoothly conducted, cohesive and coherent communication patterns frequently appear.
(ii) In case of insufficient communication, which can cause failures, one or more essential elements (features) of the above normal communication pattern is often missing.

Based on these assumptions, we focus on the characteristics of frequent patterns and the derivations obtained by removing or adding elements.

V. Analysis of Frequent Patterns

A. Detection of frequent patterns

We applied PrefixSpan [19], a sequential pattern mining method, to the FPV communication logs and examined obtained frequent patterns.

For actual mining, we used the following settings.

- The starting time of a feature was the time of the feature’s occurrence.
- The starting time of an utterance was used for the occurrence time of features in the utterance.

These settings were based on the consideration that a person already has the intention of an action, i.e., speech or behavior, just prior to that action. We, therefore, employ the starting time as an approximation of the action time. Moreover, the precise time at which a word is spoken is presumably less important than the starting time of the utterance, because the word order heavily depends on the grammar of a language. Thus, we assumed that words in the same utterance had the same occurrence time.

In sequential pattern mining, a “transaction” is a sequence of “items”. We considered each feature as an item and a transaction as a sequence of features, e.g., $S = \{F_1, F_2, ..., (F_i, ..., F_m), ..., F_z\}$, where F_i represents a feature type such as “looking around” and “request (utterance)”. A frequent pattern is also a sequence of items, i.e., a sequence of feature types, that frequently appeared in a transaction. A frequent pattern is represented as $s_j =< F_{j1}, ..., (F_{jp}, ..., F_{jr}), ..., F_{jz} >$.

We can interpret their meanings by referring actual FPV communication logs. For example, $s_1 =< (\mathrm{object\_name}, \mathrm{request}), \mathrm{movement}, \mathrm{explanation} >$ may represent that an object is mentioned and a subtask is requested in an utterance, an acting person moved to another location, and an explanation was given.

B. Temporal characteristics

An extracted frequent pattern maintains the order of the items (features) within. To examine in more detailed the temporal characteristics, we considered co-occurrence and pseudo mutual information.

Co-occurrence

First, let us consider two feature instances $f_i \in F_a$ and $f_j \in F_b$ that have a possibility to satisfy a co-occurrence relationship $R_k(f_i, f_j)$. Its occurrence probability is denoted as follows.

$$P(R_k(F_a, F_b)|F_a) = \frac{N(R_k(f_i, f_j,)}{N(f_i)} \quad \text{s.t.} f_i \in F_a, f_j \in F_b \tag{1}$$

where, $N(g)$ represents the number of occurrences of g.

Next, we deem a feature f_i has a duration $[t_i^s, t_i^e]$ with starting time t_i^s and ending time t_i^e. The modified occurrence relationship C_k incorporating duration is defined as follows:

$$C_k(f_i, f_j, \triangle t) = \begin{cases} 1 \text{ (co-occurring with offset } \triangle t) \\ 0 \text{ (otherwise)} \end{cases} \tag{2}$$

where, "co-occurring with offset $\triangle t$" means that either $t_i^e + \triangle t (\triangle t > 0)$ or $t_i^s + \triangle t (\triangle t < 0)$ is in section $[t_j^s, t_j^e]$. C takes the following value if $\triangle t = 0$.

$$C_k(f_i, f_j, 0) = \frac{\text{overlap}}{t_i^e - t_i^s} \tag{3}$$

where "overlap" represents the overlapping length of $[t_i^s, t_i^e]$ for f_i and $[t_i^s, t_i^e]$ for f_j. The formula below represents the relationship between feature $f_i (\in F_a)$ and $f_j (\in F_b)$

$$\hat{C}_k(f_i, F_b, \triangle t) = \max_{f_j \in F_b} C_k(f_i, f_j, \triangle t) \tag{4}$$

The performance measure that we take is the average of $\hat{C}_k(f_i, F_b, \triangle t)$ over all $f_i \in F_a$.

$$\tilde{C}_k(F_a, F_b, \triangle t) = \frac{\sum_{f_i \in F_a} \hat{C}_k(f_i, F_b, \triangle t)}{N(f_i \in F_a)} \tag{5}$$

Thus, $\tilde{C}_k(F_a, F_b, \triangle t)$ indicates the ratio $f_i \in F_a$ has correspondence with $f_j \in F_b$ which satisfies C_k with the time offset $\triangle t$[1]. If both F_a and F_b are items of a frequent pattern, $\tilde{C}_k(F_a, F_b, \triangle t)$ exhibits a temporal characteristic of the pattern.

Pseudo mutual information

The frequent pattern is less important when the occurrence probability of each element is large, even if co-occurrence probability of the elements of a frequent pattern is also large. The same argument holds for the case of low co-occurrence probability with low occurrence probability of each element.

We consider mutual information as representing temporal characteristics of frequent patterns in place of co-occurrence probability. More specifically, we defined pseudo mutual information $\tilde{I}_{\triangle t}(F_b; F_a)$ using $\tilde{C}_k(f_i, F_j, \triangle t)$ given in the above formulae (2)–(5).

$$\tilde{I}_{\triangle t}(F_b; F_a) = \sum_{f_j \in F_a} \sum_{f_i \in F_b} \tilde{C}_k(f_i, f_j, \triangle t) \log \frac{\tilde{C}_k(f_i, f_j, \triangle t)}{P(f_i)P(f_j)} \tag{6}$$

where $P(f_i)$ represents occurrence probability of a single feature f_i.

Because $\tilde{C}_k(f_i, F_j, \triangle t)$ was used in place of $P(f_i, f_j)$ for mutual information, pseudo mutual information ($\tilde{I}_{\triangle t}(F_b; F_a)$) is a rough estimate of mutual information.

For experiments in the following sections, we predominantly assessed pseudo mutual information in the case of $F_a = \{f_a, \overline{f_a}\}$ or $F_b = \{f_b, \overline{f_b}\}$, whether feature f_a or f_b occurred or not.

TABLE I. EXPERIMENTAL SYSTEM

worker	USB camera attached to the head with a hair band, a headset with a microphone and a headphone, notebook PC
helper	notebook PC
video communication	Skype (around 10 fps)

TABLE II. WORKER-HELPER PAIRS

pair	worker	helper	features
A	N1 (novice)	M1 (experienced)	N1 consults M1 well and carefully performs a task
B	N2 (novice)	M1 (experienced)	N2 often consults M1 often and sometimes performs a task with M1's own idea.
C	M2 (experienced)	N1 (novice)	M2 does not often ask N1's advice and performs a task with M2's own idea.
D	M3 (experienced)	M1 (experienced)	M3 does not often ask M1's idea, but tries to perform according to any given advice.
E	N1 (novice)	N3 (novice)	N1 and N3 consult each other and carefully perform a task with sufficient communication

VI. DATA COLLECTION AND FREQUENT PATTERN EXTRACTION

A. Data collection

The experiments were conducted with a prototype system as shown in Figure 1 and Table I. The video captured by the USB camera and microphone was transmitted to a helper by using Skype with QVGA quality. The voice of a helper alone was transmitted to a worker, because the image of the helper is less important[2].

We chose a cooking task in a kitchen, wherein the worker, unaware of the location of kitchenware and seasoning, cooks an unfamiliar menu. The entire task took around 30 minutes.

B. Worker and helper

To investigate how the means of communication varied with respect to working skills and personal character, we gathered several participants of varying skill levels and formed a several worker-helper pairs as shown in Table II.

C. Multimodal features

We selected features based on the possibility of automated feature detection. The ground truth data, however, were manually collected, because perfect accuracy cannot be expected at this moment.

Visual features

Table III shows the visual features used in the experiments. Those are possibly detected by image processing, for example,

[1]This measurement is based on the same idea in formula 2. However, it is not exact probability, because it includes the process of taking max in Formula 4

[2]This type of asymmetry is discussed by Billinghurst [11]. If a worker needs to look at instruction in images or behaviors of a helper, we may need a head-mount display or such carrying devices.

TABLE III. FEATURES FROM VIDEOS

feature	condition	abbreviation
gazing still	no or small camera motion	V:h
looking around	camera rotation	V:l
movement (location)	camera motion for going forward	V:m

TABLE IV. FEATURES FROM SPEECHES

features	condition	abbreviation
concrete object name	classification in thesaurus [22]	c
exophora	demonstrative (including restrictive modification)	s
role of utterance	request, question, explanation, reply	R, Q, D, T

behaviors of looking and movements can be detected by camera motion detection [18].

Speech features

Table IV shows the features extracted from the transcripts which were manually transcribed. They were semi-automatically detected by a combination of natural language processing [20], [21] and manual corrections.

D. Result of frequent pattern detection

The total length of recorded data was 112 minutes, and the average length was 22.5 minutes. 758 utterances and 221 visual features were detected. Table V shows the number of occurrences of each feature.

We applied PrefixSpan to the feature sequences detected from the recorded data. The minimum support of PrefixSpan was set to 0.1. 222 frequent patterns were extracted from the recorded data; the number of occurrences of each frequent pattern ranged from 22 to 263.

The lengths of detected patterns are from 2 to 6, and the actual number of patterns corresponding to each length is shown in Table VI. Figure 3 shows the numbers of occurrences in descending order: up to the 70th pattern, we see only trivial combinations of utterances, which show two people are talking to each other.

Descending from 80th, we find interesting patterns as shown in Figure VII. In the followings, we denote any feature, for example X, in a worker's and helper's utterances as $W : X$ and $S : X$, respectively. All patterns with the length 5 and 6 are combinations of utterances that are not informative. Some patterns of length 3 and 4 show interesting combinations of multimodal actions. $<$ S:D, W:P, S:D $>$ in the first row in Figure VII represents a worker acknowledging (back-channeling) while listening to explanations from a helper. Patterns from the second row show typical characteristics of an FPV communication, such as a helper explaining an object to a worker while they are looking at the object, and a worker

TABLE VI. NUMBER OF FREQUENT PATTERNS AT EACH LENGTH

Length	Number
1	15
2	41
3	70
4	49
5	36
6	11

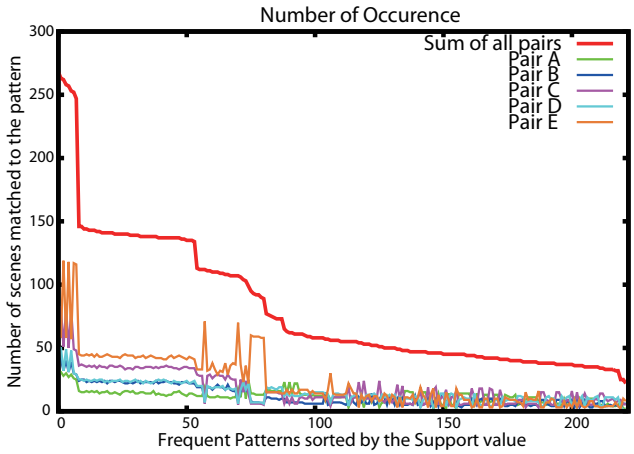


Fig. 3. Number of occurrences (frequent patterns)

moving on according to a helper's request. Such patterns show clearly that two people tried to share attention and information necessary for collaboration; for example, a worker gives a helper information necessary to provide advises.

E. Behaviors and communications related to failures

In 41 cases of the experiments, a worker failed or nearly failed. Table VIII shows some typical examples. "Lack of information" and "inappropriate timing" were the primary causes of those failures produced by insufficient communication. Concerning lack of information, not a number of failures or near–failures were caused by an absence of response or back-channeling, such as an acknowledgment by nodding. Others were caused by a lack of the helper's attention. Failures related to inappropriate timing include cases in which the worker could not request an instruction because of the helper's behavior, and many were caused by the delay in the helper's advice.

VII. ANALYSIS AND DISCUSSION

We examined relationships between the characteristics of frequent patterns and the behaviors of a worker and a helper. The idea was based on the following discussion:

(a) Frequent "fundamental patterns" are required to exchange essential information in FPV communications.
(b) Lack of information has two factors: (i) information is lacking even with one of the fundamental patterns, and (ii) One or more elements is/are missing from a fundamental pattern.
(c) Inappropriate timing has two factors: (i) timing of one or more elements of a fundamental pattern is/are delayed or advanced, and (ii) the occurrence order of the elements are changed and are not detected as a fundamental pattern

Because the purpose of this research is the analysis without deep semantics, we can only deal with the above (b)(ii), (c)(i). We leave (b)(i),(c)(ii) for future work.

TABLE V. NUMBER OF OCCURRENCES (FEATURE)

feature	symbol	occurrence		
		occurrence	worker	helper
utterance	-	757	427	330
object name	c	108	66	42
exophora	s	79	61	18
request	R	27	19	8
question	U	33	22	11
explanation	D	530	327	203
response	P	167	59	108

feature	symbol	occurrences
gazing still	h	104
looking around	l	65
location movement	m	57

TABLE VII. EXAMPLES OF RELATIONS OF THE COMMUNICATION AND FREQUENT PATTERNS

Order	pattern	typical situation	number of occurrences
79	< S:D, W:P, S:D >	helper: explanation or advice worker: promoting next utterance by back-channeling	92
119	< W:D, V:h, S:D, W:D >	worker: explanation or report of the situation helper: advice, explanation, or notification	54
172	< V:l, S:D, W:D, S:D >	worker: explanation or report of the situation helper: confirmation of the situation	42
177	< V:l, W:D, S:D, W:D >	worker: report of the situation helper: acknowledgment	40
214	< S:D, W:D, V:m >	helper: explanation of tools or foodstuffs worker: confirmation of the place, etc.	33

TABLE VIII. SITUATIONS WITH FAILURES OR NEAR-FAILURES

factor	situation	cause for the situation
lack of information in communication	wrong use of a tool by the worker	not noticed by the helper nor a question from the worker.
	the worker failed to find a tool	the helper did not recognize the situation, and did not give good advice
inappropriate timing of communication	the worker needed to change the method in the middle of a manipulation	delayed advice
	the worker chose a wrong method because the worker failed to ask a question	the helper kept speaking without being aware of the demands of the worker
a simple mistake or accident	the worker pushed down some dishes	lack of attention to the environment (not a communication problem)

A. Number of occurrences of frequent patterns

We analyze (b)(ii) in the previous section as follows. We consider the subpattern F_i^j derived by removing jth element of F_i. $N(F_i)$ is the number of occurrences of F_i, and r (= $r(F_i, j) = N(F_i)/N(F_i^j)$) represents the ratio of the occurrences of F_i^j, F_i. r_{all} the ratio for the all pairs of participants, and $r_A(F_i, j)$ – $r_E(F_i, j)$ the ratio for pairs A–E, respectively. Note that $r \leq 1$ always holds as F_i^j is detected for every F_i in PrefixSpan.

Table IX shows some examples of r. Values are underlined if a significant difference (significance level of 5%) was observed.

The first row of Table IX shows the ratio if a worker's response ($W : P$) was observed between utterances of a helper as < S:D, W:P, S:D >. Participant pair C had a significantly low ratio of r, while pair E had a significantly large r. These values coincided well with M1's behaviors and the fact N1 and N3 consulted together, as described in Table II.

The second row of Table IX indicates the pattern (< V:l, W:D, S:D, W:D >) shown in Table VII, where r represents the ratio a worker did not report or explain something after looking around. r was smaller for participant pair A and E compare to other pairs. The helpers in A and E often gave advice or explanation without waiting for a subsequent report or explanation from the workers. This suggests that the helpers noticed the workers intentions of looking around before any utterances from the workers.

Other cases with significant differences are given in the third and fourth rows of Table IX. These patterns are combinations of utterances for explanation or report (< S:D, W:D >), looking around (V:l), and moving (V:m). For participant pair A, with regard to r, the same argument as the above (for the second row) holds. In the case of participant pair D, movement of the worker was significantly frequent, due to a high level of skill and knowledge of the next necessary step. As a consequence, they often talked in advance about the foodstuffs or tools that the worker used.

B. Temporal characteristics of frequent patterns

For (c)(i) mentioned above, we examined the temporal changes of pseudo mutual information and its relationship to the characteristics of FPV communications. Figure 4 and Figure 5 show the temporal characteristics of $I_{\triangle t}(F_b; F_a)$. The "0" at the center of each graph represents the time of an F_a occurrence, with the left portion indicating that F_b occurred before F_a and vice versa. The horizontal axis represents the value of pseudo mutual information[3].

Figure 4 shows the temporal characteristics of pattern < S:D, V:l > in which "looking around" occurs after or before explanation by the helper. Participant pair A exhibited characteristics different from that of other pairs. The mutual information had small peaks (i) and (ii) around 10 seconds before and after an explanation. This feature is a product of

[3]Pseudo mutual information sometimes has negative values because it is calculated based on a rough approximation of probability.

TABLE IX. THE VALUE OF r FOR ALL DATA AND EACH PAIR'S DATA

$N(F_i)/N(F_i^j)$	r_{All}	pair				
		r_A	r_B	r_C	r_D	r_E
$N(<$ S:D, W:P, S:D $>)/N(<$ S:D, S:D $>)$	0.357	0.5	0.222	0.12	0.233	0.496
$N(<$ V:l, W:D, S:D, W:D $>)/N(<$ V:l, S:D, W:D $>)$	0.90	0.83	1.0	1.0	1.0	0.67
$N(<$ V:l, S:D, W:D $>)/N(<$ S:D, W:D $>)$	0.27	0.75	0.174	0.139	0.36	0.214
$N(<$ S:D, W:D, V:m $>)/N(<$ S:D, W:D $>)$	0.225	0.188	0.174	0.278	0.44	0.095

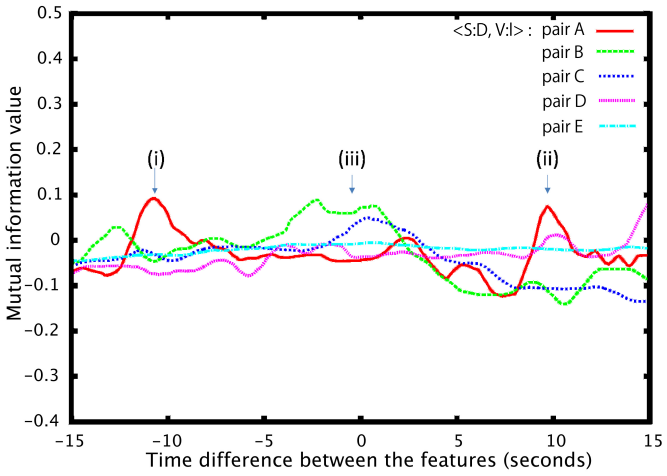


Fig. 4. Pseudo mutual information of $<$ S:D, V:l $>$

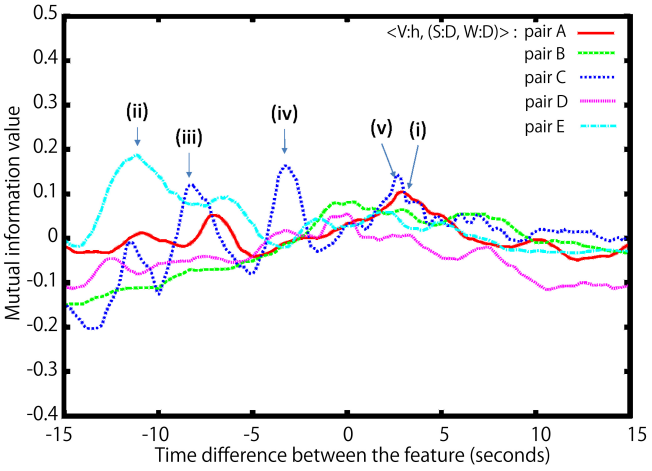


Fig. 5. Pseudo mutual information of $<$ V:h, (S:D, W:D) $>$

the behavior of participant N1 and M1: N1 looks around after hearing an utterance of M1, and M1 provides explanations realizing the N1's intention. In contrast, in pair B, explanation and "looking around" occurred simultaneously, as indicated by gentle peak (iii). Participant N2 looks around before N1 has finished an utterance, or N1 starts an explanation while N2 is still looking around. These behaviors make communication less efficient as it is difficult to recognize and confirm an object when looking around. Participant pair C had a similar feature; however, it was less significant. For Pair D and E, this feature was not observed.

Figure 5 shows the temporal characteristics of pattern $<$ V:h, (S:D, W:D) $>$ in which an explanation of a helper and a report of a worker are given after gazing still. The graph shows that participant pair A has peak (i) at 5 seconds after gazing still, and that pair E has peak (ii) before gazing still. This differences can be explained by the participants' behaviors: In pair A, helper M1 often gives information triggered by an action of worker N1; in pair E, worker N1 and helper N3 often consult before a task, then, N1 acts as N3 requests. On the contrary, the value of pair C has two or more peaks (iii), (iv), and (v) before and after gazing still. This characteristic is caused by worker M2's behavior that M2 starts speaking instantly when M2 finds or notices something. Subsequently, helper N1 cannot grasp M2's state well, and gives information on tools or the next task on an ad hoc bases. Participant pair B and D have no significant characteristics, while they are closer to A than to C.

TABLE X. MAIN FACTORS OF FAILURE OR NEAR–FAILURE

main cause	A	B	C	D	E	total
lack of communication/information	2	10	6	1	1	20
inappropriate timing	2	2	1	6	1	12
simple mistake or slip		6	2		1	9
total	4	18	9	7	3	41

C. Failures and frequent patterns

Table X shows the manual classification of the failures or near. The numbers clearly differ among participant pairs.

For participant pair B and C, there was a larger number of failures or near–failures caused by a lack of information. As we discussed in Section VII-A and VII-B, these failures are the reasonable result of pair B and C's quantitatively measured characteristics: less frequency of back-channeling or response, inappropriate timing among utterances, looking around, and gazing. Interestingly, insufficient communication easily caused a failure for pair B, which was not always the case in pair C, as the worker was a beginner in pair B and experienced in pair C.

For participant pair D, a larger number of failures or near–failures caused by inappropriate communication timing. Worker M3 in pair D is experienced in cooking and behaved as he saw fit;, consequently, helper M1 could not give appropriate messages at the appropriate time. As a result of this behavior, Figure 4 and 5 reveal no peaks of utterance (explanation or report) of the worker in pseudo mutual information.

Quantitative measurements show that participant pairs A and E carefully communicated in performing the task, resulting in a smaller number of the failures or near–failures. Some failures were caused by a situation in which the helper does not know what kind of advice is appropriate.

D. Discussion

As shown above, the numbers of occurrences and temporal characteristics of frequent patterns are a good match to the features of participant pairs, which suggest the causes of possible

failures. Features used in this analysis can be automatically detected, although not perfectly, suggesting, we can expect automatic or real-time analysis can be expected in the future.

Patterns and characteristic are at present, however, manually chosen from the obtained data. As a consequence, all possibilities are not thoroughly examined. Further research is required for such a systematic investigation with indices needed for picking up important frequent patterns and their characteristics. In other words, in Section VII-A and VII-B, the authors chose important and interesting frequent patterns for discussion; however, there is currently no method to automatically choose them. Moreover, the method of examining characteristics of chosen frequent patterns needs to be delineated.

Furthermore, the amount of data that is valid for extracting frequent patterns must also be considered. In our experiments, patterns that include questions (W:Q or S:Q) were not extracted as frequent patterns, as the number of such occurrences was relatively small. A more extensive examination of the data size and the variety of patterns is left for future work.

In this research, our aim was to deal with only low-level features without deep semantics. Although this trial was successful, semantics eventually need to be dealt with in the future in order to clarify the power and the limitation of our method.

VIII. Conclusion

This paper introduced an analysis of working support through FPV communication. First, we discussed the characteristics of FPV communication, the importance of frequent patterns, and their extraction. Next, we showed that the number of occurrences and temporal characteristics of frequent patterns appropriately expresses the characteristics of the behaviors and communications of worker–helper pairs. We conjectured that some of them were related to the smoothness of the work outcome and failures. These findings enable a useful working and communication analysis for an environment where multiple people are collaborating.

The analysis was based on multimodal features semi-automatically detected as ground truth data; yet, further study is needed to fully realize automatic data analysis. An automated process would also be useful also for real-time working support, e.g., analyzing communication patterns and providing advice to a worker or a helper using this system.

References

[1] P. Garner, M. Collins, S. Webster, D. Rose, “The application of telepresence in medicine”, BT Technolo J, Vol.15, No.4, pp.181–187, 1997

[2] “UMET”, http://www.tanizawa.co.jp/umet/,

[3] J. Siegel, R. Kraut, B. John, K. Carley, “An Empirical Study of Collaborative Wearable Computer Systems”, Proc of the ACM Conference on Computer Supported Cooperative Work (CSCW1995), pp.312–313, 1995

[4] S. Mann, “Smart Clothing: Wearable Multimedia Computing and Personal Imaging to Restore the Technological Balance Between People and Their Environments”, Proc. of the ACM Conference on Multimedia, pp. 163–174, 1996

[5] B. Rhodes, The wearable remembrance agent: a system for augmented memory, In Proc. of ISWC 1997, pp. 123–128, 1997

[6] H. Wactlar, M. Christel, A. Hauptmann, Y. Gong, ”Informedia Experience-on-Demand: Capturing, Integrating and Communicating Experiences across People, Time and Space”, ACM Comput. Surv., vol. 31, no. 9, 1999.

[7] Y. Sumi, et al., Collaborative capturing and interpretation of interactions. Pervasive 2004 Workshop on Memory and Sharing Experience, pp 1-8, 2004

[8] J. Gemmell, G. Bell, R. Lueder: “MyLifeBits: a personal database for everything”, Communications of the ACM, vol. 49, Issue 1 (Jan 2006), pp. 88–95. 2006.

[9] R. Kraut, M. Miller, J. Siegel, “Collaboration in performance of physical tasks: Effects on outcomes and communication”, Proc of the ACM Conference on Computer Supported Cooperative Work (CSCW 1996), 1996

[10] S. Fussell, R. Kraut, J. Siegel, “Coordination of communication: Effects of shared visual context on collaborative work”, Proc. of the ACM Conference on Computer Supported Cooperative Work (CSCW 2000), 2000

[11] M. Billinghurst, S. Bee, J. Bowskill, H. Kato, “Asymmetries in Collaborative Wearable Interface”, The Third International Symposium on Wearable Computers, pp.133–140, 1999

[12] R. Kraut, D. Gergle, S. Fussell, “The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-Presence”, Proc. of the ACM Conference on Computer Supported Cooperative Work (CSCW 2002), 2002

[13] D. Gergle, R. Kraut, S. Fussell, “Action as language in a shared visual space”, Proc. of the ACM Conference on Computer Supported CooperativeWork (CSCW 2004), 2004

[14] H. Clark, “Referring as a collaborative process”, Cognition, Vol.22, pp.1–39, 1986

[15] S. Norris, Analyzing Multimodal Interaction: A Methodological Framework, Routledge, 2004

[16] S. Norris, Identity in (Inter)action, De Gruyter Mouton, 2011

[17] “Failure Knowledge Database”, http://www.sozogaku.com/fkd/en/index.html

[18] S. Kubota, Y. Nakamura, Y. Ohta. Detecting scenes of attention from personal view records – motion estimation improvements and cooperative use of a surveillance camera, IAPR Workshop on Machine Vision and Applications, pp. 209–213, 2002.

[19] J. Pei, J. Han, B. Mortazavi-asl, H. Pinto, Q. Chen, U. Dayal, M. Hsu, PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth. 17th International Conference on Data Engineering (ICDE ’01), pp.215–224, 2001

[20] S. Kurohashi, D. Kawahara. Japanese morphological analysis system JUMAN version 5.1 manual. 2005.

[21] S. Kurohashi, M. Nagao. A syntactic analysis method of long Japanese sen- tences based on the detection of conjunctive structures. Computational Linguistics, 20(4) pages 507–534, 1994.

[22] National Language Research Institute Publications, Word List by Semantic Principles (Bunruigoihyou), Shuei Shuppan, 1964. in Japanese

[23] E. Schegloff, H. Sacks: Opening up closings, Semiotica, Vol.8, pp.289-327, 1973.