Establishing communication in an artificial interaction environment

We investigated how humans establish communication in a novel environment using an artificial cooperative task. Subjects are asked to solve a kind of 8-puzzle game cooperatively. One subject (director) can see the design of the puzzle and instructs how to move a tile. The other (operator) cannot see the design but can manipulate the puzzle. The director can communicate his intention only through his body movement.

We conducted two experiments. In Experiment I, the operator role is played by the experimenter, who follows the pre-determined algorithms for establishing communication. The director role is played by a recruited subject. In Experiment II, both the director and the operator roles are played by recruited subjects. In both the experiments, simple languages for communication were developed between the players, but the strategies adopted were quite different. The results correspond to two approaches to future Human-Agent communication — human-controlled communication and mind-reading communication. We should re-consider the role of mind-reading communication in designing Human-Agent interface.

I. INTRODUCTION

Humans can communicate their intentions even if there are no languages shared among them. They adopt whatever media available on the spot, and use them to communicate their intentions. We want to adopt this communication mechanism in Human-Agent (machine, artificial object, or robot) Interactions. With this mechanism we can interact with entirely novel machines more comfortably. The above mechanism, we assume, uses a human-specific faculty — mind-reading. Mind-reading is what makes human communication flexible, unbounded, and applicable in any situations[1][2]. Hence our goal can be paraphrased as “realizing a mind-reading communication in Human-Agent Interactions.”

For many years, researches from various fields have been addressing the above problem. Taking an intentional stance is a starting point for mind-reading communication. Dennett[3] discussed how a human takes an intentional stance to various non-human objects. Psychologically we seem to have an innate tendency to take an intentional stance to objects having certain characteristics[4][5]. There are also many researches relating to the emergence of communications[6][7][8][9][10]. These researches have made clear some aspects of mind-reading mechanism, even if their researches did not mention it explicitly.

In the previous paper, we investigated how humans use a color (hue) signal, or a sound (monotone sound) signal as a communication media which have no pre-defined meanings in a cooperative game environment[11]. The difference of our research to the former researches is that in our experiment the task goal is to clear the game or to get a high game points. Here communication is just a tool for achieving that goal. The purpose of this setup is that important features of mind-reading can only be observed in the unconscious communicative interactions.

The results of our experiments are that (1) humans can somehow communicate their intentions using given media, (2) what is achieved in communication is just sharing of minimum signals to solve the given task.

In this paper, we employ a body movement to solve a cooperative task. The difference of body movement from our previous media lies in the freedom and the ambiguity the body movement necessarily has. How this freedom and ambiguity modify the mode of interaction is one of our aim in this experiment.

II. HUMAN-CONTROLLED VS. MIND-READING INTERFACES

Here we want to make clear the difference of human-controlled interfaces and mind-reading interfaces. There have been two approaches to designing human-agent (machine, artificial object, or robot) interfaces. One is human-controlled interfaces, in which the algorithm of the interface agent is pre-determined (may be adaptable), and a human (user) is expected to understand the algorithm and controls the whole interactions. For a good performance, the algorithm must be simple, suited to human thinking, and flexible enough to be able to adapt to individual differences. Even in the adaptation process, the important point is that a human is expected to take control of the system — a human must ultimately determine everything in his own responsibility. Here, the role of agents in the interface design is necessarily passive.

The other is a mind-reading interface, which works like human communication expounded by Sperber and Wilson[12]. In human communication, participants are regarded equal in ability and in responsibility. Here, a strategy that a human should understand the algorithm of the interface
agent does not work — the agent may be too sophisticated to be understood at an algorithmic level. Consequently, a human must adopt a so-called “intentional stance,” inferring the future behaviors in terms of an intention or a goal of the agent. Here we are forced to assume rationality of the agent (sometimes with no sound grounds).

It is often said that in designing an interface agent, sophisticated algorithms should and can be avoided. It may partly be true. But we humans do not want others (either humans or machines) to understand us at an algorithmic level. The fact that others can predict our behavior precisely (= at an algorithmic level) is not very comfortable. In the future, with the same reason, the designer might not want users to understand his agent at the algorithmic level. Agents need some secrets for behaving cleverly. This is especially true in situations where there are conflicts of interests between users and interface agents.

Fortunately (or unfortunately), current technology still does not allow the design of such a high-ability agent. But we should prepare for the age when the sophistication of the machine is such that we can no longer rely on the assumption that machines can be predictable.

III. THE COOPERATIVE 8 PUZZLE GAME

The 8 puzzle is a game with 9 positions or slots, as shown in Fig.1. Each of 8 slots have a tile, and the remaining slot without a tile, is called “gap.” Any tile adjacent to the gap can be moved onto the gap, creating a new gap at the previous position of the moved tile. Each tile has a part of a picture on it. The aim of the game is to rearrange the tile to complete the picture on the tiles. In ordinary 8 puzzles, tiles can be translated up/down/right/left-ward, but in our cooperative 8 puzzles, tiles can also be rotated at 90 degrees to complete the picture. The pictures on the tiles are selected from Japanese “anime” characters, popular among Japanese youngsters.

The operator can move a tile but cannot see the picture on the director’s tiles. On the tiles of the operator are drawn just numbers, instead of the pictures. The move of tiles are synchronized, i.e., the operator’s move of a tile is immediately reflected in the director’s view.

A. Director and Kinect sensor device

The director instructs the operator how to move the tile using his body movement. The body movement is analyzed by Kinect. Kinect, shown in Fig.2, is a motion sensing input device developed by Microsoft for a Xbox 360 video game machine.

Kinect has an optical camera, a depth sensor, and analyzes the body configuration at the frame rate of 30 Hz. It outputs the analyzed body configuration as a skeleton graph consisting of 15 joint points, as shown in Fig.3. The director making a pose in front of Kinect is shown in Fig.4.
B. Operator

On the operator’s tile the number is drawn instead of a picture. He can see how he has moved the tiles, but the goal is to complete the picture on the director’s tiles, not to sequence the numbers on his tiles. Only after the game is cleared, or game is over, the picture on the director’s tiles is also shown on the operator’s tiles.

An operator can manipulate a tile with a three-button mouse through the following interface:

- One of the 9 slots can be focused, showing it a little brighter than other tiles. Only a focused tile can be moved (translated/rotated). If the clicked tile is not focused, a focus moves to the clicked tile and become focused. In this case, all the three mouse-button have the same effect.
- If the clicked tile is already focused, one of the following will happen depending on which mouse-button is clicked.
  - (a) Right button: the tile rotates 90 degrees rightward.
  - (b) Middle button: the tile rotates 90 degrees leftward.
  - (c) Left button: If the clicked tile is adjacent to the gap, it moves onto the gap, otherwise nothing happens.

On the operator’s screen is shown the skeleton graph of the director. The operator has to play the game consulting only the behavior of the skeleton graph. The skeleton graph shown to the operator is a mirror image, or a image from behind, so that the right/left in the skeleton graph correspond to the right/left in 8-puzzle screen.

IV. Experiment

We conducted two experiments. In Experiment I, a director is a subject recruited in our university, and an operator is an experimenter in our research group. In Experiment II, both an operator and a director are recruited subjects. First we explain the experimental setup.

A. Experiment I

In Experiment I, directors are student subjects of age 20-25. At first a subject was asked to read a two-page manual of the game explaining the aim, the setting, and how to play it. The subject was instructed that the operator is an experimenter in our research group. Next he was asked to take a pose in front of Kinect device, and saw how Kinect outputs the skeleton graph of himself. After he was sure what information is sent to the operator, he was asked to start the game. When Kinect recognizes the calibration pose of the subject, the game automatically starts and the first puzzle is shown.

A game consists of four puzzles, with increasing degrees of difficulty. The puzzle shown in Fig.1 is the view of the first one at the start of the game. When either the puzzle is completed (game clear), or 10 minutes have passed from the start (game-over), the next puzzle automatically starts. The skeleton graph of all the frames, and the operations taken by the operator were recorded with timestamps for later analysis.

The strategy of the operator (experimenter) is pre-determined, and is roughly as follows:

- If director indicate any direction (i.e right/left/up/downward) by the body movement, then the operator first tries to move the focus in that direction. This is true even if currently focused tile can be moved in the indicated direction, i.e., a focus moves to the gap (empty slot). If the director seems not satisfied by this (focus move) operation, then the operator next tries to move the tile. “Not satisfied” may be inferred by the repetition of the same action (or action sequences), some negation gestures, etc. Once this is observed, strategy is switched, and the operator always moves the tile if there is a tile movable to the indicated direction. Consequently, the focus cannot be moved through the gap hereafter. If there is no tile movable to the indicated direction, a focus is moved as before.
- If a new type of gesture for indicating a direction is observed, assign the operation “move a tile” to this new gesture. A rotation gesture always rotate the focused tile (90 degrees), for the gestured rotation direction.

Ten subjects played the game. The total time of the game for each subject is from 20 min. to 40 min. After the game each subject was asked to answer to the questionnaire about his “strategy” and intended meanings of his gestures. After the experiment, a fixed sum of money was paid as a reward.

B. Experiment II

In Experiment II, both a director and an operator are student subjects of age 20-25. The instruction given to the director is the same as Experiment I, except the instruction that the operator is also a recruited subject. The operator was also asked to read the same game manual as given to the director. Next the operator was conducted to the operator PC, and was given the explanation of operation of the puzzle. These explanations were given separately for a director and an operator. They could not meet before the end of the game.

Ten pairs of subjects played the game. The role of director and operator are randomly assigned on the spot. The first five pairs are treated as above(Ia). As to last five pairs(Ib), 10 minutes of practice time, during which the subjects played the 8-puzzle game in solitaire mode, is given before the experiment. This is to familiarize the subjects with our 8-puzzle game. Other settings are the same with Experiment I.

V. Experimental results

A. Experiment I

Through a number of try and error interactions, most subjects learned how to move a focus and a tile. Two types of communication strategies are observed.

Type A: Only one type of direction gesture is used. As tiles and a focus are both movable, the ambiguity of the target is resolved according to the following rule: If a tile can be moved, then move the tile, otherwise move a focus.

Type B: Two types of direction gesture are used. That is, one is used for moving a tile, and the other for moving a focus. Eight out of ten subjects adopted type A strategy, and the remaining two subjects adopted type B strategy.
Most subjects acquired a communication strategy in 5-10 trial-and-error steps. As subjects with type A strategy cannot move a focus directly through the gap, they also learned to move a focus avoiding the gap. This is rather clumsy, but no subjects starting from type A strategy was observed switching to more efficient type B strategy.

The results are summarized in Table I. Type in the Table means communication strategy type (A or B). The number in the table is the number of steps for solving puzzle 1-4. Italics means that the puzzle could not be solved (time over at the steps).

The progress of the games for puzzle 1 (the first puzzle) are shown in Fig. 5. In the Figure, the vertical axis is a cost of the game state (minimum steps necessary to complete the puzzle from this state, without counting focus moves). The horizontal axis is a number of steps taken by the operator, mostly corresponding one-to-one to the director’s instructions. The graph starts at 12 (leftmost), meaning that the puzzle can be solved with only 12 translation/rotations of tiles. The bottom line (cost zero) means that the puzzle is solved.

Subject 5 and 7 randomized the puzzle at an early stage of the game, making the puzzle substantially more difficult than the start. Seven out of ten subjects solved the first puzzle in less than 70 steps. The difference in performance among subjects seems to be mostly due to a communication failure, but to a skill of the director to solve the 8 puzzle itself.

When designing a game for our experiment, we expected that most subjects can solve a 8 puzzle easily. Unfortunately this is not true for every subjects. Some subjects continued to move tiles for many steps without much improvement in the game state (subject No.5, No.7).

On moving a tile the subject No.3 and 5 used their hands and bodies to indicate the location of the target tile, giving the operator redundant information, for the target tile can be fixed by the direction of the movement. They never tried to rotate a non-focused tile by this strategy, which the operator would not obey (even if he understood the intention). The subject No.6 actually tried it, and failed to get an intended results. He quickly understood the operator’s algorithm and followed it. This is inevitable response, for the operator strictly followed the algorithm explained in IV. We designed the operator behavior as a sort of a user-adaptable interface.

As the operator is an experimenter in Experiment I, the same experimenter examined the video replay of the skeleton graph, segmented a meaningful movement and classified it into several movement types (move the left hand upward etc.) without consulting the puzzle screen. Next a cross correlation table for classified movement types and the operations taken in the experiment following the movements is calculated. If the mapping from a movement type to an operation is one-to-one, for each movement type only one operation should be observed. The mismatch between idealized one-to-one mapping table, and the actual table obtained must be caused by the following two reasons:

1. The meaning of the movement changed through the progress of the experiment.
2. The classification of movement types is difficult without consulting the puzzle screen. At first we expected that the reason (1) is dominant, and the mismatch rate would decrease quickly. Actually, the data shows that the mismatch rate is rather flat over the course of the time, suggesting that the reason (2) is dominant.

Average mismatch rates for each subject pairs are shown in Table II. The total average for 10 subjects is 10.2%. The mismatch rate seems very high, but large part of the mismatch was appropriately understood by the operator using a puzzle state as a context.

### B. Experiment II

In Experiment II, both the director role and the operator role were played by recruited subjects. The operator’s behavior differed greatly from the pre-defined algorithm of Experiment I, and consequently made director’s behaviors also quite different.
The results are summarized in Table III. The progress of the games for puzzle 1 (first puzzle) are shown in Fig.6. The meaning of each items are the same as explained in Experiment I. No.1-5 correspond to pairs without, and No.6-10 to pairs with 8-puzzle practice in a solitaire mode. The 10-minute practice before the experiment does not seem to have much effect on their performance.

All the operator seemingly preferred to move a tile rather than to move a focus. For any indication of direction from the director, all the operators moved a tile (if it is movable for that direction) irrespective of the location of the focused tile. As only a focused tile can be moved, the operators first moved a focus to the movable tile, and then moved it onto the gap.

At first some director subjects were surprised at the operator’s behavior. Actually, some directors at first tried to move a focus by some gesture, but got a tile moved, and seemed perplexed. But ultimately, most directors understood the algorithm, and seemed satisfied with this strategy — i.e., the strategy to move a tile directly without regard to a focus.

This strategy is simple and intuitive for translation of a tile, but it needs another strategy for rotating a tile, for the target tile must be specified for rotation of a tile. Some pairs rotated a tile after it is translated, for after the translation the moved tile remains focused (type A). Some other pairs directly indicated the target tile by their hands and/or body positions (type B). Others first used only a translation of a tile, and after all the tiles are relocated in the correct slots, they started to rotate the tiles (type C). In one pair, the operator sequentially moved a focus to slots around the gap. The director waited until a target tile was focused (type D). These strategy types are shown in Table III.

Average steps taken for puzzle 1-4 each for Experiment I, IIa (pair No.1-5), and IIb (pair No.6-10) are shown in Fig.7. There seems to be no clear tendency in the graph. This is probably because a few “anomaly pairs” made a large contribution to the average (see Table I and III). This is also observed in Fig.6, where some pairs used nearly 6 times more steps compared to good pairs.

### Table III

**Summary of Experiment II.**

<table>
<thead>
<tr>
<th>No.</th>
<th>type</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>total</th>
<th>cleared</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>42</td>
<td>204</td>
<td>165</td>
<td>35</td>
<td>446</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>177</td>
<td>27</td>
<td>227</td>
<td>270</td>
<td>635</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>43</td>
<td>29</td>
<td>80</td>
<td>157</td>
<td>309</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>114</td>
<td>34</td>
<td>103</td>
<td>164</td>
<td>389</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>D</td>
<td>71</td>
<td>99</td>
<td>203</td>
<td>261</td>
<td>634</td>
<td>3</td>
</tr>
<tr>
<td>average</td>
<td>179.4</td>
<td>78.6</td>
<td>154.4</td>
<td>160.2</td>
<td>482.6</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>149</td>
<td>352</td>
<td>439</td>
<td>217</td>
<td>1157</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>293</td>
<td>74</td>
<td>132</td>
<td>49</td>
<td>548</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>191</td>
<td>83</td>
<td>34</td>
<td>160</td>
<td>468</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>136</td>
<td>33</td>
<td>49</td>
<td>179</td>
<td>397</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>C</td>
<td>313</td>
<td>415</td>
<td>194</td>
<td>258</td>
<td>1180</td>
<td>4</td>
</tr>
<tr>
<td>average</td>
<td>216.4</td>
<td>191.4</td>
<td>169.6</td>
<td>172.6</td>
<td>750.0</td>
<td>3.4</td>
<td></td>
</tr>
</tbody>
</table>

### VI. Discussions

**A. Different subject behaviors in Experiment I and II**

We obtained in Experiment I, and II two quite different behaviors of subjects. In Experiment I, directors inferred the operator’s algorithm and adapted to it. The operator adapted his algorithm to the director’s gesture, making a gesture-command mapping on the spot. This is so-called user-adaptation process. We expected that our algorithm for the operator is one of the best (in performance) and one of the most intuitive of operator strategies.

The system has some degrees of freedom, or parameters to be learned, and used them to adapt to the user. On the other hand, users infer the algorithm and fix unknown parameters by try and error method. This is similar to what Chomsky thought by Universal Grammar. Core grammar are pre-determined, but users can adjust various parameters to satisfy their needs. In Experiment I, we assumed that kind of adaptation process, and designed the operator’s algorithm explained in IV.

In Experiment II, operator subjects behaved quite differently from our pre-determined algorithms. All the pairs takes ultimately “Move a tile directly” strategy, ignoring a focus. As an afterthought, it seems simpler, and intuitive. But the weak point is that the rotation of tiles cannot be handled in this strategy. As we described in V, subject pairs invented various strategies to cope with this problem.
The result of Experiment I and Experiment II each represents a different kind of interface models, user-adaptation interfaces, and mind-reading interfaces. For better communication between humans and machines (agents), user-adaptation technique is widely investigated, and actually adopted in many commercial systems. On the other hand, mind-reading communication has so far not been employed in the interface design. To begin with, what is an essential part of mind-reading communication is not clear. It seems just an assembly of ad hoc behaviors invented on the spot, just an application of human general intelligence to communication. We must know the mechanism of mind-reading more deeply before applying it to the interface design. The aim of Experiment II is to obtain knowledge on the role of mind-reading in establishing communication channels.

In Experiment I, we prepared what we think is the best strategy for the operator. In Experiment II, we did not give any hints on what strategies should be taken by an operator. What we asked is just to solve the puzzle cooperatively. Subject pairs tried to communicate using mind-reading faculty, and invented communication protocols on the spot. Operator subject could employ a similar algorithm as we adopted, but they did not. They prefer ad hoc strategy based on mind-reading rather than simple and efficient adaptive algorithms. It seems that mind-reading communication is a natural strategy for any human communication task, even if it is very artificial one.

B. The necessity of mind-reading communication

We have to admit that mind-reading communication and mind-reading interface are not always easy to use, and often uncomfortable. Humans often likes to use vending machines rather than to use a human-attended shops. Using a mind-reading faculty demands, (often unconsciously) a certain amount of mental costs. It is often much easier to adjust himself to machines rather than to use mind-reading faculty to buy everyday groceries.

Then why do we investigate mind-reading communication? Our answer is that there are some situations where only mind-reading communication can achieve the goal. Features of mind-reading is in improvisation — creating everything on the spot. It may not be of best performance, but never fails completely. If there is other efficient methods to achieve a goal, it may be better. Mind-reading communication is the last resort where other means all failed.

Other important point for mind-reading is that it may be the only means for communication between agents (humans and/or machines) with conflicting interests. For, in such situations, humans (and intelligent agents) can and should tell a lie if it is profitable to them. A lie is not necessarily of egoistic nature. It may be for the benefit of all the participants (so called win-win situation). We can encourage our friend when the situation seems not very encouraging even to the speaker. If the hearer takes it for meaning the situation encouraging, it is wrong. The correct meaning is that the speaker want the hearer encouraged. If the surface meaning cannot be relied upon, mind-reading communication must enter the stage.

Human may often feel happier if he can be freed from the burden of always reading others mind. It is argued that McDonald system is an successful example of employing an non-mind-reading communication between salesperson and customers[13]. But if we abandon the mind-reading communication entirely, our society cannot be maintained. Mind-reading faculty of humans is not a blessing, but just a necessity. We cannot live without it.

VII. Conclusions

We investigated how humans communicate their intentions using body movement using a cooperative 8-puzzle game. A director instructs the operator how to move a tile through his body movement. An operator manipulates the puzzle from the skeleton graph of the director. Two kind of experiment were conducted. In Experiment I, an operator is an experimenter following a user-adaptive algorithm and directors are recruited subjects. In Experiment II, both operators and directors are recruited subjects. In Experiment I, subject directors accepted the operator’s algorithm and quickly learned how to communicate his intentions. In Experiment II, subjects invented various strategies on the spot based on a mind-reading communication.

REFERENCES