Directing a Target Person Among Multiple Users Using the Motion Effects of an Image-Based Avatar

Tsubasa Miyauchi[(✉)], Masashi Nishiyama, and Yoshio Iwai

Graduate School of Engineering, Tottori University,
101 Minami 4-chome, Koyama-cho, Tottori 680-8550, Japan
miyauti.tsubasa51@gmail.com

Abstract. We investigate whether an image-based avatar with motion can smoothly direct a target person who has requested guidance. Existing methods that use an avatar generally assume a one-to-one interaction with the user and did not fully consider how the avatar should direct a target person, which is important in interactions with multiple users. When the target person feels that the image-based avatar is facing them and feels that the avatar is directing them, the person judges that the requested guidance has been given. When a nontarget person feels that the image-based avatar is not facing them and also feels that the avatar is not directed towards them, the nontarget person then judges that the requested guidance has not come to them. To direct the target person smoothly, the action-type motion of the image-based avatar and the rotational motion of the video sequence itself relative to the display are important. We performed a subjective assessment of how the target person and the nontarget person feel about the direction by comparing the case where motions were added with the case where motions were not added. The results show that the image-based avatar with the action-type and rotation-type motions is effective for the target person but is not effective for the nontarget person.

Keywords: Image-based avatar · Motion · Direction · Target person · Nontarget person

1 Introduction

We have developed an image-based avatar system [1] that interacts naturally with users. Several image-based avatar systems have been produced over the last decade [2,3]. Specifically, we considered a navigation system that guides a user and consists of an avatar shown on a stationary display. The avatar serves as an alternative to a real guide in an information center located in a public space, e.g., at an airport or in a bus/rail station.

Before describing this system, we consider an actual navigation situation involving people in an information center, where a real guide interacts with

Fig. 1. Navigation situation.

the users waiting around him/her. Figure 1 shows an example of this type of situation. Based on our knowledge of cognitive science [4], the people can be categorized into the following three roles: a guide that provides directions, a participant that receives these directions, and a side participant who is waiting to receive directions from the guide. A side participant becomes a participant when they are given directions by the guide.

Here, we consider how the avatar gives information to a single side participant when multiple side participants are present and waiting to begin interaction with the avatar. First, multiple side participants are waiting around the avatar for directions, as illustrated in Fig. 2(a). In order to provide clear separation of the roles of the multiple side participants, we also newly define the following roles: a target person who is receiving directions and a nontarget person who is not receiving directions, as illustrated in Fig. 2(b). When a target person feels that they are being given directions by the avatar, this person changes from a side participant into a participant, as illustrated in Fig. 2(c). The nontarget person remains in their role as a side participant as long as they do not feel that they are being directed by the avatar. It is important that the avatar only directs the target person in the presence of the nontarget person.

We now discuss how to design a method such that the avatar only directs the target person, even when a nontarget person is present. Existing methods [5–9] have used special displays with embedded motion mechanisms. However, we cannot use this method easily in a stationary display. Caridakis et al. [10] used a stationary display with video sequences that included motion effects when the avatar spoke to the user. We believe that the motion effects provide an important cue when the avatar and the user interact via a stationary display. Therefore, we focused on the use of motion effects in the video sequences for the image-based avatar. However, this method does not sufficiently consider the problem of how to direct a target person when in the presence of a nontarget person.

In this paper, we investigate the use of motion effects in an image-based avatar to enable this avatar to give directions smoothly to a target person in the presence of another nontarget person. To compare methods that add motions to
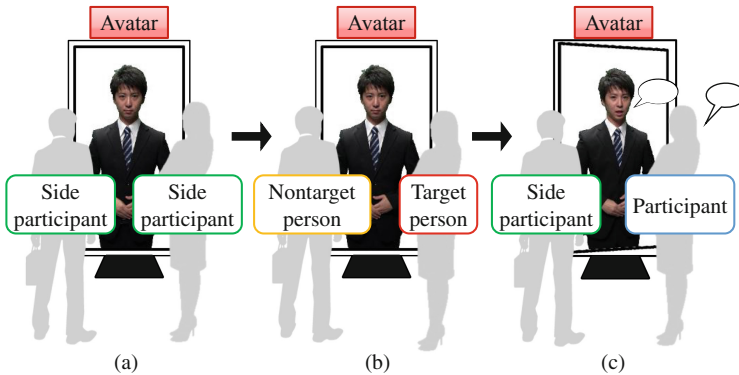
Fig. 2. User role transitions in the proposed navigation system.

the image-based avatar with a method without movement of the image-based avatar, we evaluated the following two hypotheses:

H1: A target person feels that they are facing the image-based avatar. The target person then feels that they are being given directions by the image-based avatar.

H2: A nontarget person does not feel that they are facing the image-based avatar. The nontarget person thus does not feel that they are being given directions by the image-based avatar.

We evaluated the effects of an action-type motion (where the face and body of the avatar move to face the target person) and a rotational motion (where the display rotates to face the target person).

2   Navigation Situation

We assume a situation in which the side participants stand on the right and left sides of the participant in the information center. Figure 3 shows the flow that occurs in this situation. At t_1, a participant is guided by an image-based avatar. At t_2, a side participant who wants to use the image-based avatar arrives the information center and stands in an empty space at the side of the participant. At t_3, another side participant arrives at the information center and stands on the empty side opposite the first side participant. At t_4, the participant leaves the information center after receiving the image-based avatar's guidance. At this point, the side participants stand on the right and left sides of the position that had been occupied by the participant. To direct the target person only, we add motion effects to the image-based avatar in this situation.
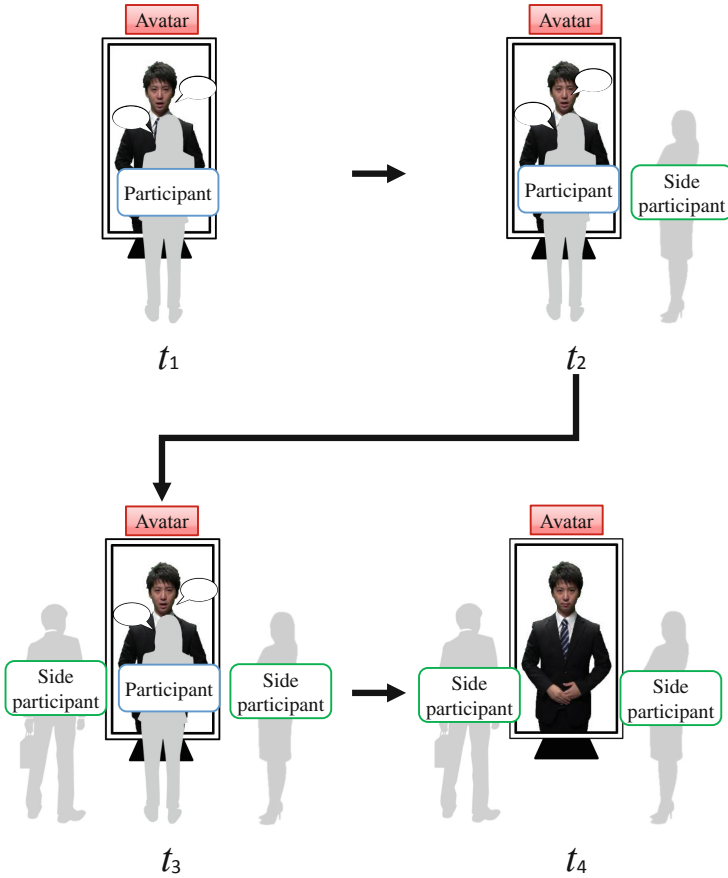
Fig. 3. Flow of the situation that we consider in this paper.

3   Motion Effects

3.1   Overview

To direct a target person only, we must consider what motion must be added to the image-based avatar. Because the image-based avatar resembles the appearance of a person, the image-based avatar is then expected to interact like a person. However, the image-based avatar is also expected to produce an interaction like that between a person and a machine because the image-based avatar is displayed on a display. We therefore added a guide motion and a rotating display to the image-based avatar. We describe the details of each motion below.

First, we explain how a guide moves to direct a target person only. A guide talks to a target person after facing that target person. It is important that the guide faces the target person. Figure 4 shows the way in which a guide moves to direct the target person only. The target person feels that they are facing a guide
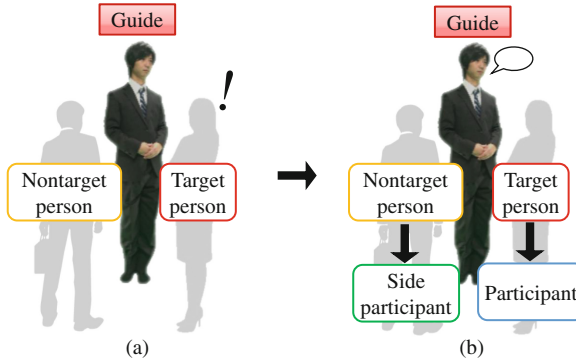
Fig. 4. How the guide moves to direct the target person only.

because that guide is facing them (Fig. 4(a)). Furthermore, the target person feels that they are being given directions because that guide is talking to them (Fig. 4(b)). In contrast, the nontarget person does not feel that they are facing the guide because that guide is facing the target person (Fig. 4(a)). Furthermore, the nontarget person does not feel that they are being given directions by that guide because the guide is talking to the target person (Fig. 4(b)). There are motions in the face, the body and the eyes in the case where the guide faces the target person.

Below, we consider addition of the motions of a guide to the image-based avatar. When using a video sequence for an image-based avatar, we must consider the Mona Lisa effect [11,12], which occurs whenever users see an avatar that is displayed on a flat panel. This effect causes the users to feel that an avatar who is facing the camera is actually gazing directly at them. If the avatar faces the camera, then both the target person and the nontarget person will feel that they are facing the avatar simultaneously. To alleviate the Mona Lisa effect, our method uses the fact that the avatar rotates both its face and its body while gazing at the camera. When an image-based avatar talks to a target person after
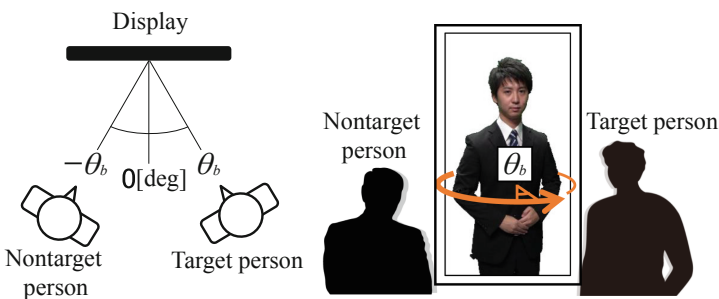

Fig. 5. Angle parameter θ_b for an action-type motion.
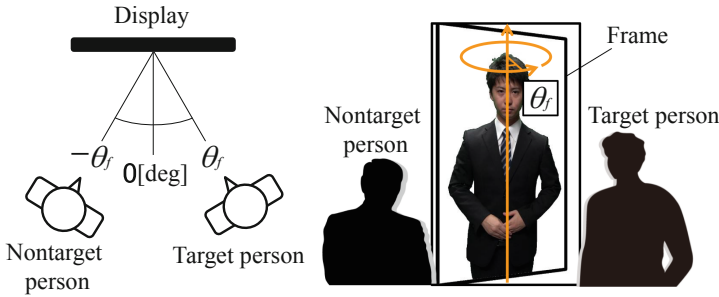
Fig. 6. Angle parameter θ_f for a rotational motion

rotating its face and body, the target person then feels that they are being given directions by the image-based avatar.

Next, we explain how to move a rotating display to direct a target person only. The display rotates towards the target person. It is important that the rotating display faces the target person. In this paper, we represent the physical frame of the rotating display in the video sequence of the image-based avatar. We can change the appearance of the region of the frame in which the avatar is located using projective transformation of the video sequence. As a result, the target person feels that they are facing an image-based avatar. When the image-based avatar talks to a target person after the projective transformation, the target person then feels that they are being given directions by the image-based avatar.

3.2   Action-Type Motion

Action-type motion is used to describe the scenario where the image-based avatar has a guiding motion added to the video sequence. In this case, the image-based avatar rotates its face and body while its gaze remains fixed towards the camera. Note that the image-based avatar rotates its face and body in conjunction because rotation of its face and body individually would occur only rarely in the information center. Figure 5 shows the angle parameter θ_b for an action-type motion. θ_b is the angle of the rotating body. θ_b sets the front direction to 0°. When the target person is standing on the right side of the avatar, θ_b has a positive value. When the target person is standing on the left side of the avatar, θ_b has a negative value.

3.3   Rotational Motion

Rotational motion is used to describe the situation where the motion of a rotating display is added to the image-based avatar in the video sequence. Changes

in the appearance of the subject region in the frame and the frame itself are expressed by projective transformation. Figure 6 shows the angle parameter θ_f of the rotational motion. θ_f is the angle of the rotating frame. θ_f sets the front direction to $0°$. When the target person is standing on the right side of the avatar, θ_f has a positive value. When the target person is standing on the left side of the avatar, θ_f has a negative value.

3.4 Combination of an Action-Type Motion with a Rotational Motion

Next, we need to consider combination of an action-type motion with a rotational motion. A rotational motion is added to an image-based avatar during an action-type motion. This combination of the action-type motion and the rotational motion has an angle parameter composed of θ'_b and θ'_f. We set $\theta'_b = \theta_b/2$ and $\theta'_f = \theta_f/2$.

4 Subjective Assessment

4.1 Experimental Conditions

We performed a subjective assessment to investigate the hypotheses described in Sect. 1. In addition, we performed another subjective assessment to investigate the impression that is made by the image-based avatar. Twenty subjects (17 males and 3 females, with an average age of 22.2 ± 1.1 years) participated in this assessment. We compared the following four motion methods for the image-based avatar:

M1: No motion effects
M2: Action-type motion
M3: Rotational motion
M4: Both motion types

Figure 7 shows examples of these methods.

The 20 subjects were split randomly into pairs to view a video sequence of an image-based avatar for each method. One subject was assigned the role of the target person and the other subject was assigned the nontarget person role. The two subjects then stood to the side of the display. Figure 4 shows the standing positions of the two subjects. After viewing the video sequence, the subjects then answered the following questions:

Q1: Did you feel that you faced an image-based avatar?
Q2: Did you feel that you were given directions by an image-based avatar?
Q3: Did you feel that you interacted smoothly with the image-based avatar?
Q4: Did you feel that the image-based avatar interacted politely with you?
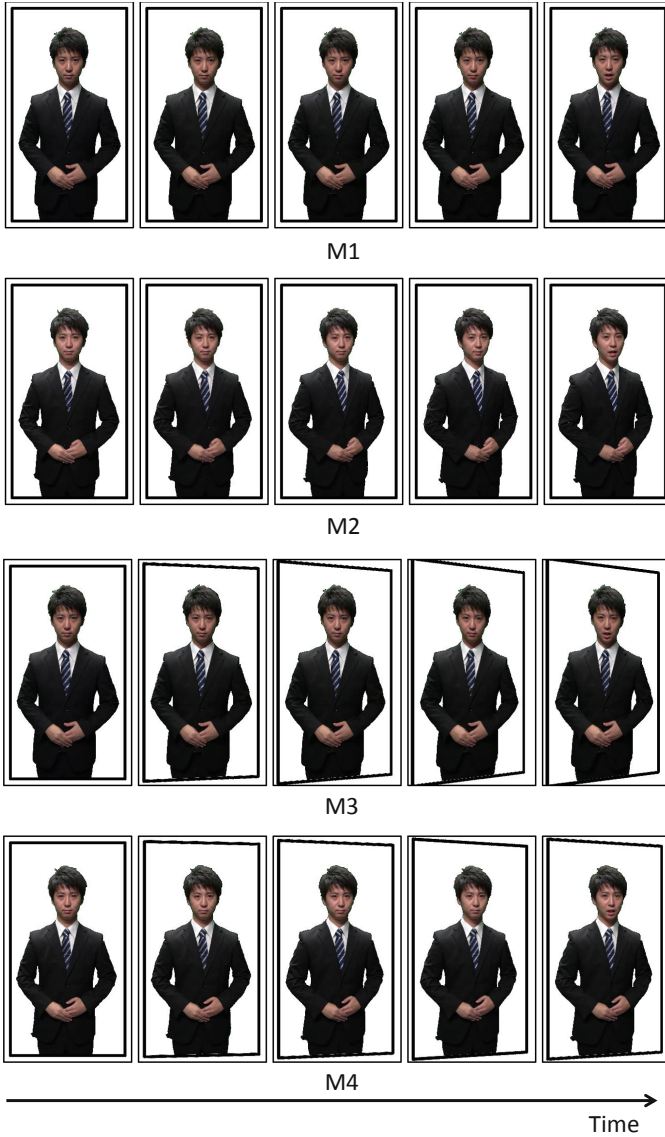Q5: Did you feel that the image-based avatar interacted nicely with you?

Fig. 7. Image-based avatars M1 to M4 with and without motion effects for the subjective assessment.

Each subject provided a rated score using four response levels (-1.5: disagreeable; -0.5: slightly disagreeable; 0.5: fairly agreeable; 1.5: agreeable) for each question. We also asked the reverse questions to Q1 to Q4. We set $\theta_b = 15[\text{deg}]$ and $\theta_f = 15[\text{deg}]$. We used two-way analysis of variance (ANOVA) and the Wilcoxon signed-rank test to evaluate the test results.
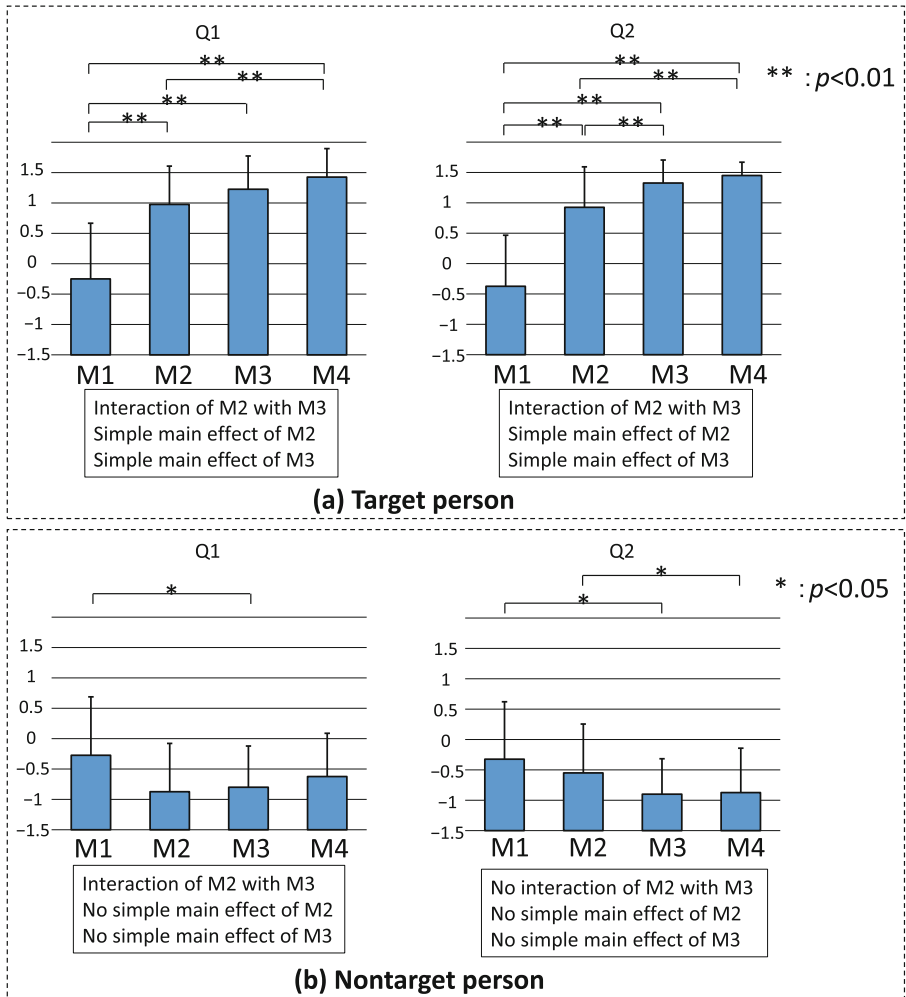
Fig. 8. Results of subjective assessments of Q1 and Q2.

4.2 Results of Subjective Assessment

Figure 8(a) shows the subjective scores for Q1 and Q2 for the target persons. The high subjective scores indicate agreement among the target persons. Additionally, there are significant differences among M1 and the other three methods. Therefore, we can claim that H1, as described in Sect. 1, is valid under the condition that motion effects were used. Figure 8(b) shows the subjective scores for Q1 and Q2 for the nontarget persons. The low subjective scores also indicate agreement among the nontarget persons. There were no significant differences among M1 and the other three methods. Therefore, we cannot claim that H2, as described in Sect. 1, was valid under the condition that motion effects are used.
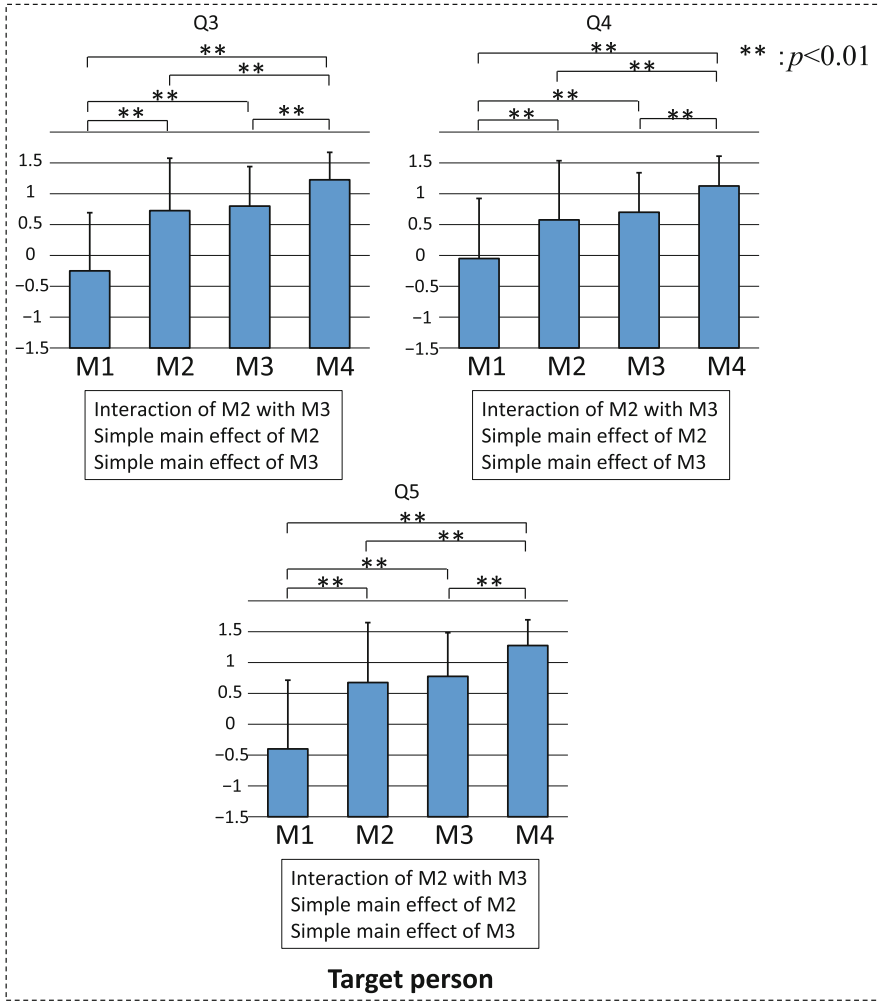
Fig. 9. Results of subjective assessments of Q3 to Q5 for the target persons.

Figure 9 shows the subjective scores for Q3 to Q5 for the target persons. The high subjective scores again indicate agreement among the target persons. Additionally, there are significant differences among M1 and the other three methods. Therefore, we can claim that the image-based avatar made a good impression under the conditions where motion effects were used. Figure 10 shows the subjective scores for Q3 and Q5 for the nontarget persons. The low subjective scores also indicate agreement among the nontarget persons. There were no significant differences in this case among M1 and the other three methods. Therefore, we cannot claim that the image-based avatar made a good impression under the conditions where motion effects were used.
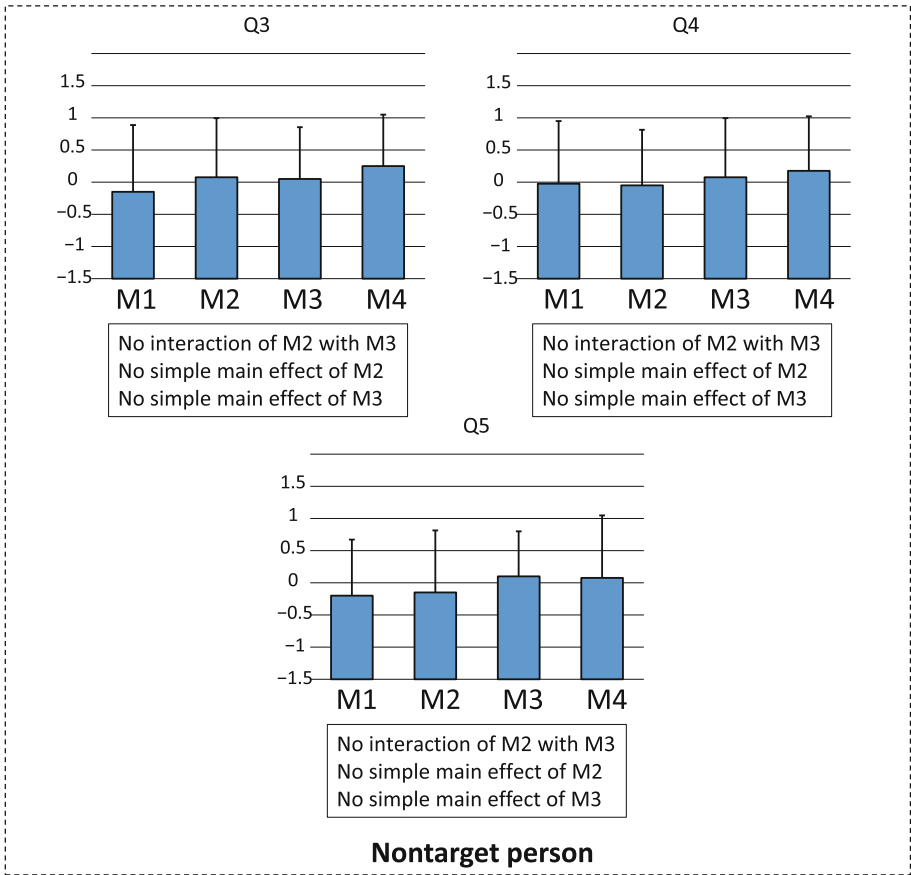
Fig. 10. Results of subjective assessments of Q3 to Q5 for the nontarget persons.

5   Conclusions

We have proposed a method to ensure that an image-based avatar only directs a specific target person. We added an action effect and a rotation effect to an image-based avatar. We then performed a subjective assessment to compare the methods that add the effects to the image-based avatar with a method without movement in the image-based avatar. The results of a subjective assessment showed that a target person feels that they are being given directions by the image-based avatar when the motion effects are used. In addition, the image-based avatar also made a good impression on the target person during the use of motion effects. However, the results of the subjective assessment did not show that the nontarget person did not feel that they were being given directions by the image-based avatar. In future work, we will consider motion effects in another situation in an information center, and we will also compare motion effects and special displays.

References

1. Miyauchi, T., Ono, A., Yoshimura, H., Nishiyama, M., Iwai, Y.: Embedding the awareness state and response state in an image-based avatar to start natural user interaction. IEICE Trans. Inf. Syst. **E100-D**(12), 3045–3049 (2017)
2. Artstein, R., et al.: Time-offset interaction with a Holocaust survivor. In: Proceedings of the 19th International Conference on Intelligent User Interfaces, pp. 163–168 (2014)
3. Jones, A., et al.: An automultiscopic projector array for interactive digital humans. In: Proceedings of ACM SIGGRAPH Emerging Technologies, p. 6:1 (2015)
4. Clark, H.H., Carlson, T.B.: Hearers and speech acts. Language **58**, 332–373 (1982)
5. Otsuki, M., Kawano, T., Maruyama, K., Kuzuoka, H., Suzuki, Y.: ThirdEye: simple add-on display to represent remote participant's gaze direction in video communication. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI 2017, pp. 5307–5312 (2017)
6. Kawaguchi, I., Kuzuoka, H., Suzuki, Y.: Study on gaze direction perception of face image displayed on rotatable flat display. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, pp. 1729–1737 (2015)
7. Yankelovich, N., Simpson, N., Kaplan, J., Provino, J.: Porta-person: telepresence for the connected conference room. In: Extended Abstracts on Human Factors in Computing Systems, CHI 2007 (2007)
8. Adalgeirsson, S.O., Breazeal, C.: MeBot: a robotic platform for socially embodied presence. In: Proceedings of 5th ACM/IEEE International Conference on Human-Robot Interaction, pp. 15–22 (2010)
9. Onishi, Y., Tanaka, K., Nakanishi, H.: Embodiment of video-mediated communication enhances social telepresence. In: Proceedings of the Fourth International Conference on Human Agent Interaction, pp. 171–178 (2016)
10. Caridakis, G.: Virtual agent multimodal mimicry of humans. Lang. Resour. Eval. **41**(3–4), 367–388 (2007)
11. Masame, K.: Perception of where a person is looking: overestimation and understimation of gaze direction. Tohoku Psychologica Folia **49**, 33–41 (1990)
12. Kendon, A.: Some functions of gaze-direction in social interaction. Acta Psychologica **26**, 22–63 (1967)