Automatic Analysis of Nonverbal Mirroring Communication

Oky Dicky Ardiansyah Prima, Yuta Ono
Graduate School of Software and
Information Science, Iwate Pref. Univ.
Takizawa, Japan
email: prima@iwate-pu.ac.jp,
g236s001@s.iwate-pu.ac.jp

Kumiko Hosogoe, Miyu Nakano
Faculty of Social Welfare,
Iwate Prefectural University
Takizawa, Japan
email: hosogoe@iwate-pu.ac.jp,
g221r007@s.iwate-pu.ac.jp

Takashi Imabuchi
Office of Regional Collaboration,
Iwate Pref. Univ.
Takizawa, Japan
email: t_ima@ipu-office.iwate-pu.ac.jp

***Abstract*—Nonverbal communication plays an important role in social interaction. Mirroring, an action that mimics the nonverbal behavior patterns of their interaction partners, captures the attention of the Human-Computer Interaction (HCI) community. This action can help building rapport with others by making communication more effective and reflective. This study proposes a computer vision-based system that detects mirroring and analyzes the time lag during a face-to-face communication. Our approach consists of the following steps: (1) human pose estimation; (2) hand gestures quantization; (3) action detection based on Dynamic Time Warping (DTW); (4) estimation of mirroring time lag based on the cross-correlation. For this study, we recorded twenty face-to-face communication scenes using an omni-directional video camera with and without mirroring performed by the imitator. Results show that the DTW was able to detect actions having distinct gestures, whereas the cross-correlation was able to estimate the time lags for reactive mimicry of the imitator during the conversation.**

Keywords- mirroring communication; nonverbal communication; human pose estimation; DTW; cross-correlation.

I. Introduction

To improve communication skill, it is important to pay attention to eye contact, gestures, postures, body movements, and voice tones. These nonverbal actions can provide clues, additional information, and meaning in addition to verbal communication. Moreover, using these actions that reflect the behavior of the talking partner can help to create a strong connection with both side during the conversation. These techniques are called nonverbal mirroring. This study extends our previous research on analysis of communication mirroring using vision cameras [1]. Nonverbal mirroring during face-to-face communication can be used to show empathy and positive reaction to counterparts. Nonverbal behavioral mimicry can occur with little or no awareness but can occur during more than 30% of a given interaction [2]. More specifically, nonverbal mirroring can be distinguished from imitation behavior, which is an event in which two people act the same regardless of timing, and complementary behavior, which is an event in which two people act differently [3]. In this paper, we limited the scope of nonverbal mirroring to imitation behavior. The results of this study will complement studies on mirroring facial expressions in the generation of rapport scales.

The areas of the human brain that are activated by observation and execution of the same actions are called the "mirror neuron system." Functional Magnetic Resonance Imaging (fMRI) of frontal and parietal regions of the brain indicated that these regions are most consistently involved during mirroring [4]. For actions that are considered mirroring, each action taken by a partner are reciprocated by the coordinated manner with time lags. Studies have been conducted to define the time lag before mirroring occurs. Hale et al. (2020) suggested that 400-1,000ms is a plausible time range for reactive mirroring in a natural conversation [3]. However, mirroring might happen on a longer timescale, 2-10s [5] or 7s [6] at most.

Traditionally, measuring nonverbal mirroring had to be done manually by annotating of characteristic gestures of the subject and similar gestures of the counterpart from recorded videos of face-to-face communication. The resulting repetitive behavior, its duration, and response latency are quantized and used for further analysis, such as rapport-based behavior analysis [7]. BECO2, an integrated behavioral coding system, is widely used in Japanese universities to train students about behavior coding [8]. The system allows observers to record and analyze the occurrence and duration of actions by pressing the keyboard keys corresponding to each category. Because there is a lot of ambiguity in judgments of specific actions, observers tend to make inconsistent judgments, which reduces the quality of measurements.

The analysis of nonverbal mirroring has been studied for some time, but little research has been done on how to automate the analysis as an alternative to manual coding. To date, there is no practical software application program that can automatically detect mirroring from a video of face-to-face communication and calculate the time difference until the mirroring occurs. Speech and video processing technologies may contribute to the efficient analysis of conversational scenes. On the one hand, speech processing can reveal nonverbal behaviors such as speech, stress acts, and speech rate based on speech signals [9], but on the other hand, video processing can measure facial information and gestures [10].

Since mirroring is a complex phenomenon [11], we made a first attempt to build a framework which quantify gestures from a video of face-to-face communication in order to automatically detect the presence of mirroring [1]. In this paper, we further enhance the framework by improving

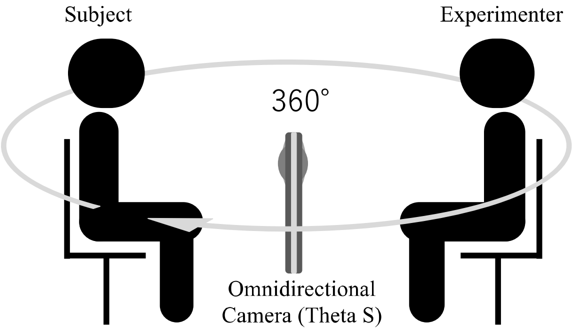


Figure 1. Experimental setting.

gesture detection and adding the ability to estimate the overall time lag of the detected mirroring.

The rest of this paper is organized as follows. Section II presents related research on behavioral coding and gesture analysis using computer vision techniques. Section III introduces the proposed framework for analyzing the presence of communication mirroring. Section IV describes the results. Finally, Section V presents our concluding remarks.

II. Related Work

The development of automatic mirroring detection involves building the mimicry dataset. MAHNOB is a public mimicry dataset consisting of a collection of multisensory audiovisual recordings of fully synchronized naturalistic dyadic interactions. The recordings were made under controlled laboratory conditions using 15 cameras and three microphones to obtain the most favorable conditions possible for analyzing the observed behavior [12]. Bilakhia et al. (2015) applied classifiers such as cross-correlation, generalized time warping, and Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) to face and head movement data in the MAHNOB dataset [13]. However, the mirroring detection performance of these classifiers was poor, suggesting that more advanced learning methods are needed to deal with the variability in the dataset.

Some studies have attempted to detect mirroring using special cameras. Terven et al. (2015) introduced mirroring detection based on head-gestures using computer vision-based wearable devices [14]. A camera embedded in the wearable device was used to detect facial features of the partner during a face-to-face communication. Hidden Markov Models (HMMs) was used to recognize similar head-gestures. However, the mirroring detection performance was affected by the amount of head-gestures that occur in each data case. Jaana et al. (2014) developed an automated behavioral analysis system using a single omnidirectional camera. This system analyzed facial expressions, head nods, utterances based on facial features extracted from the camera [10]. While the system is not specifically designed to detect mirroring, it opens a way to simplify the video recording process during face-to-face communication by using an omnidirectional camera to analyze all participants in a conversation.

Body movements can be automatically accessed using computer vision-based methods, such as Motion Energy Analysis (MEA) [15] and OpenPose [16]. MEA measures motion by counting color changes in successive frames within a predefined region of interest, whereas OpenPose measures key points on the human body, hands, face and feet. Schoenherr et al. (2019) evaluated the performance of various time series analysis methods on nonverbal synchronous data quantified by MEA [17]. Schneider et al. (2019) proposed a gesture recognition system [18] using human posture obtained from a single camera using OpenPose. This system combined Dynamic Time Warping (DTW) and One-Nearest-Neighbor classifier to classify the time series data.

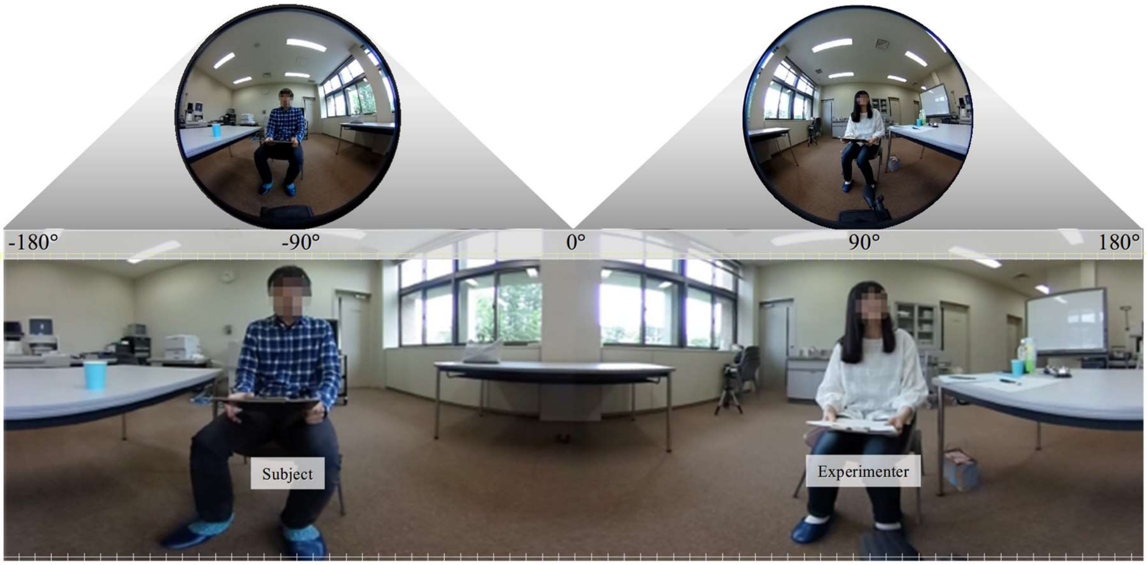


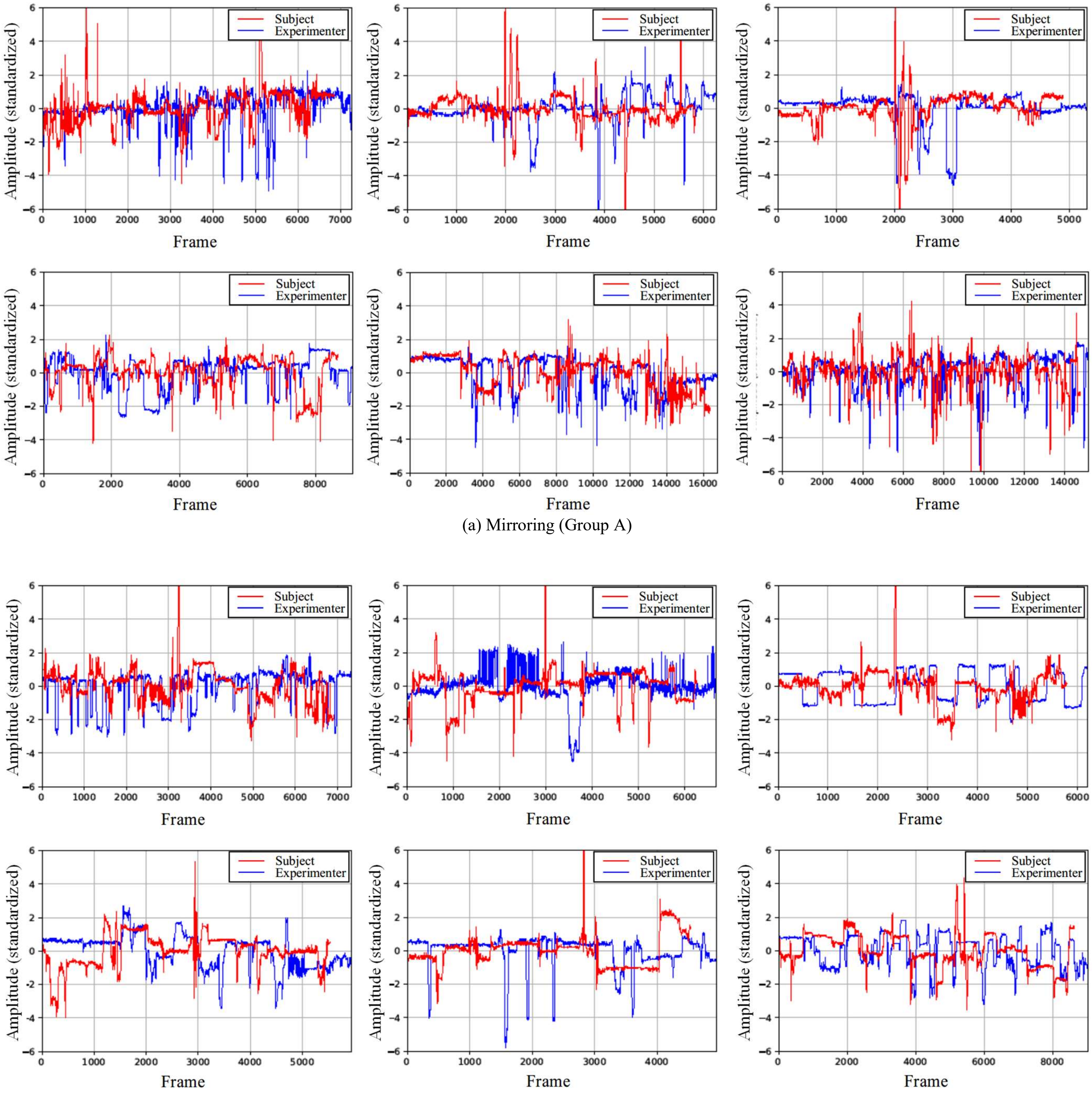
Figure 2. Face-to-face communication captured in a panoramic image

III. Analysis of Nonverbal Mirroring Communication

We propose a method to automatically analyze nonverbal mirroring communication from the recorded movements of pairs of participants (dyads): a subject and an experimenter. Our method uses DTW to detect and classify characteristic movements and uses cross-correlation to estimate the overall time lag of mirroring movements during a conversation. Here, we chose to focus on hand-gesture data only, based on a pilot study that revealed stronger similarities between the dyads than whole-body posture data.

A. Participants

46 students (25 males and 21 females) who have had part-time work experience in multiple faculties at Iwate Prefectural University were interviewed for approximately 5 minutes. The experimenter was a student in the same grade as the subjects and had been fully trained in behavioral mirroring. Subjects, who were not normally close to the experimenter, were recruited through the snowball sampling method.



(a) Mirroring (Group A)

(b) Without Mirroring (Group B)

Figure 3. An example of Wrist data for Group A and B.

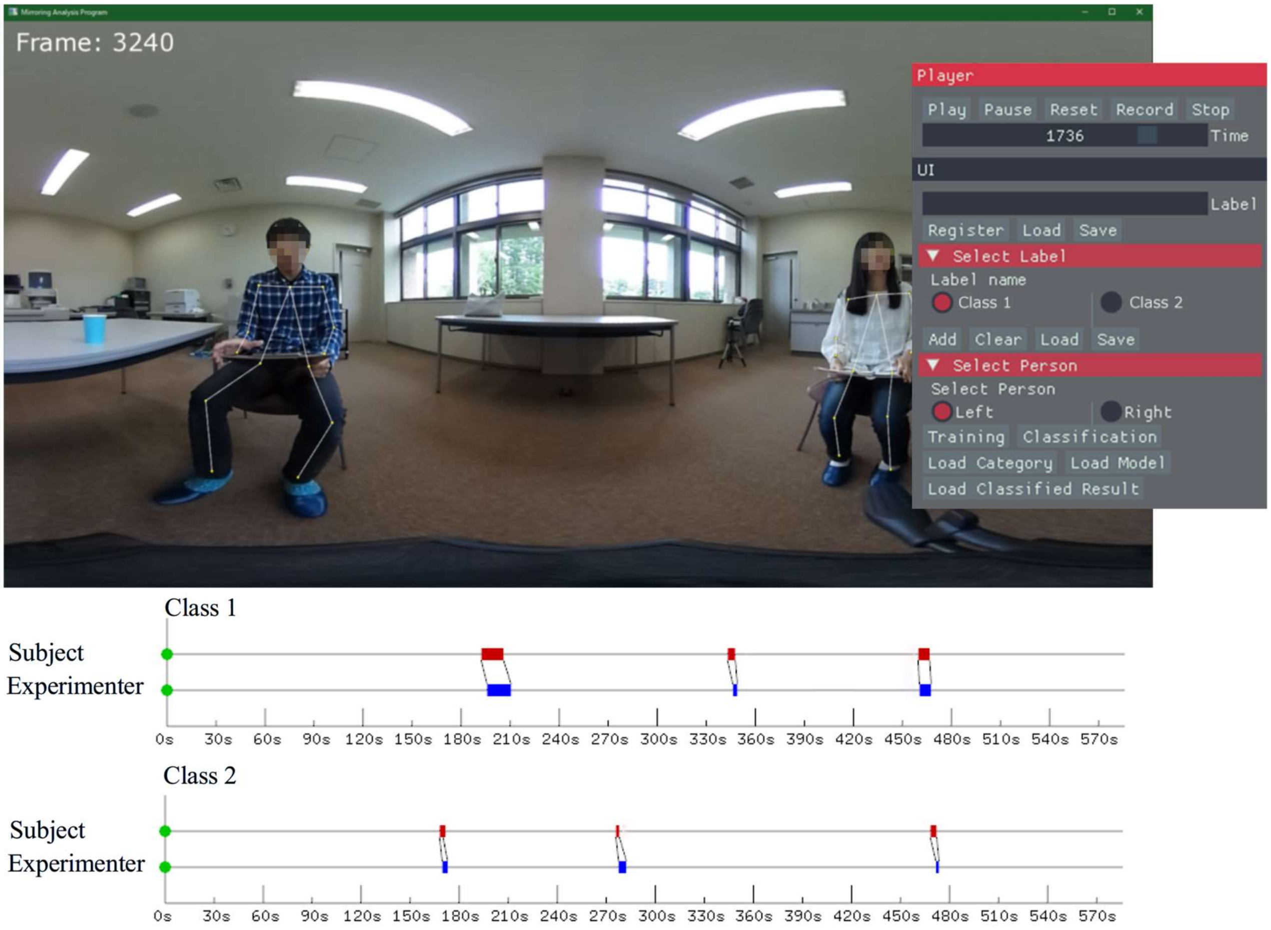


Figure 4. The Graphical User Interface (GUI) created to facilitate the selection and visualization of the training dataset in this study.

During the interview, the subjects were asked about their work experience. The experimenter intentionally mirrored 25 subjects (14 males and 11 females) and did not mirror the remaining 21 subjects (11 males and 10 females). Hereinafter, we refer to the former as Group A and the latter as Group B. This division was done randomly. Six subjects who did not perform the hand gestures were excluded from the analysis. Finally, a total of 40 conversations, 20 in group A (10 males and 10 females) and 20 in group B (11 males and 9 females), were included in the analysis.

B. *Laboratory Setup*

Two chairs were set up in the room facing each other for the subjects and the experimenter, as shown in Figure 1. To simplify the video recording process, an omnidirectional camera (Ricoh Theta S) [19] was placed between the experimenter and the subject, and video recording was performed at 30 Hz. Subjects were seated after completing the informed consent form. Subjects and experimenters were given a clipboard to take the necessary notes on. The interviews were conducted while holding this clipboard. This clipboard restricted the subject's hand gestures and allowed us to efficiently extract only those gestures that are important for communication mirroring. The subject was expected to generate hand gestures with one hand while holding the clipboard in the other hand, or to generate hand gestures with both hands by placing the clipboard on his or her lap.

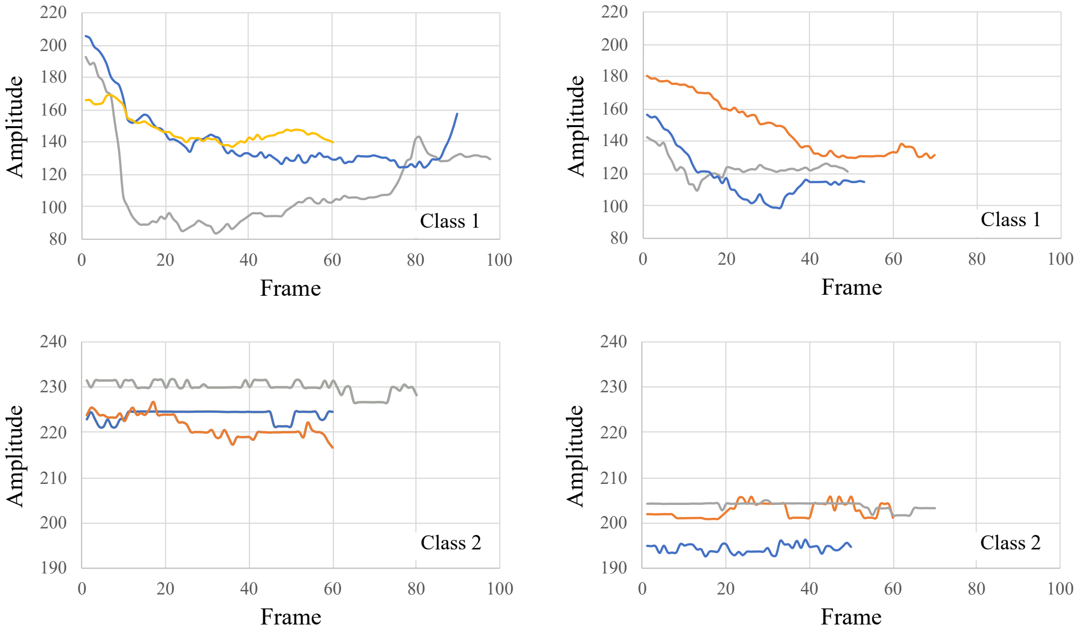
C. *Pre-processing*

1) Panoramic Image Projection

Ricoh Theta S generates two fisheye images to represent a 360° image. We merged and warped these images to produce a panoramic image, as shown in Figure 2, which allows the experimenter and the subject in the image to be seen from the front. The panoramic image is presented as a rectangular image of a 360° image. No special effort was made to produce this image. Everything was done using the built-in Ricoh utility program.

2) Quantitation of Hand Gestures

OpenPose with the 18-keypoint Coco body model was used to estimate the body posture of the dyads. For the purpose of this paper, only the positional information of the wrist joint is extracted from the body posture. The pixel coordinates of each joint were calculated from the panoramic images and these coordinates were normalized with the neck joint as the



(a) Subject

(b) Experimenter

Figure 5. Two classes of gestures created from dyads for this study.

origin. We quantified the wrist data as follows. Let $W_n(x_l, y_l, x_r, y_r)$ represents coordinates of both wrists at n^{th} frame, wrist data at this frame is calculated by

$$Wrist_n = \|W_n\|_2 = \sqrt{x_l^2 + y_l^2 + x_r^2 + y_r^2} \ . \quad (1)$$

Equation (1) reduces the dimensionality of the data at both wrists, but our preliminary results show that this data transformation can eliminate ambiguous gestures that do not correspond to mirroring.

In this study, we use wrist time series data to represent the gesture. Figure 3 shows an example of wrist data for Group A and B, measured in this study. These figures show that the variability of the dyads' wrist data was similar in Group A, but not in Group B. Here, we did not limit the duration of the interviews, which resulted in different frame lengths of the wrist data measured in each interview.

3) Building Training Dataset for DTW

Given that natural mirroring does not require a faithful reproduction of the opponent's hand gestures, we defined two classes of hand gestures based on their movement characteristics. We considered that highly variable gestures are subject to mirroring and less variable gestures are not. Hereinafter, we refer to them as class 1 and class 2, respectively.

A Graphical User Interface (GUI) was created to facilitate the selection and visualization of wrist data ranges for building training dataset, as shown in Figure 4. The slider can be used to determine the onset and offset of each gesture. In addition, when the slider is moved, the frame and the posture of both persons corresponding to the slider are displayed in real time. Training data individually selected from subjects and experimenters are drawn in a timeline. The vertical lines connecting the subject and the experimenter indicates that the subject's gestures are mirrored.

TABLE I. LIKELIHOOD BETWEEN GESTURES OF THE SUBJECT AND THE EXPERIMENTER.

No.	Class Label (Subject)	Predicted Class label (Experimenter)	Maximum Likelihood
1.	1	1	0.599
2.	1	1	0.714
3.	1	1	0.716
4.	2	2	0.912
5.	2	2	0.765
6.	2	2	0.766

One of the most difficult aspects of creating a training dataset is to define the number of classes of gestures in the dataset. After reviewing all recorded videos, none of the subjects generated two-handed gestures with the clipboard on their lap. When generating a gesture, the subject always holds the clipboard with one hand. Interestingly, when the gesture was not generated, the subject continued to hold the clipboard with both hands. Based on these observations, we classified the subjects' gestures into two classes.

Figure 5 shows the wrist data, describing the class 1 and class 2 hand gestures created from the dyads, respectively. Each class contains three time series data. In class 1, the

TABLE II. Results of cross-correlation analysis for 20 mirrored conversations in this study

No.	Time Lags		Maximum Correlation	Critical Value
	Frame	Time (s)		
1.	-79	-2.6	0.476	0.027
2.	-60	-2.0	0.725	0.021
3.	-47	-1.6	0.287	0.029
4.	-67	-2.2	0.433	0.030
5.	-68	-2.3	0.174	0.025
6.	-38	-1.3	0.419	0.023
7.	-147	-4.9	0.715	0.020
8.	-83	-2.8	0.301	0.033
9.	-72	-2.4	0.298	0.029
10.	-97	-3.2	0.489	0.018
11.	-110	-3.7	0.731	0.016
12.	-62	-2.1	0.737	0.023
13.	-21	-0.7	0.315	0.022
14.	-55	-1.8	0.877	0.033
15.	-49	-1.6	0.669	0.017
16.	*-481*	*-16.0*	*0.639*	*0.021*
17.	*-812*	*-27.1*	*0.257*	*0.032*
18.	*933*	*31.1*	*0.312*	*0.030*
19.	*433*	*14.4*	*0.156*	*0.021*
20.	*1,279*	*42.6*	*0.289*	*0.031*

TABLE III. Results of cross-correlation analysis for 20 non-mirrored conversations in this study

No.	Time Lags		Maximum Correlation	Critical Value
	Frame	Time (s)		
1.	12,271	409.0	0.284	0.050
2.	-2,736	-91.2	0.215	0.037
3.	475	15.8	0.686	0.040
4.	1,483	49.4	0.425	0.033
5.	689	23.0	0.241	0.032
6.	219	7.3	0.203	0.032
7.	230	7.7	0.340	0.024
8.	-751	-25.0	0.245	0.033
9.	41	1.4	0.513	0.035
10.	-1,802	-60.1	0.335	0.041
11.	1,505	50.2	0.257	0.030
12.	662	22.1	0.244	0.021
13.	1,032	34.4	0.197	0.023
14.	-2,027	-67.6	0.147	0.034
15.	2,777	92.6	0.160	0.024
16.	1,025	34.2	0.235	0.025
17.	4,210	140.3	0.265	0.077
18.	*-5*	*-0.2*	*0.325*	*0.027*
19.	*-107*	*-3.6*	*0.474*	*0.035*
20.	*-219*	*-7.3*	*0.244*	*0.027*

moment the hand leaves the clipboard was determined to be onset, and the moment the hand returns to the clipboard is to be offset. On the other hand, in the class 2, the onset and offset were determined when the clipboard is held in both hands for a period.

The gestures of the experimenter are relatively shorter than those of the subjects. This can be interpreted as a result of the experimenter confirming the subject's gesture and then simply imitating the gesture.

4) Detection

We performed DTW using the Gesture Recognition Toolkit (GRT) of Gillian and Paradiso (2012) [20] and applied maximum likelihood from the warping distance to estimate similarity to the training data. Table I shows the likelihood between gestures of the dyads shown in Figure 5. This data was normalized before it was inputted into the GRT. The DTW was able to correctly match the same gesture between the dyads, even though the length of the subject's gesture is different from the length of the experimenter's gesture. Although more gesture training data would be desirable, this study focuses on a basic analysis of the extent to which the simplest gestures can be used to detect what appears to be mirroring.

In this study, we used the average length of the training data frame as the maximum warp amount during the calculation of DTW. We considered that this DTW's window width is sufficient for our purpose.

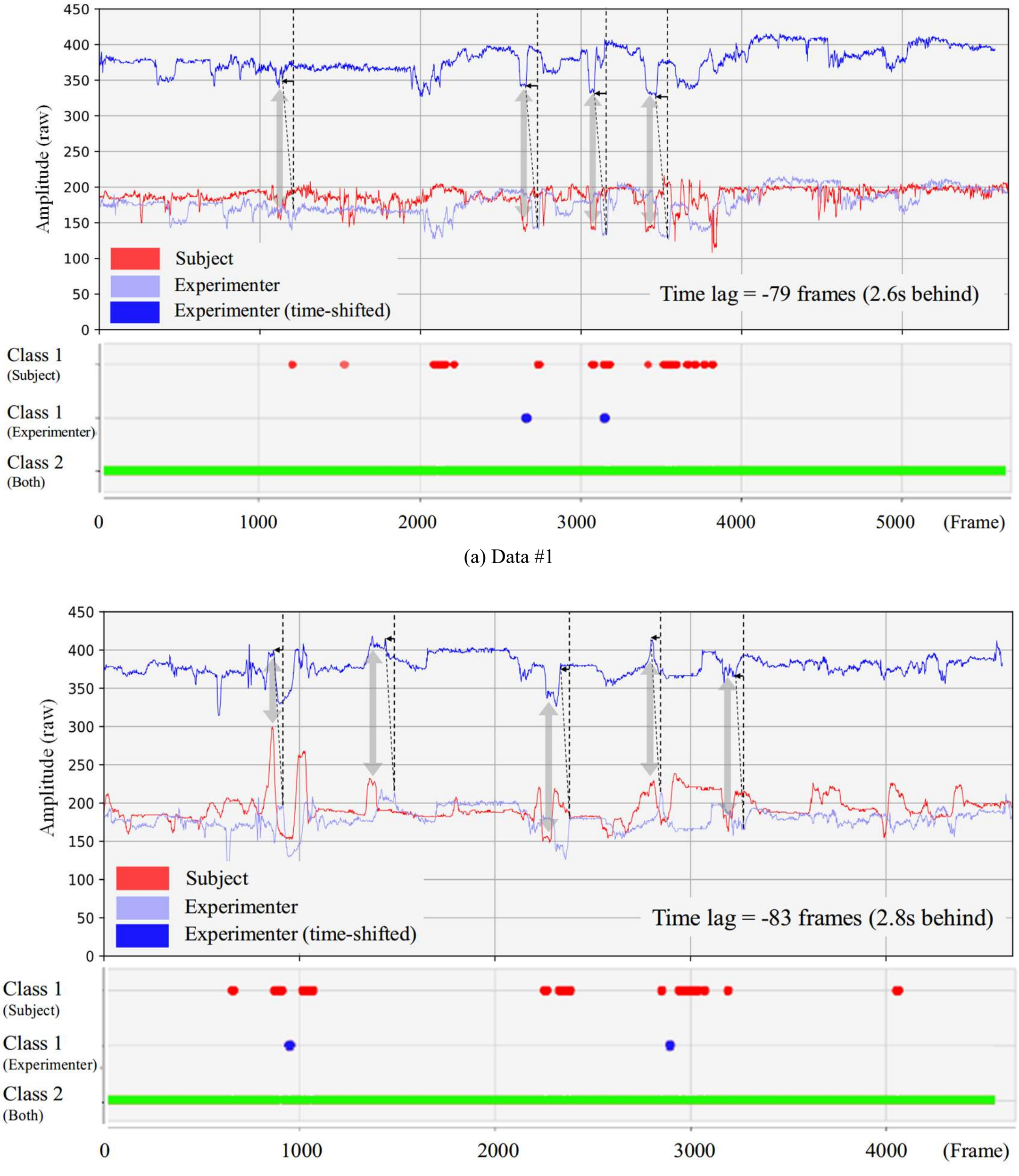
5) Cross-correlation

*C*ross-correlation is useful for aligning two time series, one of which is lagged relative to the other, since its peak occurs at the lag where the two time series are most correlated. Cross-correlation ρ at delay d between the subject's and the examiner's hand gestures is define as

$$\rho(d) = \frac{\sum_{i=1}^{N}\{(x_i - \bar{x})(y_{i-d} - \bar{y})\}}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_{i-d} - \bar{y})^2}} \qquad (2)$$

Here, x_i and y_{i-d} are series of the subject's and the examiner's hand gestures, respectively. $\bar{x}$ and $\bar{y}$ represents means of x_i and y_{i-d}.

Following [3] and [5], we assume that the mirroring occurs at around 0.4-7s. In other words, if a time lag longer than 7s is calculated, it means that no mirroring has occurred in the data. From this perspective, cross-correlation can be the easiest way to determine the presence or absence of mirroring.



(a) Data #1

(b) Data #8

Figure 6. Detection and cross-correlation of the two classes in the two mirrored conversations.

IV. Results

A. Mirroring Analysis

For automatic detection of two classes, we used only training data collected from subjects, as shown in Figure 5(a). The experimenter's gestures described in the previous section were only for validating the subject's gestures. The behaviors of subjects and experimenters corresponding to each class were collected and presented in a time series of the input data.

Cross-correlations were calculated for the subject's and experimenter's wrist data. Here, we normalized the values of the cross-correlation to take values from -1 to +1. The maximum value of the cross-correlation function indicates the point in time where the data are most aligned (delay time).

Figure 6 shows the results of the detection and cross-correlation of the two classes in the two interview scenes. The top of the figure shows the raw data of the dyads' wrist movements. The experimenter's data were then shifted based on the time lag between dyads as measured by cross-correlation. To make the shifted data easier to observe, the shifted data was moved upward. The gray arrows indicate some of the areas where the subject's wrist data and the shifted experimenter's wrist data are similar. The bottom row of the figure shows two classes of detection from the dyads' wrist

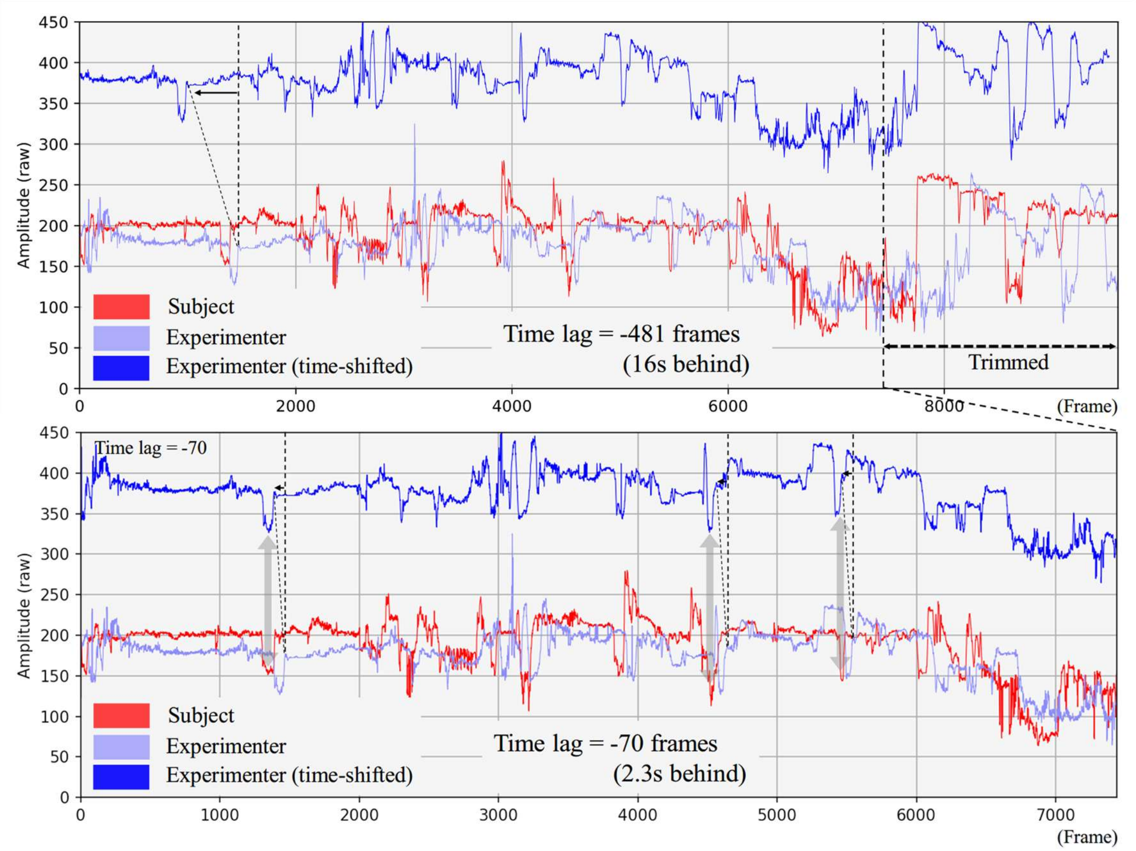


Figure 7. Refinement of the cross-correlation result by data trimming (data #16).

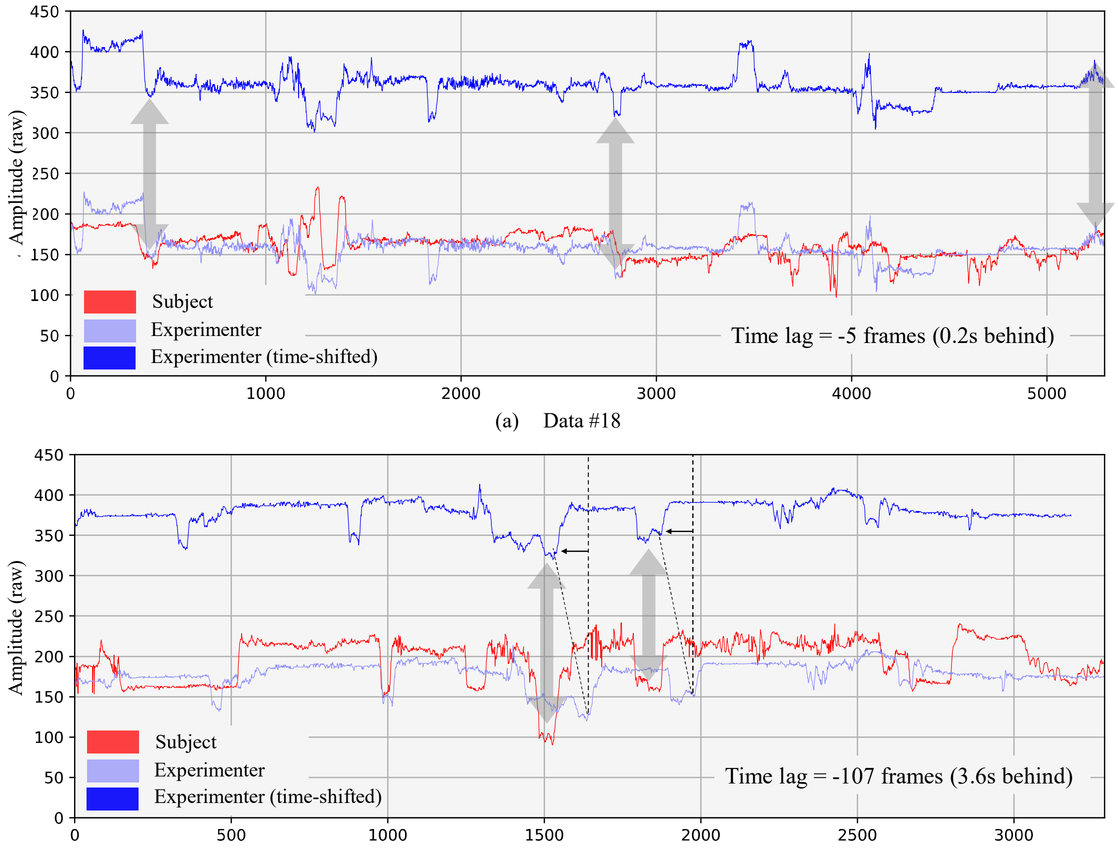
data. Red points represent class 1 detections for the subject, whereas blue points for the experimenter. Given that these detected gestures have similar features, it is likely that these blue dots mirror the red dots. As can be seen from the figure, the number of class 1 detections of the subjects was higher than that of the experimenters. It is a reasonable result since the incidence of the mirrored gestures should not exceed the mirrored target. It should also be noted that the class 1 behaviors detected from the experimenter occurred after the same class of behaviors detected from the subjects. Green dots represent class 2 detections of the dyads. Gestures that could not be classified as class 1 were classified as class 2, resulting in a higher frequency of class 2 detections. In this mirroring analysis, we did not target class 2, which has few variable gestures, so we integrated the resulting detections of class 2 of the dyads, as shown in Figure 6.

Table II shows results of cross-correlation analysis for 20 mirrored conversations in this study. In the first 15 conversations, we found that the experimenter's wrist data are delayed between 0.7-4.9s than the subject's wrist data. However, the remaining five conversations measured a time lag that did not meet the mirroring condition (numbers in italic). These are partly due to the inability of the experimenter to mirror the subject properly during conversations, but also due to the increased variation in behavior, such as having a drink during a conversation. Figure 7 shows that the measurement delay of conversation #16 can be corrected from 481 frames (16s) to 70 frames (2.3s) by trimming a portion of the data.

Table III shows results of cross-correlation analysis for 20 non-mirrored conversations in this study. In the first 17 conversations, the measured time lag showed that no mirroring occurred during the conversation. However, the time lag in the remaining three conversation scenes suggests that mirroring behavior occurred (numbers in italic). Figure 8 shows that similar patterns between the dyads in conversation #18 and #19. This implies that the experimenter may unconsciously engage in mirroring behavior.

B. *Perceived Empathy*

Perceived empathy was measured using the 16-item empathy understanding subscale of the Barrett-Lennard Relationship Inventory (BLRI) [21]. All items in the perceived empathy assessment were scored on a six-point Likert scale, with 1 = strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = slightly agree, 5 = agree, and 6 = strongly agree. Higher scores represented more empathy perceived by subjects. Cronbach’s coefficient alpha is 0.81, indicating high reliability.



(a) Data #18

(b) Data #19

Figure 8. Similar patterns between the dyads in the non-mirroring conversations.

TABLE IV. MEAN AND STANDARD DEVIATION OF THE PERCEIVED EMPATHY SCORES

GENDER	WITH MIRRORING	WITHOUT MIRRORING
MALE	4.37 (0.52)	4.23 (0.49)
FEMALE	4.77 (0.48)	4.65 (0.37)

To evaluate whether there was a difference in the mean score on the perceived empathy scale by gender and the presence of mirroring, we conducted a two-factor analysis of variance. The means and standard deviations are shown in Table IV. The results showed a significant main effect on the gender factor ($F(1, 36) = 8.04$, $p < 0.5$), but not on the presence of mirroring factor ($F(1, 36) = 0.72$, *n.s.*). No significant interactions were observed ($F(1, 36) = 0.01$, *n.s.*). The present results indicate that female subjects were more emphatic with and without mirroring during the interview. To promote empathy by mirroring communication, a longer interview than the present experiment would be necessary.

C. Debriefing

Once the interview was completed, the experimenter provided the subject with accurate and pertinent information about the nature of this experiment. During the debriefing process, subjects are informed about what the hypothesis for the experiment was as well.

After the debriefing, the 20 subjects who had been mirrored were asked if they noticed that they were being mirrored or if they felt that the conversation was unnatural. Six subjects said that they noticed that they were being mirrored. They were aware of being mirrored because they were already knowledgeable about mirroring. All 20 subjects did not find it unnatural to be mirrored. In this short interview, the recorded video shows that there is almost no pause in the conversation between the two parties, with the subject actively answering the experimenter's questions. This may explain why the subjects did not feel unnatural in their conversations.

V. CONCLUDING REMARKS

Many studies have been done to automatically detect mirroring during conversations, but to the best of our knowledge, it has not reached a practical level [13][14][22]. In order to automatically detect mirroring behavior during conversations, we attempted to make the gestures on the dyads as simple as possible. Having the subjects hold the clipboard during the conversation was effective in limiting their hand gestures. This allowed the experimenter to properly imitate the subjects' gestures.

The DTW and cross-correlations used in this study are commonly used in time series data analysis. In general, a mimic gesture is one in which the performer tries to mimic the actions of the other person as accurately as possible. However, there is a difference between similar behaviors perceived by humans and similar behaviors seen from sensor information. For this reason, we successfully identified distinct gestures using DTW without being too obsessed with small movements by making the information of the hand gestures one-dimensional using the L2 norm. On the other hand, cross-correlation analysis successfully estimated the time lag of mirroring behavior in conversations. Interestingly, cross-correlation analysis was able to detect the unintended mirroring even if the experimenter did not intend to mirror the subject.

Nonverbal mirroring communication helps to create a strong connection between the two parties during a conversation, but in the present experiment, subjects' empathy levels were not significant with or without mirroring communication. A longer interview than the present experiment would be necessary to promote empathy through mirrored communication.

Finally, since the observed data contain a lot of noise, removing the noise can improve the accuracy of the analysis. The mirroring time lag can be properly determined by trimming a portion of the data, as in this study.

Acknowledgment

We thank the faculty of Social Welfare, Iwate Prefectural University, Japan, for funding this project. We also thank the editor and four anonymous reviewers for their constructive comments, which helped us to improve the manuscript.

References

[1] K. Hosogoe, M. Nakano, O. D. A. Prima, and Y. Ono, “Toward automated analysis of communication mirroring,” The Thirteenth International Conference on Advances in Computer-Human Interactions, ACHI2020, pp. 15-18, 2020.

[2] J. L. Lakin and T. L. Chartrand, “Using nonconscious behavioral mimicry to create affiliation and rapport,” Psychological Science, 14(4), pp. 334–339, 2003.

[3] J. Hale et al., “Are you on my wavelength? Interpersonal coordination in dyadic conversations,” Journal of Nonverbal Behavior, 44 (1), pp. 63-83, 2020.

[4] P. Molenberghs, R. Cunnington, and J. B.Mattingley, “Is the mirror neuron system involved in imitation? A short review and meta-analysis,” Neuroscience and Biobehavioral Reviews, 33 (7), pp. 975-980, 2009.

[5] N. P. Leander, T. L. Chartrand, and J. A. Bargh, “You give me the chills: Embodied reactions to inappropriate amounts of behavioral mimicry,” Psychological Science, 23(7), pp. 772–779, 2012.

[6] J. W. Robinson, A. Herman, and B. J. Kaplan, “Autonomic responses correlate with counselor–client empathy.” Journal of Counseling Psychology, 29(2), pp. 195–198, 1982.

[7] C. F. Sharpley, J. Halat, T. Rabinowicz, B. Weiland, and J. Stafford, “Standard posture, postural mirroring and client-perceived rapport.” Counselling Psychology Quarterly, 14(4), pp. 267–280, 2001.

[8] Behavior coding system, DKH Co. Ltd., https://www.dkh.co.jp/product/behavior_coding_system/ [retrieved: August 31, 2020]

[9] K. Otsuka and S. Araki, “Audio-visual technology for conversation scene analysis,” NTT Technical Review, 7(2), pp. 1-9, 2009.

[10] Y. Jaana, O. D. A. Prima, T. Imabuchi, H. Ito, and K. Hosogoe, “The development of automated behavior analysis software,” Proc. SPIE 9443, Sixth International Conference on Graphic and Image Processing (ICGIP), pp. 1-5, 2014.

[11] T. L. Chartrand and J. A. Bargh, “The chameleon effect: the perception-behavior link and social interaction,” Journal of Personality and Social Psychology, 76(6), pp. 893–910, 1999.

[12] MHI-Mimicry database, https://mahnob-db.eu/mimicry/ [retrieved: August 31, 2020]

[13] S. Bilakhia, S. Petridis, A. Nijholt, and M. Pantic, “The MAHNOB mimicry database: a database of naturalistic human interactions,” Pattern Recognition Letters, 66, pp. 52–61, 2015.

[14] J. R. Terven, B. Raducanu, M. E. Meza-de-Luna, and J. Salas, “Head-gestures mirroring detection in dyadic social interactions with computer vision-based wearable devices,” Neurocomputing, 175, pp. 866–876, 2015.

[15] K. Grammer, M. Honda, A. Juette, and A. Schmitt, “Fuzziness of nonverbal courtship communication unblurred by motion energy detection,” Journal of Personality and Social Psychology, 77 (3), pp. 487-508, 1999.

[16] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields,” Computer Vision and Pattern Recognition, pp. 7291-7299, 2017.

[17] D. Schoenherr et al., “Quantification of nonverbal synchrony using linear time series analysis methods: Lack of convergent validity and evidence for facets of synchrony,” Behavior Research Methods, 51(1), pp. 361–383, 2019.

[18] P. Schneider, R. Memmesheimer, I. Kramer, and D. Paulus, “Gesture recognition in RGB videos using human body keypoints and dynamic time warping,” Lecture Notes in Computer Science, vol. 11531, pp. 281–293, 2019.

[19] Ricoh Theta S, https://theta360.com/en/about/theta/s.html [retrieved: August 31, 2020]

[20] N. Gillian and J. A. Paradiso, “The gesture recognition toolkit,” Journal of Machine Learning Research, 15, pp. 3483–3487, 2014.

[21] G. T. Barret-Lennard, “Dimensions of therapiat responses as causal factors in therapeutic change,” Psycological Monographs, 76 (43), pp. 1-36, 1962.

[22] S. Michelet, K. Karp, E. Delaherche, C. Achard, and M. Chetouani, “Automatic imitation assessment in interaction,” Lecture Notes in Computer Science (Included in Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7559 LNCS, pp. 161–173, 2012.