PATTERNS THAT INDUCE ADVICE IN REMOTE COOKING SUPPORT VIA FIRST PERSON VISION COMMUNICATION

Kanako OBATA†, Yuichi NAKAMURA†, Takahiro KOIZUMI†, Kazuaki KONDO†, Yasuhiko WATANABE‡

†Academic Center for Computing and Media Studies, Kyoto University, JAPAN
‡Faculty of Science and Technology, Ryukoku University, JAPAN

ABSTRACT

In this study, we introduce a novel scheme for distance learning that aims to teach people how to cook, thereby reaching people who are unable to join online cooking classes, supporting rehabilitation and care for single-living elderly people, and handing down dietary knowledge and culture. Our scheme utilizes a framework of working support through first person vision (FPV) communication, in which a worker with a head-mounted camera works under the guidance of an experienced mentor monitoring the FPV from a distance. To make this scheme fully effective, we investigated the communication that takes place in this environment, *i.e.*, how typical response patterns and absence pauses induce advice. Through experiments, temporal characteristics of such features are found to be tightly related to the occurrences of advice.

1. INTRODUCTION

People who live alone tend to eat out, or purchase boil-in-the-bag foods, therefore consuming meals with high quantities of salt and fat and fewer vegetables. Such dietary habits easily undermine their health. Cooking by themselves would greatly improve their situations, providing enough nutrition for better body and mental health. Cooking is also known to be good for physical and mental rehabilitation and care for people with dementia or higher brain dysfunction, because cooking requires task planning, manual skills including hand operations, and training on attention distributions.

With the recent trend of nuclear families and people living alone, our cultures face difficulties in handing down cooking skills, dietary knowledge, and food culture that used to be kept in families and communities. People with jobs during the day do not typically have enough time to join cooking classes. Moreover, family members often find it difficult to take care of the elderly while they are cooking, especially when they live apart from the family. Without enough care and supervision, cooking could be dangerous and possibly causes serious accidents involving fire, cutlery, boiled water, and so on.

To provide opportunities to learn to cook and hand down aspects of one's culture, we propose a cooking support system using first -person vision (FPV) technologies. As illustrated in Figure 1, our scheme utilizes a framework of working support through FPV communication in which a worker with a head-mounted camera works under the guidance of an experienced mentor monitoring the FPV from a distance. 1. One important advantage of this framework is that a worker can use both hands without the need to hold a camera while the FPV is transmitted to the helper. Another advantage is that the helper can see what the worker sees, responding and referring to what the worker's attention is on, which cannot be realized with fixed surveillance cameras. Using our scheme, a person cooking can get help or care from his or her family or trainer from a distant location. We also expect that cooking becomes more enjoyable without a feeling of complete loneliness in the kitchen.

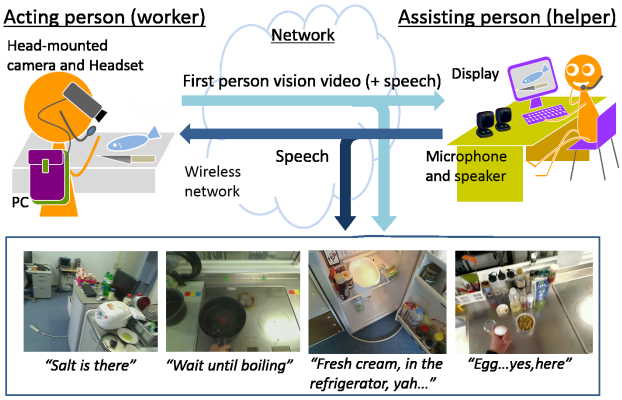


Fig. 1. Overview of a working cooking support system using FPV communication

However, to exploit these advantages, we need to consider communication factors. Communication channels in distance-based cooking support are limited compared with ordinary situations wherein a helper is in the same kitchen; a helper is not able to physically touch or handle food and kitchenware, implying that we need to pay close attention to ensure communication is smooth and flexible enough to keep cooking in good and safe conditions. Furthermore, the communication channels must be natural enough such that users feel comfortable handing down cultural knowledge.

On the basis of the above considerations, we investigated the relationships between communication patterns and occur-

rences of advice[1] in distance-based cooking support. More specifically, we gathered statistics on occurrences of advice, *i.e.*, response patterns that typically consist of pairs of actions and corresponding reactions, as well as absence pauses. Our results showed that temporal characteristics of those communication patterns were tightly related to the occurrences of advice, thus suggesting system and communication requirements to achieve better teaching and care in distance-based cooking support systems.

2. RELATED WORKS

Various forms of mobile video communication have been explored from the early days of wearable computer devices. In [3], Kraut et al. performed a case study of bicycle maintenance using video communication. While video communication did not significantly affect work efficiency and accuracy, the behavior of the workers and experts changed significantly with the addition of video communication, in particular drastic increases in proactive behavior on the part of the experts. In [4], Fussel et al. reported a similar experiment in which they observed significant differences in the use of visual information, *i.e.*, deictic expressions often appeared that made conversation simpler, though the worker and helper had differences in this tendency. In [6], Kraut et al. devised a collaborative online jigsaw puzzle and analyzed the behavior of a worker and a helper ; the shared video space caused differences in the acknowledgment of understanding and behavior as well as deictic expressions.

These working support schemes have not yet been applied to cooking. In cooking situations, we have little knowledge as to what occurs, how it occurs, and why. Moreover, an important difference that we propose is the means of analyses. Previous works argued differences in efficiency and communication style among different styles of working support. In contrast, our study focuses on a more detailed analysis of relationships between communication behaviors and opportunities for giving advice. A crucial aspect of our work is quantitative analysis based on low-level features that can be automatically detected.

3. COMMUNICATION PATTERNS

Figures 2 – 4 show typical situations with our distance-based cooking support system using FPV communication. In each case, a worker is performing unfamiliar cooking tasks with the assistance of an expert. We considered how utterances and behaviors were related and how advice was given or failed to be given. Figure 2 shows a case in which a task was performed smoothly with sufficient information exchanged via FPV communications. The helper gave advice in responding to the explicit request of the worker, who was clearly aware of the need for information. In such a case, we had few problems or accidents.

[1]We consider advice as a general word here that represents instruction, suggestion, tips, and any other helpful information for a recipient.

(a-1) (a-2) (a-3)

Worker (a-1): “Bread flour...? Do I use this, too ?” (pointing to the plastic bag of bread flour)
Helper (a-2): “Yes, it’s for dusting a cake tin. We are going to use weak flour for the cake.”
Worker (a-3): ”Ah, I see.”

Fig. 2. Example of communication and advice (a)

(b-1-1) (b-2-1) (b-3-2)

Worker (b-1-1): (shaking salt shaker)
Helper (b-1-2): “Rotate it, not shaking...”
Worker (b-1-3): “OK, I got it!”

Worker (b-2-1): (looking at two bags of sugar with his hands)
Helper (b-2-2): “The red one is caster sugar, and the blue one is granulated sugar. Use blue one.”

Helper (b-3-1): “Oil the pan, first.”
Worker (b-3-2): “OK. Done.”
Helper (b-3-3): “Preheat the pan a little...”
Worker (b-3-4): “OK...”
Helper (b-3-5): “You can make oil flat and smooth with kitchen paper”

Fig. 3. Example of communication and advice (b)

In all cases shown in Figure 3, the helper voluntarily gave advice while watching the worker’s behavior, implicitly interpreting such behaviors as requests for advice. In general, a worker often needs to concentrate on the task at hand, and does not have enough time to explicitly ask questions or even be aware of the needs for advice. In (b-1-1), the helper noticed that the salt did not come out, because the worker handled the salt shaker in the wrong way. In (b-2-1), the helper noticed that the worker was not sure which sugar should be used; the hand and head movements were relatively small in this case, showing that the helper might have needed to help the worker make a decision. In (b-3), sufficient advice was induced because of the good tempo of communication that confirmed the worker was listening to the helper.

Conversely, the helper failed to give advice in the cases shown in Figure 4. In (c-1), the helper was not encouraged to give advice or missed the timing to give advice, because the interaction between them was not coherent or smooth. In (c-2), the helper did not notice the need to give advice, because the worker introduced a topic that was different from the previous one.

Overall, the problem is how to keep and reinforce the conditions in situations similar to (b) and avoid situations similar to (c). For this purpose, we need to consider the steps of giving advice, *i.e.*, a helper notices the need to give advice, thinks

(c-1-2) (c-1-4) (c-2)

Helper (c-1-1): "Slice it (cake) into two pieces."
Worker (c-1-2): "...too difficult..." (without response or backchanneling)
Helper (c-1-3): "Ha-ha-ha!" (laughing and missing the chance for advice on how to slice the cake)
Worker (c-1-4): "Ah, ... ugly curve!"

Helper (c-2-1): "Spread butter to the cake pan."
Worker (c-2-2): "I love oil spray. Buy one please. It makes cooking easy."
Helper (c-2-3): "Hmm, it sounds convenient." (missing the change for advice on explaining how much butter the worker needs to use)

Fig. 4. Example of communication and advice (c)

of the appropriate advice, and gives it.

We suggest that there are two communication patterns that are tightly related to those factors. First, we have the *pause* in which a helper needs enough time to watch, think, and give advice. A pause, *e.g.*, a head motion pause or a pause in conversation, gives good opportunities to watch, think, and give advice. Second, we have the *response* in which appropriate responses with a good tempo in relation to the other's speech encourage the helper to give advice and suggest appropriate timing for giving advice.

Note that one aspect of the pause is the time gap between one element and another, *i.e.*, as part of a responding action. We therefore consider that we can partially analyze pauses depending on the patterns of response; however, another aspect of the pause in which nothing is occurring in one or some modalities does not match the above idea. We regard this aspect of the pause as an absence pause and handle it independent of response patterns.

4. ANALYSIS OF COMMUNICATION PATTERNS

4.1. Pause and Response Patterns

Communication patterns of pause and response described in Section 3 are composed of elements of interactions, such as utterances, motions, hereafter called features. The purpose of our research is, to investigate relationships between combinations of features and occurrences of giving advice; however, the number of feature combinations can be large, most of which are useless. Below, we explain how feature patterns are chosen from a vast amount of combinations.

1. Extract frequent patterns, *i.e.*, feature combination patterns that frequently appear in actual communication.
2. Extract response patterns,*i.e.*, basic action-reaction patterns that are included in selected frequent patterns.
3. Statistically analyze the relationships between advice and temporal characteristics of selected response patterns.

Table 1. Example of features

modality	property	symbol
advice (utterances)	now, future	A:n, A:f
other utterances	request, question, explanation, reply	R, U, D, P
gazing still	no or small camera motion	V:h
looking around	camera rotation	V:l
movement (location)	camera motion for going forward	V:m
movement (hand)	hand motion	V:b

Frequent patterns are useful for the following reasons:

- If we collect enough samples in which tasks are performed in conjunction with good advice, their logs include enough samples of good communication patterns inducing advice.
- Irrelevant or irregular communication patterns may occur even in ordinary cases; nevertheless, the number of occurrences is relatively smaller than that of relevant ones.

Note that some common pause patterns, each of which is composed of a preceding feature and a following feature with a time gap in between, are also extracted as response patterns if they frequently appear in the sample data. We can therefore obtain, both pause and response patterns using the above method.

4.2. Component Features

As communication elements that carry information from one person to an other, we consider the following features in three modalities:

Speech: utterance of advice (for now, for future), request, query, explanation, and so on.
Visible actions: hand motions/pauses, location movements, gestures.
Observational actions: staring/gazing still, looking around.

We categorized visible actions and observational actions as different modalities though their original source is visual information; more specifically, visible action primarily represents what a worker is doing, and observational action directly expresses what information a worker requires.

Table 1 shows the features that we used in our experiments, detected by natural language processing and image processing, for example, detecting movements via camera motion detection [9].

4.3. Frequent Pattern Extraction

In this subsection, we outline frequent pattern extraction (please refer to [10] for further details) . In sequential pattern mining, a transaction is a sequence of items occurring in order of their occurrence times. We considered each feature as an item and a transaction as a sequence of features, *e.g.*, $S = \{F_1, \ldots, (F_i, \ldots), \ldots, F_z\}$, where

Table 2. Examples of frequent patterns

frequent pattern	av. #	av. time (sec)
<(W:D) (S:D) (W:P)>	6.12	15.70
<(S:R) (W:D) (S:D)>	7.50	24.03
....	...	....
<(V:h) (W:D) (S:D)>	7.81	20.94
<(S:D) (S:D) (W:P)>	7.87	20.28
<(S:R) (V:h) (S:D)>	7.89	26.15
<(W:D) (S:R) (S:D)>	7.89	26.15
....	...	....
<(S:D) (W:D) (S:R)>	8.23	25.10
<(S:D) (V:h) (W:D)>	8.33	20.88
....	...	....

av. #: average number of other features occurring in the matched period
av. time: average length of matched period (sec)

F_j represents a feature type such as looking around or 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 $s_j =< F_{j1}, \ldots, (F_{jp}, \ldots), \ldots, F_{jz} >$.

We applied PrefixSpan [11], a sequential pattern mining tool, that allows simultaneous occurrences of two or more items. Table 2 shows actual frequent patterns extracted from our distance-based cooking support logs. Abbreviations, W, S, and V represent a worker, a helper, and the worker's visual information, respectively. Other symbols are summarized in Table 1. As an example, W:D indicates a worker's utterance of a description (*i.e.*, situation).

4.4. Response Patterns

Response patterns are selected according to the following criteria:

- A response pattern is a combination of two features included in frequent patterns with three or more features.
- A response pattern that has the possibility of a fewer number of other co-occurring features is preferred.

Both criteria are effective for eliminating vast amounts of useless patterns. The former is necessary because frequent patterns with two elements frequently occur irrespective of their relationship if each of the elements frequently occurs on its own; here we need to eliminate them. The latter originates from the idea of frequent patterns identified by PrefixSpan; it allows other features to simultaneously occur while a frequent pattern is occurring, *i.e.*, from the time the first feature started until the time the last feature finished. Table 2 shows the statistics from this viewpoint. If the time length becomes longer, the possibility of incoherent situations in which all elements do not share the same topic or focus[2] increases. We therefore prefer response patterns with smaller numbers of co-occurring features. Table 3 shows response patterns selected from the frequent patterns shown in Table2.

[2]"Focus" can be an ambiguous word; here, we use "focus" to mean "focus of attention."

Table 3. Response patterns extracted from frequent patterns

response pattern	ATG (sec)	SDG (sec)
<W:D S:D>	-0.207	2.008
<S:D W:P>	-0.257	2.121
<S:R W:D>	0.128	2.517
<W:U S:D>	-0.499	1.785
<S:D W:D>	-0.127	2.187
<V:h S:D>	-0.270	2.492
<S:R V:h>	0.055	2.695
<W:D S:R>	0.004	2.626
<W:D S:P>	-0.198	2.328
<S:D V:h>	0.040	2.661
<V:l S:D>	-0.069	2.449
<S:D V:m>	0.131	2.680

ATG: Average length (time) of the gap between two features
SDG: Standard deviation of the gap between two features

4.5. Absence Pauses

As discussed in Section 3, we consider absence pauses independent of time gaps within response patterns. Let $F_r(t_i)$ denote that one or more instances of feature F_r occurs. Then, an absence pause is any instance of feature set $\{F_r\}$ not occurring within period $t_i \leq t \leq t_i + \Delta t$, *i.e.*, $\{\neg F_r(t)|^{\forall} F_r \in \{F_r\}, t_i \leq t \leq t_i + \Delta t\}$. We denote this as $M(\{F_r\}, t, \Delta t)$.

5. ANALYSIS BY COMBINING ADVICE AND COMMUNICATION PATTERNS

Let us denote an occurrence of a response pattern by $R(F_p, F_q, \Delta t_1)$ where feature instances $f_p \in F_p$ and $f_q \in F_q$ occur with time gap Δt_1 between their occurrences, *i.e.*, the difference between end time (t_{pe}) of f_p and the start time of f_q. A positive Δt_1 indicates an ordinary gap, whereas a negative Δt_1 means that those two features have overlap in their occurrences. An occurrence of giving advice is denoted by $A(t)$.

The co-occurrences of response patterns and advising are evaluated by mutual information as follows:

$$I_r(R(F_p, F_q, t_{pe}, \Delta t_1); A(t_{pe} + t_2)) \text{ s.t. } 0 \leq t_2 \leq t_{th} \quad (1)$$

Here, t_{th} is the threshold value for the time difference in which a response pattern and advice are related to one another. Similarly, mutual information between an absence pause and advice can be represented by the following formula:

$$I_p(M(\{F_r\}, t, \Delta t_1); A(t + t_2)) \text{ s.t. } 0 \leq t_2 \leq t_{th} \quad (2)$$

Here, t_{th} is the same as above.

Mutual information indexes given in equations (1) and (2) are good clues to identifying the correlation between the gap length of a response pattern and advice and the correlation between the gap length and advice, respectively. For example, if I_r or I_p is significantly large at t_1, a gap with t_1 length has good characteristics for inducing advice. In contrast, if

I_r or I_p remains small with no significant variations with the change of t_1, we cannot find a relationship between a gap or pause length and advice. By examining I_r for every response pattern and I_p for every absence pause, we can detect how response patterns and pauses affect advice.

6. EXPERIMENTS

6.1. Data Collection

Purpose of the experiments: Given the hypotheses on the relationships between advice and the temporal characteristics of response patterns and absence pauses in the discussions of Sections 3 and 4.1, we conducted experiments to verify such ideas and possibly find other relationships between advice and those communication patterns.

System and task: Our experiments were conducted using the prototype system shown in Figure 1. The video captured by the USB camera and microphone was transmitted to a helper via Skype with QVGA quality. The voice of the helper was transmitted to the worker. We chose some cooking tasks in a kitchen wherein the worker, unaware of the location of kitchenware and seasoning, cooks an unfamiliar dish. Each task required approximately 30 min.

Workers and helpers: We gathered 12 cooking support samples. To give experimental data enough variety, we gathered eight participants of varying skill levels and formed worker helper pairs. Most participants performed two or more samples changing recipes and roles as worker and helper.

Communication logs and ground truth: We chose features depending on the possibility of automated feature detection; however, the ground truth data were manually collected because perfect accuracy cannot be expected initially.

Response patterns and absence pauses: Frequent patterns were extracted from the data for which partial results were already shown in Table 2. Response patterns shown in Table 3 were also used. Features in Table 1 were used to check for absence pauses.

6.2. Results

Since a helper can only give advice using speech, we only need to consider advising utterances for advice. The total number of advising utterances is 212. In the following discussion, we only discuss cases that have enough samples, because the number of advice utterances is not enough for some specific response patterns and absence pauses.

Response patterns: Figure 5 shows some typical examples of relationships between advice utterances and time gaps of response patterns. Figure 5(a) shows a simple case of <S:D W:P> in which a helper gives an explanation utterance and a worker replies to it. We observe that mutual information shows larger values around -1.5 second, which means advice is facilitated if two utterances overlap before the first utterance finishes. This negative time gap maintains a good tempo of conversation causing the helper to feel that the worker is listening. Conversely, a positive gap tends to suppress advice, especially over +1 second, which presumably causes the tempo to worsen.

Figure 5(b) shows a multimodal case of <S:D V:h> in which a worker stops his or her head motion, *i.e.*, gazing still at something, after a helper's explanation utterance. Mutual information has a peak at approximately +1.5 second, and advice is clearly promoted at that time implying that a head pause after an explanation often suggests that a worker does not fully understand the explanation and needs more advice. In contrast, in cases in which head motion stops before a helper finishes his or her speech, no clear relationships to advice were observed.

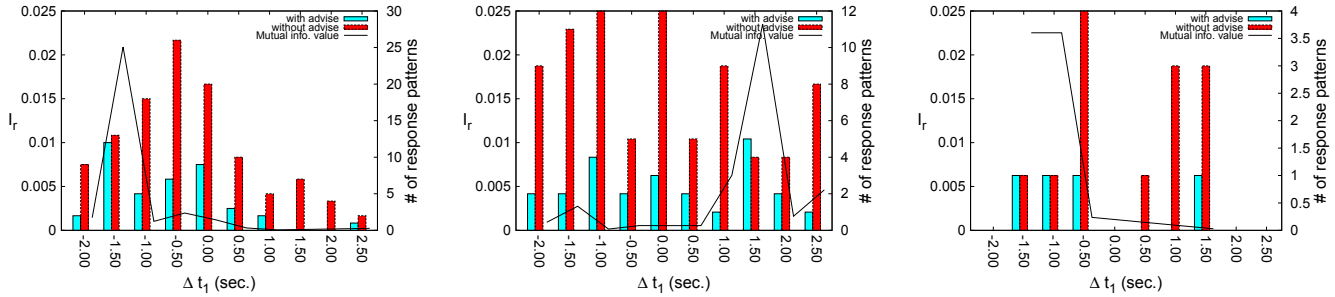
In other cases, such as the case of <S:D W:D> in Figure 5(c), there are not enough samples as shown characteristics. We therefore need to gather larger numbers of samples, which is left for future work.

In general, some response patterns have peaks of mutual information with a negative gap, for which we can assume that the negative gap, *i.e.*, the overlap, results in a good tempo and feeling of active communication. Others response patterns have peaks with a positive gap for which we can assume that the gap is the helper's need to observe and understand the situation and decide what advice to give, if any.

Absence Pauses: Figure 6 shows typical examples of relationships between advice utterances and the length of absence pauses. Figure 6(a) shows a simple pause of utterances,*i.e.*, $\{\text{S:D} \wedge \text{W:D}\}$. Here, pauses of 1 or 3 seconds promote advice, which matches our suspicion that silence implies a request for advice. More interestingly, we can see the difference between advice for the current situation and advice for the future; Figure 6(b) shows that advice for the current situation is more promoted by a longer pause at approximately 3 seconds, whereas Figure 6(c) shows that advice for the future is promoted by a short pause at approximately 1second. These results imply that a helper needs time to notice the necessity of advice for the current situation.

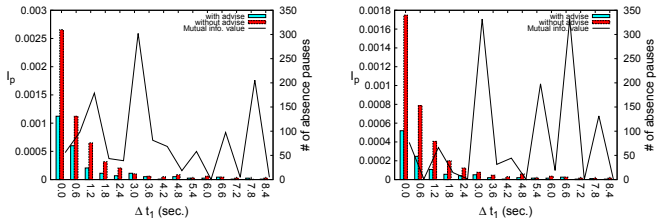
Finally, Figure 6(d) provides multimodal case {S:D, V:h}, which is a combination of a worker's head motion pause and an absence of helper explanation. We observe in the figure that mutual information is large at approximately 1 second and 3 seconds; however, co-occurrences are, significantly different across 3 seconds. Advice is suppressed if the pause is 2 seconds or less and promoted if it is 3 seconds or more, implying that a helper clearly feels the request for advice with a head motion pause longer than 3 seconds, whereas a pause with shorter time would be interpreted as another sign, for example, just checking the location or position of something.

The above results generally match our expectations on advice and communication patterns. An appropriate response with a good tempo and appropriate pauses tends to facilitate advice; otherwise, advice is suppressed. To confirm these findings, we need further statistical verification of these ideas,

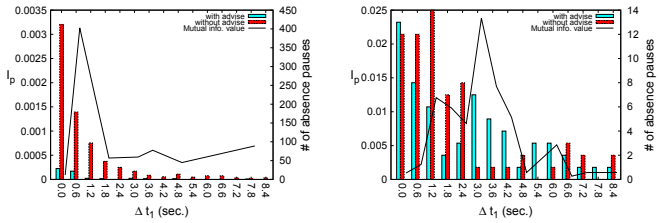


(a) <S:D,W:P> (b) <S:D,V:h> (c) <S:D,W:D>

Fig. 5. Relationships between advice utterances and time gaps of response patterns, with line showing mutual information and bars showing the number of actual occurrences of the pattern with or without advice utterances.)



(a) S:D ∧ W:D (b) S:D ∧ W:D, current



(c) S:D ∧ W:D, future (d) S:D ∧ V:h

Fig. 6. Relationships between advice utterances and absence pause, with lines showing mutual information and bars showing the actual number of occurrences of pauses with or without advice utterances.

which is left for future work.

7. CONCLUSION

In this study, we introduced the concept of distance-based cooking support and its realization via FPV communication. We considered the relationships between communication patterns that induce advice essential to distance-based cooking. Our experiments delineated that temporal characteristics of response patterns and absence pause patterns have tight relationships to the occurrences of advising utterances, *i.e.*, those interaction patterns induce or suppress advice depending on the time gap of a response pattern or the length of a pause.

For future work, we need to statistically verify the above findings with a wider variety of data and consider a variety of features that possibly affect advice. Moreover, we need to discuss design and implementation issues of a distance-based cooking support system that consistently maintains good communication conditions.

8. REFERENCES

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

[2] 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

[3] 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

[4] 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

[5] M. Billinghurst, S. Bee, J. Bowskill, H. Kato, “Asymmetries in Collaborative Wearable Interface”, The 3rd International Symposium on Wearable Computers, pp.133–140, 1999

[6] 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

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

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

[9] S. Kubota, Y. Nakamura, Y. Ohta. “Detecting Scenes of Attention from Personal View Records”, IAPR Workshop on Machine Vision and Applications, pp. 209–213, 2002.

[10] Y. Nakamura, T. Koizumi, K. Obata, K. Kondo, Y. Watanabe, “Behaviors and Communications in Working Support through First Person Vision Communication”, 11th IEEE International Conference on Ubiquitous Intelligence and Computing, 2014

[11] 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