Investigation of Motion Video Enhancement for Image-based Avatars on Small Displays

Tsubasa Miyauchi[1], Wataru Ganaha[1], Masashi Nishiyama[1 [0000−0002−5964−3209]], and Yoshio Iwai[1 [0000−0003−2248−4563]]

Graduate School of Engineering, Tottori University
101 Minami 4-chome, Koyama-cho, Tottori, 680-8550 Japan
miyauti.tsubasa51@gmail.com, {nishiyama, iwai}@tottori-u.ac.jp

Abstract. We investigate a method for enhancing the motion video sequences of an image-based avatar so that the body motion can be perceived as natural on a small display. If some avatar motions are too small, then the users cannot perceive those motions when the avatar is viewed on a small display. In particular, the motion of an upright posture that the avatar uses when waiting to start interacting with the user is very small. In this paper, we enhance the motion of the upright posture so that the user naturally perceives the movement of the avatar as human-like, even on a small display. To do this, we use an existing method for phase-based video motion processing. This method allows us to control the amount of avatar movement using a pre-defined enhancement parameter. The results of our subjective assessment show that the users sometimes perceived the avatar's motion as natural on a small display when the body sway motions of the avatar were appropriately enhanced to the extent that no significant noise was included.

Keywords: Avatar · Body sway · Motion video enhancement.

1 Introduction

Interactive systems that use human-like avatars are an active topic in human-agent interaction research and have many potential applications, such as conversational avatars in a museum [1], avatars who talk about the past [2], and speechinteractive guidance avatars [3]. These avatars can utilize the large displays that we often see in many places (e.g., in stations, shopping plazas, office entrances, and airports), and they can automatically communicate with users without the constraints of time and place (e.g., at an automatic information desk late at night). In particular, we focus on navigation systems that synthesize human-like realistic avatars by exploiting an image-based technique [2, 4] in which video sequences interact with users. To develop such navigation systems, it is important to provide a navigation guide that is intuitive for users, such as people who are not familiar with complex technology, to use. In this paper, we focus on the use of image-based techniques to generate realistic human-like avatars that can interact with users naturally.

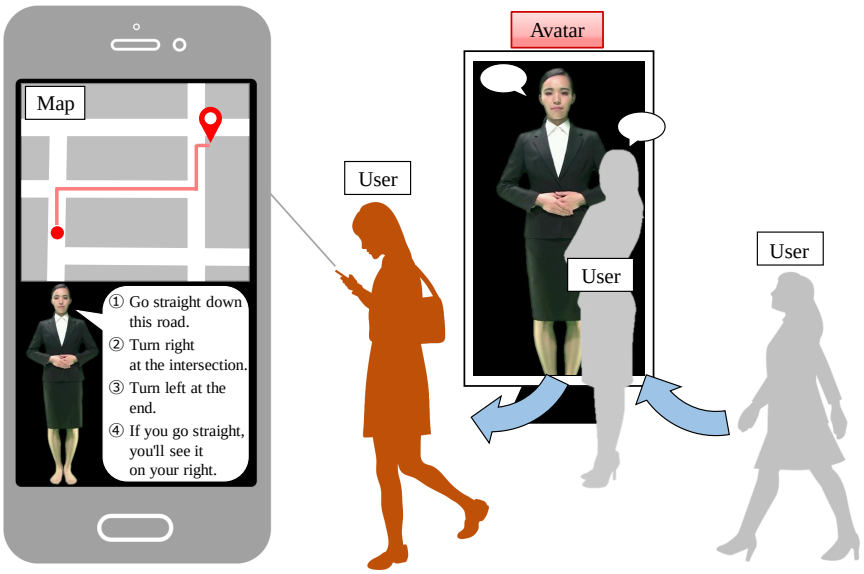


Fig. 1. Overview of our navigation system using image-based avatars on large and small displays.

Figure 1 shows an overview of our navigation system using image-based avatars. We use such an avatar on a large display as digital signage. A user stands in front of an avatar on this display and talks to it to receive directions. In our system, an avatar interacts with a user using human-like behavior. Here, we consider a situation in which it will be difficult for the user to remember these directions (e.g., a path to the destination is very complex). To help the user in this situation, our system then guides the user on his/her mobile device, such as a smartphone, after the user has finished talking with an avatar on a large display. Existing research [5] shows that an avatar that points at a map significantly increases the likeability of the system. Therefore, our system displays an avatar and a map on the user's mobile device. The avatar on the small display continues to guide the user until the destination is reached to prevent the user becoming lost en route.

We consider how to generate an intuitive video sequence of an image-based avatar, which is an important element in a navigation system. We divide the video generation into generation for large displays and generation for small displays. Because a life-sized avatar can be displayed on a large display, we can easily reproduce the movement of a real person in the avatar. In contrast, when displaying an avatar on a mobile device, the avatar motions become too small because the screen size of the mobile device is much smaller. As a result, the user cannot naturally perceive the avatar's motions. So that the avatar motions

can be perceived by the user, we need to adaptively control the amplitude of the avatar's movement. Here, we consider a situation in which an avatar with an upright posture waits to start interacting with the user before providing directions. The avatar may appear to be completely stationary, that is, moving little or not at all, on a small display. In this situation, a user incorrectly feels that the avatar system is broken and interaction cannot begin. Therefore, we need to design a video generation method for displaying an avatar with a natural-looking upright posture on a mobile device.

When we observe the motions of standing people with an upright posture, their bodies constantly swing around a certain position. In these motions, people are subconsciously swinging their bodies to avoid overloading some muscles. A method for reproducing body sway motions in an image-based avatar has previously been proposed [6]. It generates continuous and natural body sway from a short video sequence of an avatar with an upright posture. However, the existing method was designed for an image-based avatar on a large display. If we use the existing method on a small display, a user may perceive the avatar motions as very small and unnatural. We therefore need to enhance avatar motions with an upright posture on a small display so that a user feels that the motions are more natural.

In this paper, we investigate the intuitive video generation of image-based avatars with an upright posture on a small display. Specifically, we employ an existing method for phase-based video motion processing and evaluate an enhancement parameter used in the method. To do this, we test the following hypothesis.

Hypothesis: Users perceive the motion of an image-based avatar as more natural if it is enhanced using an enhancement parameter of phase-based video motion processing that is appropriate for a small display.

2 Importance of an upright posture for avatars

Existing research [6, 7] has found that, to synthesize realistic avatars for interactive systems, the states of the avatar must be considered. Two states of the avatar are important for the avatars of interactive systems. The first is the action state, in which the avatar speaks with a user, and the second is the wait state, in which the avatar maintains a standing pose. In particular, the wait state has an important role in the interaction between avatars and users because the avatars do not continuously communicate with the users. In contrast, the wait state is needed to handle periods before and after interaction. For example, if the wait state is not adequately applied in the avatars, the users cannot judge whether to begin communication with the avatars or not.

The wait state of the avatars enhances the nonverbal communication between users and avatars, as illustrated in Fig. 2. To achieve this, we synthesize human-like micro movements. Consider a standing pose in the wait state; we often see standing receptionists at airports, hotels, and information offices. When people maintain a standing pose, they continuously move their body slightly around

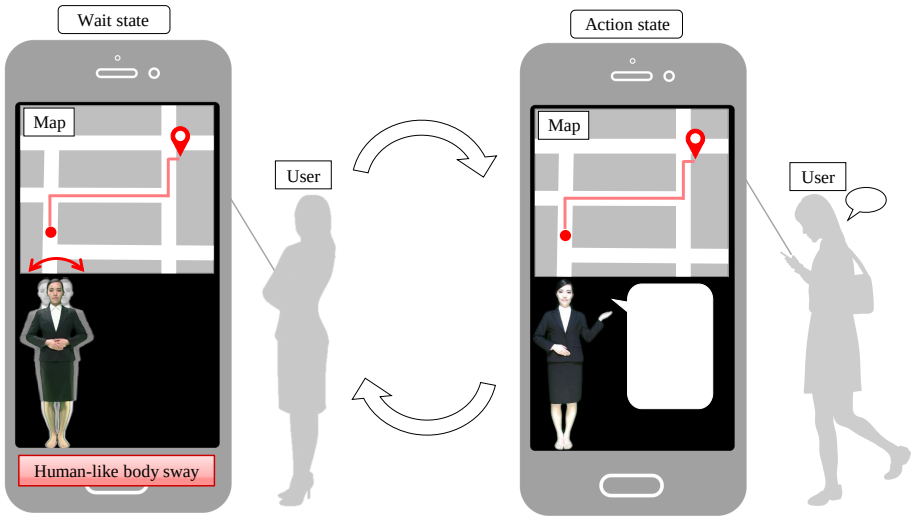


Fig. 2. We consider the body sway of an image-based avatar in the wait state to enhance the nonverbal communication between it and a user in an interactive system.

its center line. As described in [8], people unconsciously spread the burden of standing across their muscles by making micro movements to avoid overloading one set of muscles. This physiological phenomenon is called body sway.

We consider reproducing body sway in an image-based avatar. Body sway is a very small movement. When the image-based avatar is displayed on a large display, the user can visually recognize the movement of the body sway, so we can reproduce the body sway motions of an actual human in the image-based avatar. When displaying the image-based avatar on a small display, the user cannot visually perceive the body sway motions of an actual human. Then, the user misunderstands that the image-based avatar is stopping and the navigation system is out of order. We therefore need to enhance body sway motions of the image-based avatar on a small display so that a user feels that the motions are more natural.

3 Method for enhancing body sway motions

Here, we explain the method of enhancing body sway motions in an image-based avatar. In the most basic approach to this method, a data set is created by taking images of an image-based avatar standing upright with a large body sway motions. However, this method would need to use different data sets depending on whether a large or small display is used, and a large amount of data sets would need to be prepared. In addition, body sway is performed unconsciously by humans, so it is difficult to consciously increase the amplitude of body sway

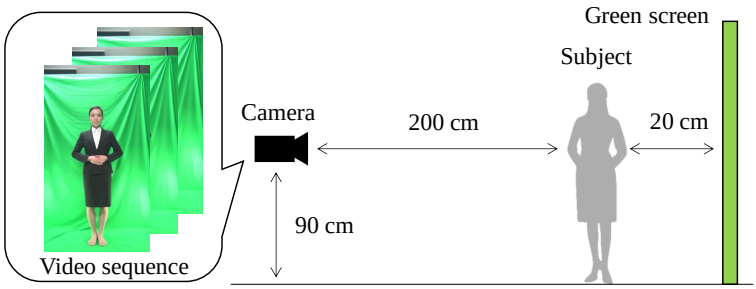


Fig. 3. Camera setup for acquiring the avatar images.

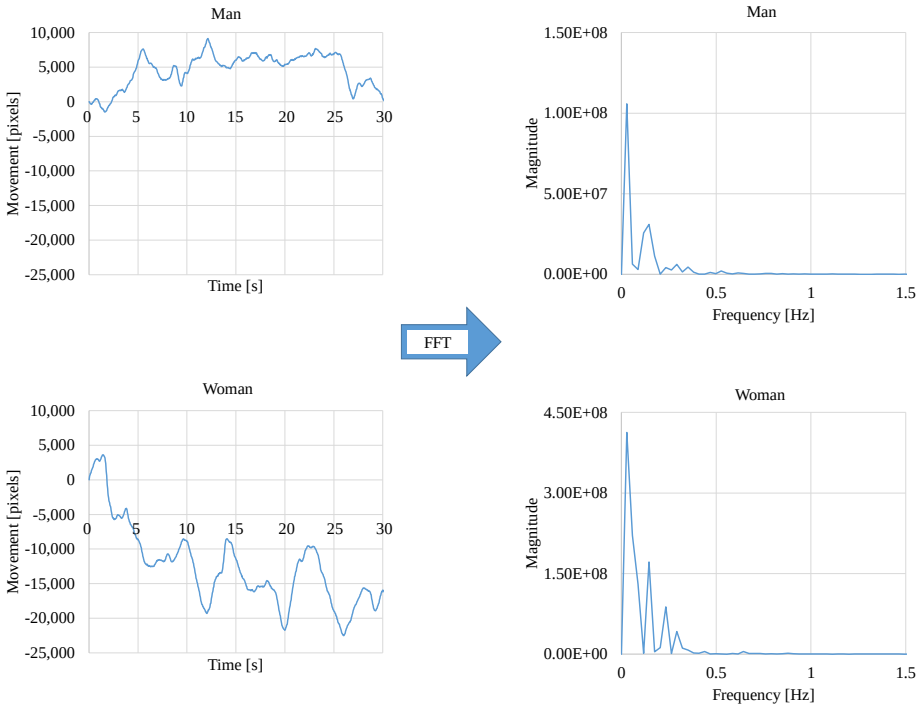
motions. Therefore, we enhance the body sway motions of the image-based avatar in the images instead.

One method for enhancing periodic motion in images is phase-based video motion processing [9]. In this paper, we use this method to enhance body sway motions. One study [10] found that the frequencies of most body sway movements range from 0 to 1.5 Hz. Here, we measure the actual body sway motions of an image-based avatar in the images, and the frequencies of the body sway are investigated. Figure 3 shows the setup for obtaining the image-based avatar images. The person region of the image-based avatar is extracted from the image-based avatar images using background subtraction. The body sway motions are measured by comparing the person regions in the images on the time axis. Specifically, we measure the difference between the person region of the image at the reference time and the person region at the comparison time. Figure 4 (a) shows the result for head motions in the image (one man and one woman). The vertical axis of the graph represents the difference in the person regions of the reference and comparison images. Figure 4 (b) shows the results for body sway frequency obtained using the fast Fourier transform (FFT) of the difference results. We confirmed that the frequencies of body sway movements range from 0 to 1.5 Hz from the results for both the male and female avatars. Similar results were obtained for parts other than the head. Therefore, phase-based video motion processing enhances frequencies in the range of 0 to 1.5 Hz. We control the amplitude of motion enhancement using parameter alpha, where a larger value of alpha generates larger motions.

4 Subjective assessment

4.1 Conditions of the subjective assessment

We conducted a subjective assessment in which we displayed image-based avatars with enhanced motions. We set the frequencies to range from 0 to 1.5 Hz, which



(a) Temporal signals measured from body sway

(b) Frequency analysis of body sway

Fig. 4. Measured results of body sway and the FFT results of body sway.

represents most body sway movements, as described in previous section. We used both male and female image-based avatars in the subjective assessment. Figures 5 and 6 show the video sequences of the male and female image-based avatars, respectively. The length of each video was 20 seconds. We compared the following experimental conditions:

M1: No motion enhancement.
M2: Enhancement using an alpha of 0 to generate a few motions.
M3: Enhancement using an alpha of 1 to generate weak motions.
M4: Enhancement using an alpha of 2 to generate relatively large motions.
M5: Enhancement using an alpha of 3 to generate very large motions.

We used the paired-comparisons method. Twenty-two participants (17 males and five females, with a mean age of 22.2 years) evaluated five avatars (M1, M2, M3, M4, and M5) presented on displays. The avatars were adjusted to the following sizes:

S1: 71.8 mm in height

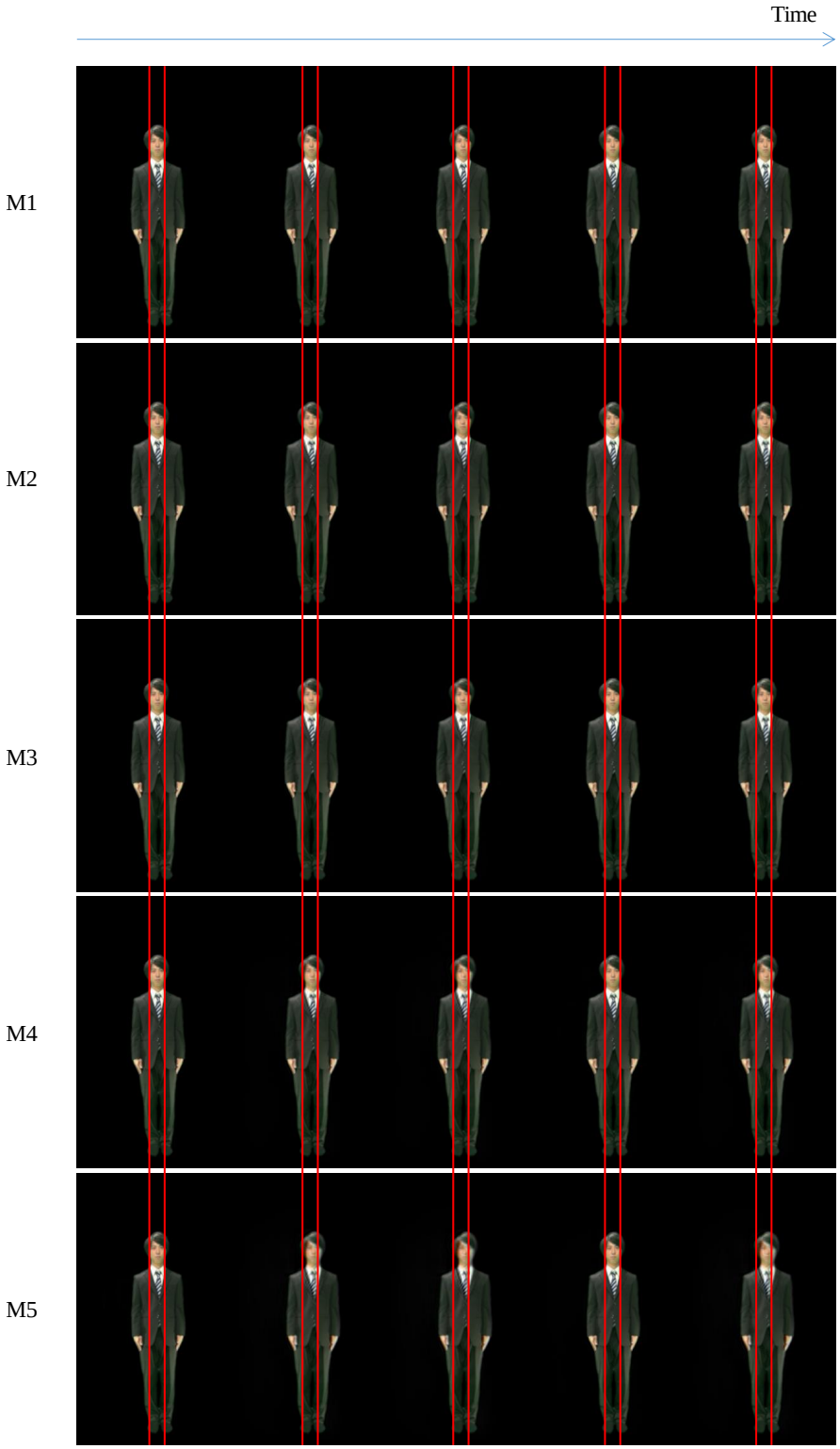


Fig. 5. Examples of video sequences of the male avatar used in our subjective assessment.

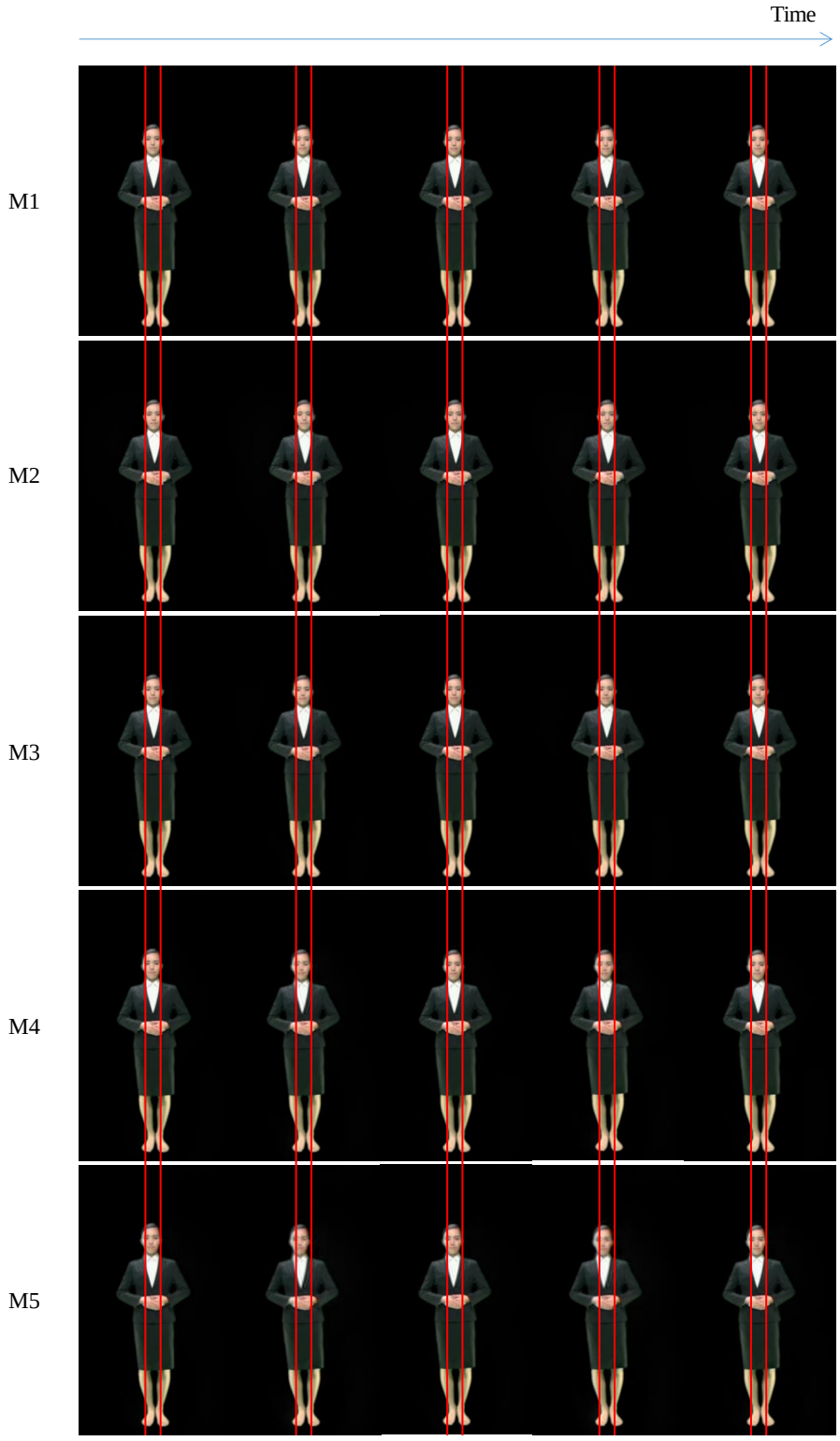


Fig. 6. Examples of video sequences of the female avatar used in our subjective assessment.

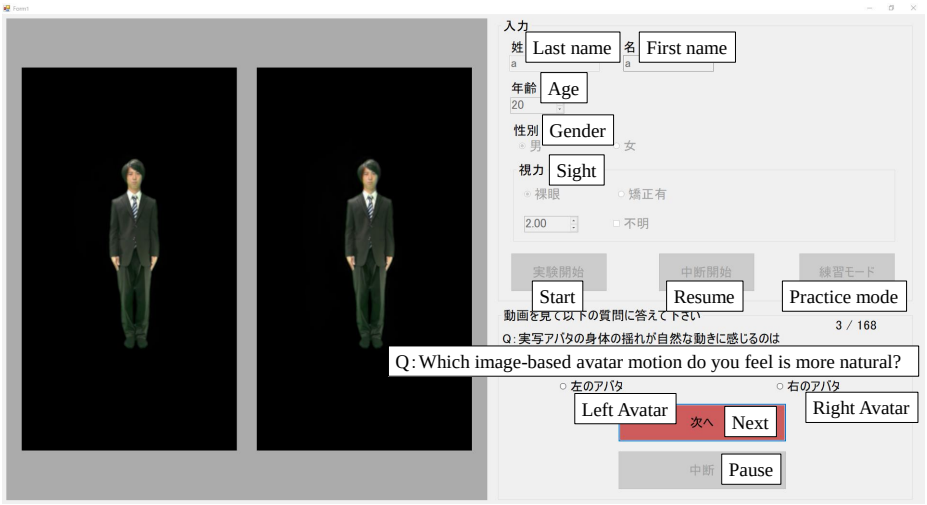


Fig. 7. Questionnaire used in our subjective assessment.

S2: 44.2 mm in height

That is, S1 was half as large as a 4-inch display, and S2 was half as large as a 6.5-inch display. We displayed pairs of avatars and asked the subjects to answer the following question:

Q: Which image-based avatar motion do you feel is more natural?

Figure 7 shows the question form. The participants evaluated the two image-based avatars displayed on the left side of the question form. Figure 8 shows the setup for each participant in our subjective assessment. The participants sat in a chair and used the question form displayed on the PC on the desk. The participant's eyes were 115 cm from the ground and 61 cm from the display on average. The average angle of the display was 119 degrees. The participants were presented with randomly generated pairs of five avatars (${}_5C_2 = 10$) of two sizes and two types of avatar (male and female), 40 times in total. The participants then selected the video sequence in the pair that he or she felt was more natural.

4.2 Result of subjective assessment

Figure 9 shows the result of the subjective assessment of the five types of body sway motion enhancement. The graphs express one axis of subjective evaluation from the number of votes obtained. A higher subjective score indicates agreement among the subjects, and vice versa. First, we focus on the results of the male avatars. For display size S1, M1 had the highest rating. We believe that the participants felt that the body sway motions of the avatar were sufficient without motion enhancement. In contrast, for S2, M3 and M4 had a higher rating than

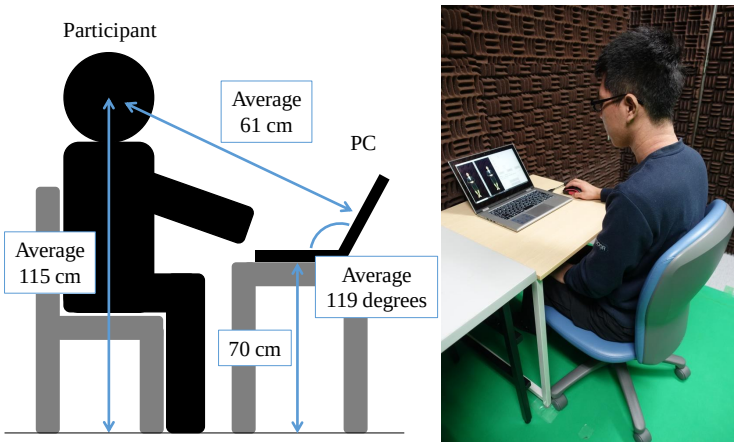


Fig. 8. Experimental setting for our subjective assessment.

M1. We believe that the participants felt that the motion enhancement was agreeable when the display size was smaller. The avatar with very large motion enhancement, M5, was unnatural because of continuous motions, so it received a low rating. M2 also received a low rating because it was too small.

Next, we focus on the results of the female avatars. For display size S1, M1 had the highest rating. Again, we believe that the participants felt that the body sway motions of the avatar were sufficient without motion enhancement. For display size S2, the ratings of M2, M3, and M4 were close to the rating of M1, but M1 had the highest rating. We believe that this result was caused by the shortcomings of the phase-based video motion processing. When enhancing motions, phase-based video motion processing introduces noise into the movie, and it can generate a large amount of noise depending on the movie. Figure 10 shows the noise of the female image-based avatar. When we enhanced motions of the female image-based avatar, high levels of noise were generated in the videos for M3, M4 and M5. Therefore, M3, M4, and M5 had ratings lower than that of M1. M2 received a low rating because the avatar's motion was too small.

In the results of the male avatar, M3 and M4 had a higher rating than M1 for S1. In the results of the female avatar, the ratings of M3 and M4 were closer to that of M1 in S2 than in S1. If we could enhance the body sway motion of the image-based avatar without adding noise, the participants might perceive the avatar motion as natural. In conclusion, the participants may have perceived the avatar motion as natural when it was enhanced with the appropriate value of alpha.

Fig. 9. Results of the subjective assessment. We used the paired comparisons method.

5 Conclusion

We enhanced the motions of an avatar with an upright posture on a small display so that a user feels that the motions are more natural. We used phase-based video motion processing to enhance the body sway motions. In a subjective experiment, the participants perceived the avatar motion as natural on a small display when we enhanced body sway motion of the image-based avatar and there were low levels of noise. In contrast, the participants did not perceive the avatar motion as natural when high levels existed in the video.

As part of our future work, we plan to develop a method for enhancing the body sway motion of image-based avatars that does not introduce noise.

References

1. Robinson, S., Traum, D., Ittycheriah, I., and Henderer, J.: What would you ask a conversational agent? Observations of human-agent dialogues in a museum setting. In: 6th International Conference on Language Resources and Evaluation (LREC), pp. 28–30. (2008).
2. Artstein, R., Traum, D., Alexander, O., Leuski, A., Jones, A., Georgila, K., Debevec, P., Swartout, W., Maio, H., and Smith, S.: Time-offset interaction with a holocaust survivor. In: 19th International Conference on Intelligent User Interfaces (IUI), pp. 163–168. (2014).
3. Lee, A., Oura, K., and Tokuda, K.: MMDAgent—A fully open-source toolkit for voice interaction systems. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). pp. 8382–8385. (2013).

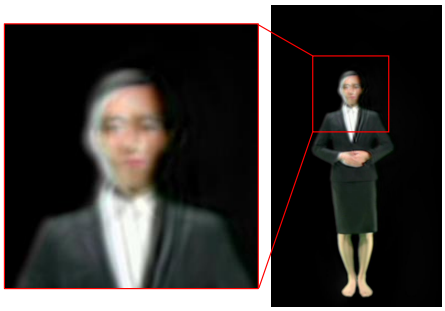


Fig. 10. Noise introduced when generating video sequences of the female avatar used in our subjective assessment.

4. Jones, A., Unger, J., Nagano, K., Busch, J., Yu, X., Peng, H. I., Alexander, O., Bolas, M., and Debevec, P.: An automultiscopic projector array for interactive digital humans. In: ACM SIGGRAPH Emerging Technologies, pp. 6:1. (2015).
5. Inoue, M., Shiraiwa, A., Yoshimura, H., Nishiyama, M., and Iwai, Y.: Evaluating effects of hand pointing by an image-based avatar of a navigation system. In: 20th International Conference on Human–Computer Interaction (HCII), Part II, pp. 370–380. (2018).
6. Nishiyama, M., Miyauchi, T., Yoshimura, H., and Iwai, Y.: Synthesizing realistic image-based avatars by body sway analysis. In: 4th International Conference on Human-Agent Interaction (HAI), pp. 155–162, (2016).
7. Miyauchi, T., Ono, A., Yoshimura, H., Nishiyama, M., and Iwai, Y.: Embedding the awareness state and response state in an image-based avatar to start natural user interaction. IEICE Transactions on Information and Systems E100-D, pp. 3045–3049 (2017).
8. Mark, L. S., Warm, J. S., and Huston, R. L.: Ergonomics and Human Factors: Recent Research. Springer (1987).
9. Wadhwa, N., Rubinstein, M., Durand, F., and Freeman, W.: Phase-based video motion processing. ACM Transactions on Graphics 32. (2013).
10. Kim, J., Eom, G., Kim, C., Kim, D., Lee, J., Park, B., and Hong, J.: Sex differences in the postural sway characteristics of young and elderly subjects during quiet natural standing. Journal of Geriatrics & Gerontology International 10, pp. 192–198 (2010).