

Communication Support Based on Interaction Analysis

Naomi Yamashita[†]

Abstract

This article describes the difficulties of human-computer interaction (HCI) research and the approach taken at NTT Communication Science Laboratories to alleviate them and establish findings for this research field. The main difficulty is that HCI research results are difficult to assess objectively owing to a lack of common yardsticks with other research fields.

1. Difficulties of HCI research

Human-computer interaction (HCI) research originates from the interactions between humans and information technology (IT), and it aims to develop technologies that bring greater comfort and convenience to our lives. HCI research is generally difficult; in particular, many researchers are currently struggling with evaluation methods. For example, if a system has been constructed with the expectation that it will improve the quality of interactions, it is surprisingly difficult to assess whether or not this expectation has been successfully met.

One of the reasons it is so difficult to evaluate an HCI-related technology is that no one has yet defined common yardsticks for interaction properties such as speed and correctness. Consequently, researchers themselves must find answers to questions such as what sort of state would indicate that an interaction has been improved, what would need to change and how before we can regard the objective as having been achieved, and what should things be compared with in order to justify the conclusion that there has been an improvement. The danger here is that different researchers can have very different views on these issues since there is no such thing as an absolutely correct value judgment.

As a result, each researcher uses his or her own value judgments as the basis for deciding when a

particular interaction has been improved from the user's viewpoint and ends up creating systems based on those personal judgments. In this way, each researcher creates systems according to his or her own individual preferences, which results in many different systems whose similarities and differences are difficult to judge.

When a system is completed, it is then evaluated by setting tasks. Even when a system can be defined as an *improvement* according to everyone's values, it is not easy to assess whether or not this is actually the case. For example, suppose that someone has made a system that can be used to make communication more enjoyable. Since there are no yardsticks that can be used to measure enjoyment objectively, this is typically done through questionnaires—where users try out the system and then say whether or not they enjoyed using it. However, researchers generally have strong opinions about their own systems and tend to have a strong belief in their usefulness. Consequently, even when users are surveyed by means of interviews or questionnaires, researchers are prone to phrase their questions so as to obtain results that favor their own systems. Even if there are no problems with the way the researchers phrase their questions, when users are asked to compare systems where a particular function is present and where it is not present, it is known that most of them will respond that the system with the function is better, regardless of whether or not they actually felt that it was better.

Through this sort of user testing, the expected results are obtained in most cases, and the researchers

[†] NTT Communication Science Laboratories
Soraku-gun, 619-0237 Japan

will report that their objective has been achieved. Unfortunately what normally happens next is that the system is not subjected to review and few further improvements are made.

My purpose here is not to belittle the diverse values of other researchers or the innovative systems that they have created. However, I feel that there is a problem whereby researchers can become so engrossed in the ingenuity of their own systems that they overlook or misapprehend the relevance of existing systems. Since only superficial evaluations are performed (in many cases), there is also a similar problem whereby improvements are overlooked or cannot be identified and are not connected with subsequent steps. For these reasons, the field of HCI unfortunately gives researchers in other fields the impression that it has gained little knowledge as a research field.

To allow HCI research to develop as a healthy research field, it is vital that evaluation and analysis are implemented in an accepted rational fashion and that systems are created with an awareness of the relevance of previous research, instead of superficial evaluations being performed on disparate systems created according to the vagaries of individual researchers. It is also necessary to have a common platform that all researchers can agree upon.

Of particular importance is the generalization of results obtained experimentally and by other means, and the accumulation of these results as knowledge. The results obtained from individual experiments are often limited and fragile, and it is not at all that rare for slight changes in experimental conditions and tasks to lead to different results. We thus need a way of generalizing experimental results by performing a comparative study with previous research so that they can be converted into meaningful knowledge. Through the accumulation of these results it should become possible to form a common platform by strengthening and storing knowledge covering the entire field of HCI research.

Below, I introduce the approach that my coworkers and I have taken so far. A characteristic of this approach is that we first observe and analyze how users use the system to understand how the system affects the actions of users, and we then consider the overall points of contact between humans and information technology. Since this approach makes it possible to suggest improvements on the basis of an understanding of the system and its role in collaboration, we can expect it to result in promising proposals.



Fig. 1. Two workers and two helpers at a remote site working on a mentoring task through the t-Room.

2. The t-Room video communication system

The t-Room is a video communication system that was developed at NTT Communication Science Laboratories [1]. It was designed for the purpose of reproducing live images of remote users and enabling them to engage in collaborative tasks as they move around a room. It can accommodate multiple users at a single location and transmit the positions in which remote users are standing almost exactly, wherever they are in the t-Room.

The t-Room hardware configuration is as follows. The system consists of multiple cameras arranged so as to surround the space where the users are, a number of vertical display devices, a central table consisting of two display devices, and a video camera that captures images of this table (Fig. 1). When multiple t-Rooms with the same configuration are set up in different locations and connected by audio and video communication links, it is possible for users in the t-Rooms at remote locations to share the same work space. A physical object on the table is also reproduced in two dimensions at the same position on the remote table, thereby making it possible to perform collaborative work while standing in positions around the table, not only for the users in the same t-Room but also for the remote users.

In a conventional system, the displays are connected and arranged so as to avoid any loss of video between the table display device and the tall display devices used to display the upper bodies of remote users, but in a t-Room system, a space is provided for people to enter between the table display and tall display devices, and the room is arranged so that people can move around freely in this space.

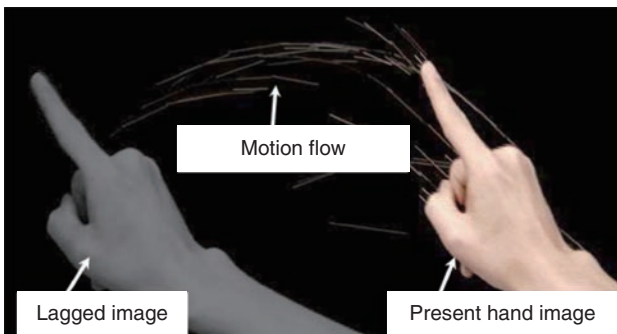


Fig. 2. Representation of t-Room.

3. Problem discovery through observation

Through our observations of preliminary tests, we have discovered a problem that can obstruct the gesture recognition of remote users in remote cooperative work using a t-Room. From related research and a more thorough analysis, we have identified two factors behind this problem. These two factors are not limited to the t-Room, but are also common to conventional remote table sharing systems.

- **Diverted attention:** When users are concentrating on work or paying attention to other things, they tend to miss the gestures of remote users. For example, in a remote instruction scenario, when workers are preoccupied with something in front of them, they will often fail to notice a gesture made by a remote user attempting to draw their attention to something else.
- **Occlusion:** When the table surface is covered by users' hands or by other objects, the video of a remote gesture displayed on the table surface may be buried underneath these objects and hidden from view. This phenomenon (occlusion) is observed particularly frequently in remote collaboration involving a real physical object, and it tends to have a detrimental effect on work efficiency. For example, in a remote instruction scenario, part of the gesture video of the remote user is buried by the object, and the worker may be unable to correctly apprehend the content of the instruction. A worker may sometimes attempt to recognize the instructor's gesture (pointing position) by lifting up the object, but by the time he or she has picked up the object, the gesture will most likely have finished and may no longer be visible.

4. Function proposed as a result of observation

In the above observation, we noticed that even if a user misses or is unable to see a remote gesture, he or she is often still aware of the gesture's existence and will take steps to search for the gesture. We therefore came up with a method called Remote Lag that delays the video of a remote user's gestures by a few hundred milliseconds when it is superimposed on the live video image (Fig. 2). That is, we thought that by making skillful use of delayed video of remote gestures, it should be possible to recover from states where gestures cannot be seen, even in cases where a remote gesture has been overlooked due to occlusion or a break in attention.

5. Experimental design

To evaluate the effects of Remote Lag in remote collaboration, we conducted a user study using the t-Room system. The research questions that we specifically wished to clarify in this experiment were as follows:

- (1) Does Remote Lag alleviate the problem of being unable to see remote gestures?
- (2) Does Remote Lag improve the efficiency of speech in remote instruction? For example, is there a reduction in the number of questions and confirmations uttered by the workers and in redundant instructions given by the instructors?
- (3) What sort of cognitive load does Remote Lag place on users?

A total of 56 participants took part in this experiment (28 in a t-Room in Kanagawa and 28 in a t-Room in Kyoto). The participants were split into 14 groups of 4 individuals: 2 in Kanagawa and 2 in Kyoto. Each group worked on a group task related to remote work instruction on a t-Room table. This task was performed twice: once without Remote Lag and once with Remote Lag.

The group task consisted of remote work instruction using real objects, which is a common subject for remote collaborative work using a table. This task involved achieving an objective by means of a specialist (instructor) issuing instructions to a novice (worker) at a remote location. In this experiment, we performed remote collaboration tasks with four people by creating a situation where two specialists with different specialist knowledge who were located in Kanagawa provided instructions to two novices working in Kyoto.

Specifically, this collaboration task consisted of

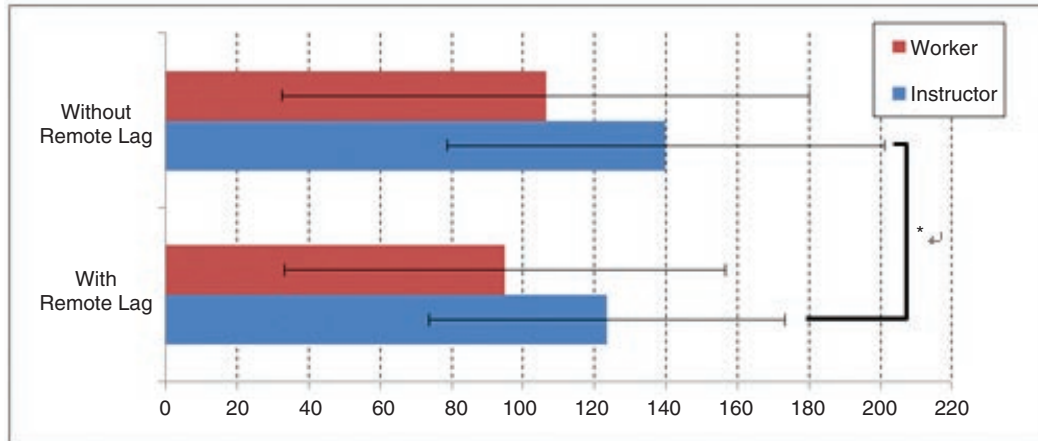


Fig. 3. Mean number of utterances per task for each experimental role.

assembling and arranging Lego blocks. The assembly of Lego blocks mainly involves having workers manipulate blocks by rotating and connecting them according to instructions from the instructors. This is often used in experiments involving collaboration on real objects. The requisite number of Lego pieces were placed on the table in the same t-Room as the workers, so that only the workers were able to manipulate them.

6. Evaluation of proposed function

To investigate in detail how the addition of Remote Lag affects the collaboration, we filmed the progress of the experiment by setting up video cameras in overhead positions at the entrances and opposite the entrances to both t-Rooms. First, we analyzed the video images to check whether or not the Remote Lag display alleviated the problem of being unable to see remote gestures. As a result, we discovered a number of situations in which Remote Lag was useful for alleviating the two problems mentioned above.

To verify the overall extent to which Remote Lag reduces redundant utterances in the remote collaboration instructions, we performed separate calculations of the average number of utterances made per task for each role and for both with and without Remote Lag (Fig. 3). As a result of comparing the number of utterances by using a paired t-test*, we found that com-

pared with the tasks performed without Remote Lag, the tests performed with Remote Lag required significantly fewer utterances from the instructors ($t[27]=2.74$, $p<.05$). Next, to investigate whether or not Remote Lag helps to reduce the number of utterances associated with questioning and confirmation, we counted the number of question and confirmation utterances and compared the results obtained with and without Remote Lag. Here, utterances associated with questions and confirmations were those made for the purpose of requesting or confirming information that was insufficient or unclear during the task. As a result of investigating the ratio of such utterances among all of the utterances made during the task, we found that the ratio of questioning or confirming utterances made by the workers was 23% when Remote Lag was used, which was less than the ratio of 30% when Remote Lag was not used. By using a paired t-test, we found that the tasks conducted with Remote Lag were accomplished with significantly fewer question and confirmation utterances from the workers than ones without Remote Lag ($t[27]=2.71$, $p<.05$).

From these results, it appears that the Remote Lag display improves the visibility of gestures by the instructors and makes it easier for the workers to understand the instructions being given to them by the instructors, thereby reducing the frequency of questions and confirmations.

Finally, we evaluated the workload imposed on the workers by measuring the subjective workload due to the Remote Lag display. To perform these measurements, we used the NASA TLX (task load index)

* t-test: A general term for statistical hypothesis tests that make use of the fact that statistical quantities follow Student's t-distribution when the null hypothesis is assumed to be correct.

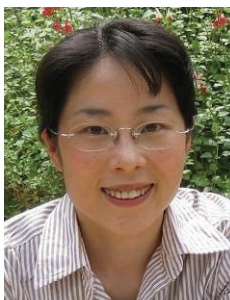
method. NASA TLX evaluates the overall workload in terms of six indicators (mental demand, physical demand, temporal demand, work performance, effort, and frustration). The overall workload is calculated in the range from 0 to 100 (lower scores are better) based on the weighted average of each indicator. This method is widely used for evaluating all manner of interfaces between humans and digital equipment as a method for evaluating workloads subjectively. By performing a corresponding t-test, we showed that the time pressure felt by the workers was significantly lower with Remote Lag than without Remote Lag ($t[27]=1.72$, $p=.097$). We can infer that this was because there is no need for extra effort in tracking realtime remote gestures because the delayed video provided by Remote Lag makes it possible to follow realtime gestures even if the remote gestures of the instructors and workers are missed.

7. Summary

In this article, I discussed the importance of making functional proposals based on an understanding of human actions, introducing our t-Room research as a case in point. In the future, I will continue making system designs and functional proposals on the basis of an understanding of human behavior, which I hope will be of use to others.

Reference

- [1] K. Hirata, Y. Harada, T. Takada, S. Aoyagi, Y. Shirai, N. Yamashita, and J. Yamato, "The t-Room—Toward the Future Phone," *NTT Technical Review*, Vol. 4, No. 12, 2006.
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200612026.pdf>



Naomi Yamashita

Communication Environment Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

She received the B.Eng. and M.Eng. degrees in applied mathematics and physics and the Ph.D. degree in informatics from Kyoto University in 1999, 2001, and 2006, respectively. She joined NTT in 2001. Her primary interests lie in the areas of computer-supported cooperative work and computer mediated communication. Using a combination of quantitative and qualitative research methods, she aims to uncover the nature of human discourse/interaction and propose guidelines for designing new communication technologies.
