Hand-over Motion Model Based on Timing between Voice Utterances and Release Motions of Humans

Mitsuru Jindai
Graduate School of Science and Engineering
University of Toyama
3190, Gofuku, Toyama, Toyama
930-8555, JAPAN

Shunsuke Ota
Graduate School of Systems Engineering
Okayama Prefectural University
111 Kuboki, Soja, Okayama
719-1197 JAPAN

Yusuke Ikemoto
and Tohru Sasaki
Graduate School of Science and Engineering
University of Toyama
3190, Gofuku, Toyama, Toyama
930-8555, JAPAN

Abstract—As society ages, it is expected that robots will play important roles in the context of social welfare and service in the home and office. These robots must be able to coexist and cooperate with humans. One of the most important tasks that these robots may perform is handing over objects to a human. Many people hand over objects to each other accompanied by voice greetings; the hand-over motion is one of the embodied interactions. In the embodied interaction between a human and a robot, the timing between motions and voices affects motion characteristics that are preferred by humans. Therefore, in this paper, we propose a hand-over motion model based on the timing between the voice utterances and the release motions of humans. This model generates a motion in which a robot hands over an object to a human accompanied by a voice greeting. In this model, a hand-over motion is generated based on the analyses of human hand-over motions; in particular, the timing between the voice utterances and the release motions of humans is analyzed. Then, the model generates the release motion of a robot in response to a voice utterance from a human. Furthermore, a hand-over robot system that uses the proposed hand-over motion model is developed. The effectiveness of the proposed model is demonstrated by sensory evaluation using the proposed hand-over robot system.

Keywords—Handshake, Robot-Human System, Embodied Interaction, Human Emotion

I. INTRODUCTION

As society ages, it is expected that robots will play important roles in the context of social welfare and service in the home and office [1]. These robots must be able to coexist and cooperate with humans. One of the most important tasks that these robots may perform is handing over objects to a human, and this motion of the robot must be not only physically safe but also psychologically comfortable.

The hand-over motion as it pertains to human emotions has been discussed [2], [3], [4], [5], and such research has proposed models and motion characteristics that are preferred by humans. These models and motion characteristics are based on hand-over motions without an accompanying voice greeting. However, in hand-over motions between humans, such behavior is often accompanied by voice greetings.

Nevertheless, hand-over motion is an embodied interaction. In human embodied interactions, humans synchronize their embodied rhythms as an embodied entrainment. It is important that the embodied rhythms are synchronized by voices and motions in embodied interactions between a human and a robot [6]. It has been discussed that in handshake motions between a human and a robot, the timing between hand motions and voice greetings is important [7], [8]. The timing affects motion characteristics that are preferred by humans. The hand-over motion is similar to the handshake motion as an approaching motion of hands. Thus, in the hand-over motion between a human and a robot, a robot should generate a hand-over motion in which the timing between hand motions and human voices is considered.

Therefore, in this paper, we propose a hand-over motion model based on the timing between the voice utterances and the release motions of humans. This model generates a motion in which a robot hands over an object to a human accompanied by a voice greeting. In this model, a hand-over motion is generated based on the analyses of human hand-over motions; in particular, the timing between the voice utterances and the release motions of humans is analyzed. Then, the model generates the release motion of a robot in response to a voice utterance from a human. Furthermore, a hand-over robot system that uses the proposed hand-over motion model is developed. The effectiveness of the proposed model is demonstrated by sensory evaluation using the proposed hand-over robot system.

II. ANALYSIS OF HUMAN HAND-OVER MOTION

A. Hand-over experiment

Human hand-over motions were measured using a threedimensional motion capture system (VICON). Five reference markers (right hand, right wrist, right elbow, right shoulder, and left shoulder) were attached to the subjects. These positions were measured using ten cameras. The accuracy of this system was ±1 mm for a sampling rate of 120 Hz.

The experimental setup is shown in Fig.1. In this experiment, the handing side and the receiving side were both predetermined. The object to be handed over was a cylinder with a diameter of 40 mm, height of 330 mm, and mass of 630 g. The giver (i.e., the person hand-over the object) and receiver stood face-to-face at a distance of 1000 mm. For initial postures, the giver bent his or her elbow naturally to a position where the object was held upright. The receiver stretched his or her right arm downward. The giver grasped the lower end of
the object to be delivered in his or her hand and kept the object upright throughout the hand-over motion. The experiment was performed for a hand-over motion accompanied by a voice greeting. In this motion, the giver started to move his or her hand at an arbitrary time, and the receiver moved his or her hand in response to this motion. Each pair handed over the object 10 times, with the object returned to the giver after each repetition. The subjects were 30 healthy students (15 pairs) aged between 20 and 24. All the subjects were right-handed.

### B. Angular velocities of human joints

Fig. 2 shows an example of the velocity profiles for the giver when the hand is extended toward the receiver to deliver the object. The figure shows the rotation velocities of the elbow and shoulder on the Y- and Z-axes. It can be seen in the figure that the angular velocity profiles of the elbow and shoulder form a bell-shaped pattern having one peak. The average movement time, average maximum velocity, and average position of maximum velocity at movement time (peak position) are provided in Table I. The table indicates that the peak position of the elbow motion is around 50%. In the elbow motion, the velocity pattern becomes a symmetrical bell-shaped profile. However, the peak positions of the shoulder motion on the Y- and Z-axes are 39% and 33%, respectively. Therefore, in shoulder motion, there are peak positions in the first half of the motion.

### C. Timing between voice utterance and release motion

Fig. 3 shows an example of the pressure of the finger and the voice utterance of a giver and a receiver. From this figure, a release motion is performed after a voice utterance from the giver side. In the same manner, on the receiver side, a gripping motion is performed after a voice utterance from the receiver. In particular, the giver performs the release motion after the voice utterance of the receiver. From this result, it is clear that the giver performs the release motion according to the voice utterance of the receiver. In embodied interactions, humans synchronize their embodied rhythms by voice greetings. The hand-over motion is one of the embodied interactions. Therefore, it is evident that humans synchronize their embodied rhythms by their voices, and perform hand-over motions at precise timing.

![Fig. 2. Angular velocity of joint](image)

![Fig. 3. Timings between gripping force and voice utterance](image)

### III. Hand-over Motion Model

A hand-over motion model is proposed in order to generate a hand-over motion with humans. The model generates a motion in which a robot hands over an object to a human accompanied by a voice greeting. The model involves an extending motion, a following motion, a switching control,
and a release motion. The robot trajectory is estimated by considering the extending and following motions. The extending motion generates the first half of the hand-over motion. The following motion generates the latter half of the hand-over motion so that the hand-over can be performed at the position of the hand extended by the human. The switching control changes from extending to following motions to generate a smooth hand-over motion. The release motion is performed according to the voice utterance from the human.

A. Generation of hand-over motion

1) Extending motion: When a robot approaches a human, it is necessary for the robot’s hand to have an acceptable trajectory in which the velocity and acceleration change smoothly. The minimum jerk model [9], which minimizes equation (1), can accurately reproduce the point-to-point motion of human hands. In this equation, \((x, y)\) is the position of a human hand and \(T\) is the movement time. In this model, the acceleration has smooth changes, and the velocity pattern has a symmetrical bell-shaped profile, as shown in Fig.5. Thus, the preferred trajectory of the robot is expected to be generated using the minimum jerk model.

\[
C = \frac{1}{2} \int_0^{\frac{T}{T_f}} \left( \frac{d^3x}{dt^3} \right)^2 + \left( \frac{d^3y}{dt^3} \right)^2 \, dt \tag{1}
\]

![Fig. 5. Velocity pattern of Minimum jerk model](image)

Therefore, in the hand-over motion model, the hand-over motion of the robot is generated according to the minimum jerk model. In the minimum jerk model, the robot velocity \(V(t)\) is calculated from the maximum velocity \(V_{\text{max}}\) and the movement time \(T_f\) using equation (2). \(V_{\text{max}}\) is calculated using equation (3). Here, \(r_0\) and \(r_f\) are the initial and target positions, respectively. The velocity pattern of the minimum jerk model has a symmetrical bell-shaped profile. Thus, it is difficult to adjust the peak position for the velocity pattern. Therefore, on the basis of the minimum jerk model, a bell-shaped velocity pattern is proposed in which the peak position can be adjusted.

\[
V(t) = V_{\text{max}} \cdot \frac{16}{T_f^2} \left( t^4 - 2 \cdot T_f \cdot t^3 + T_f^2 \cdot t^2 \right) \tag{2}
\]

\[
V_{\text{max}} = \frac{15}{8 \cdot T_f} \left( r_f - r_0 \right) \tag{3}
\]

The movement time, maximum velocity, and peak position of the targeted bell-shaped velocity pattern are denoted by \(T_{\text{max}}, V_{\text{max}}\) and \(P\), respectively. \(P\) is the value of the peak position when \(T_{\text{max}}\) is normalized to 100 %. First, a minimum jerk model is generated (Fig.6) in which the movement time is \(2 \cdot T_{\text{max}}\). and the maximum velocity is \(V_{\text{max}}\). The maximum velocity \(V_{\text{max}}\) is calculated using equation (3). Second, a minimum jerk model in which the movement time is \(2 \cdot T_{\text{max}} \cdot (1 - P)\), and the maximum velocity is \(V_{\text{max}}\), is generated in the same manner. Finally, the first half of the sequence of the first minimum jerk model and the latter half of the sequence of the second minimum jerk model are combined to achieve the target bell-shaped velocity pattern. In this bell-shaped velocity pattern, the position, velocity, and acceleration changes are smooth and the peak position is adjustable. Therefore, to produce a hand-over motion similar to that of humans, the rotational motions of the robot’s shoulder are generated using this bell-shaped velocity pattern.

![Fig. 6. Generation of bell-shaped velocity pattern](image)

2) Following motion: The extending motion alone does not allow an object to follow the hand position of the human, so the hand-over motion cannot be performed at the position of the hand extended by the human. Therefore, a following motion that follows the position of the human hand is used along with the extending motion. The following motion is generated such that the object arrives at the human hand position.
3) Switching Control: Precisely combining the extending and following motions enables the robot to perform a hand-over motion at the position of the hand extended by the human. Depending on the distance by which the human hand moves and the distance of the object from the human hand, the robot switches from the extending motion to the following motion.

In the case of this model, movement is mostly due to the extending motion. In the latter half of the movement, the robot switches to the following motion to move an object to the human hand position and to deliver the object to the human hand. This results in a hand-over motion at the position of the human hand through a natural extending motion.

4) Release motion: In the hand-over motion of humans, many humans release an object after hearing a voice utterance from the receiver. It is evident that humans synchronize their embodied rhythms by voice. Therefore, in this model, the robot releases an object after hearing a human voice.

B. Motion generated using proposed model

The trajectory of the motion performed by the subject on the giver side was compared to the trajectory generated by the proposed hand-over motion model to demonstrate the effectiveness of the proposed model. Here, the initial position, target position, and movement time of the trajectory of the robot’s hand movement were matched with those of the movement of the human hand.

Figs.7-9 show the trajectory generated by the proposed model. Figs.7 and 8 show the paths on the X-Z and X-Y planes, respectively. Fig.9 shows the velocity profiles.

The path of the hand in the model is a smooth curve whose shape is similar to the path of the human hand, as shown in Figs.7 and 8. In Fig.9, the velocity profile of the model shows smooth changes. Moreover, its shape is similar to the velocity profile for human hand movements. These results indicate that the hand-over motion model can accurately represent the human hand motion for the giver side.

Fig. 7. Path of model (X-Z Plane)

IV. HAND-OVER ROBOT SYSTEM

A hand-over robot system that uses the proposed hand-over motion model is developed. The developed hand-over robot system is shown in Fig.10. The robot’s arm is fabricated according to the average size of a human arm [10]. It has four degrees of freedom (two degrees for the shoulder joint and one degree each for the elbow and wrist joints). The coordinate system of the robot is shown in Fig.10. The origin of the coordinate system is defined as a point at the center of the shoulder. Furthermore, a hand with five fingers is constructed, as shown in Fig.11. This hand is driven by a pneumatic mechanism. The mechanism involves nonlinear behavior owing to the nonlinear behavior of the compressibility of air. Therefore, the robot grips an object gently like a human would do using the hand.

A microphone is used for the measurement of the human voice. Furthermore, a magnetic sensor (FASTRAK) is used for the measurement of the position and angle of the human hand. The hand-over motion model is used to calculate the desired position of the robot based on the obtained hand position of the human. Inverse kinematics is used to calculate the joint angle from the obtained desired position of the robot. The robot is controlled according to the obtained joint angle. Each joint is driven by an AC servomotor through a decelerator. The joint angle is measured by a pulse signal from the rotary encoder installed in the servomotor. The sampling time of the system is 10 ms. The robot follows the desired position accurately.

V. EXPERIMENT OF HAND-OVER MOTION MODEL

An experiment was performed in order to analyze the hand-over motion of the robot as preferred by humans. In the hand-over motion model, the robot released the object in response to a voice utterance from a human. Therefore, the timing between the voice utterance of the human and the release motion of the robot is important for a smooth hand-over motion. Thus, the timing preferred by humans was determined experimentally. Fig.12 shows the experimental scene.

A. Experimental method

In analysis of hand-over motions of humans, many humans release an object from between 0.0 s to 0.6 s after hearing
a voice from the receiver. Furthermore, on average, humans release an object at 0.3 s after hearing a human voice. Therefore, three modes were used in the experiment. In mode (a), the release motion is performed simultaneously with hearing a human voice. In mode (b), the release motion is performed at 0.3 s after hearing a human voice. In mode (c), the release motion is performed at 0.6 s after hearing a human voice.

A paired comparison was first carried out. Subsequently, a seven-point bipolar rating was determined. The paired comparison was performed for all six pairs of combinations. The seven-point bipolar rating with a scale from -3 (not at all) to 3 (extremely) was determined for the following four items: “Ease of receiving,” “Certainty,” “Security,” and “Politeness.” The subjects used the modes randomly. The subjects were 30 healthy students aged between 20 and 24 years.

### B. Experimental results

1) **Paired comparison:** The result of the paired comparison is shown in Table II. The table shows the number of subjects that preferred the column mode to the row mode. Table II indicates that mode (b) was preferred by the largest number of subjects. Furthermore, the Bradley-Terry model [11] was fitted to the results to analyze the results quantitatively by using equation (4).

\[
P_{ij} = \frac{\pi_i}{\pi_i + \pi_j}
\]

\[
\sum_i \pi_i = \text{Const.} (= T_{\text{Total}} : 100)
\]

\(\pi_i\) is the intensity of the preference for model \(i\). \(P_{ij}\) is the probability of the judgment that \(i\) is better than \(j\).

Using this model, the results of the paired comparison were expressed by the intensity of preference \(\pi\), as shown in Fig.13. The suitability of the model was validated by the goodness-of-fit and likelihood ratio tests. The figure indicates that mode (b) was rated as the best mode.

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>11</td>
<td>30</td>
<td></td>
<td>41</td>
</tr>
<tr>
<td>(b)</td>
<td>49</td>
<td></td>
<td>54</td>
<td>103</td>
</tr>
<tr>
<td>(c)</td>
<td>30</td>
<td>6</td>
<td></td>
<td>36</td>
</tr>
</tbody>
</table>

2) **Seven-point bipolar rating:** The results of the seven-point bipolar rating are shown in Fig.14. Mode (b) was evaluated as better than modes (a) and (c) by a significant difference of 1 % in the “Ease of receiving” item. This result agrees with the result of the Bradley-Terry model. However, in the “Certainty” and “Politeness” items, mode (c) was evaluated as better than modes (a) and (b) with a significant difference of 1 % or 5 %. Furthermore, mode (b) was evaluated as better than mode (a) by a significant difference of 1 %. In the “Security” item, modes (b) and (c) were evaluated as better than mode (a) by a significant difference of 1 %.

From the results, mode (b) was found to be the preferred mode. Furthermore, in the seven-point bipolar rating, all items of mode (b) were evaluated as positive. Therefore, it was clarified that the proposed hand-over motion model can generate hand-over motion with humans without eliciting feelings of aversion in them, thus demonstrating the effectiveness of the hand-over motion model.
Experimental results show that humans received an object easily when the release motion of the robot was performed at 0.3 s after hearing a human voice. However, from the viewpoints of certainty and politeness, humans preferred the hand over in which the release motion was performed at 0.6 s after hearing a human voice. From this result, it was evident that feelings of certainty and politeness increased by lengthening the time for which both the human and the robot were holding the object.

VI. CONCLUSIONS

In this paper, we proposed a hand-over motion model based on the timing between the voice utterances and the release motions of humans. This model generates a motion in which a robot hands over an object to a human accompanied by a voice greeting. In this model, a robot generates the release motion in response to a voice utterance from a human based on the analysis of the timing between the voice utterances and the release motions of humans. Furthermore, a hand-over robot system that uses the proposed hand-over motion model was developed. The effectiveness of the proposed model was demonstrated by sensory evaluation using the hand-over robot system.

The experimental results can be summarized as follows:

- In the analysis of hand-over motions between humans, humans moved their hands in a smooth trajectory. Furthermore, givers performed their release motion according to the voices of receivers. Therefore, it was evident that humans synchronize their embodied rhythms by their voices, and perform hand-over motion at precise timing.

- The hand-over motion with a voice between a human and a robot could be generated using a hand-over robot system that used the proposed hand-over motion model. Furthermore, the generated motion was experimentally found to be preferred by humans. Therefore, the effectiveness of the proposed hand-over motion model was demonstrated.

- When the release motion of the robot was performed at 0.3 s after hearing a human voice, humans received an object easily. Furthermore, it was found that feelings of certainty and politeness increased by lengthening the time for which both the human and the robot held the object.

ACKNOWLEDGMENT

This work was supported by KAKENHI Grant Number 25330239 of the Japan Society for the Promotion of Science (JSPS), Japan.

REFERENCES


