Ping-Pong Robot to Control Motivation of a Human Player

NAKAYAMA Masamune, KURISU Takanori, MIZUNO Yuta, MIYAKE Youichiro and YASE Satoshi

We are developing a ping-pong robot called FORPHEUS that can keep a table tennis rally going and interact with people to generate an appeal for harmonization, which is the future relationship between humans and machines. Although the performance of the table tennis robot has been improving year by year, a player's motivation to maintain a rally tended to decrease because of the tendency for rallies to be monotonous.

Therefore, we propose an interaction system to control the player's motivation to continue a rally. In our work, the table tennis robot can measure a player's motion and vital signals to estimate skill and emotion. In addition, we also implemented a ball-return plan to make the player feel comfortable and focused on using the Meta AI of Square Enix. This system controls the motivation to continue a rally by more than 80% of players and realizes harmonization on teams of bringing out the players' maximum capabilities and promoting their growth.

1. Introduction

Under the philosophy championed by Kazuma Tateishi, OMRON's founder, which goes, "Man should leave what machines can do to machines and enjoy activities in more creative areas," we have considered the man-machine relationship. We expect the man-machine relationship to change in step with social and technological changes and go through the substitution, collaboration, and harmony stages in that order. The substitution stage means the state in which machines perform tasks conventionally carried out by human hands. Meanwhile, the collaboration stage refers to a state in which machines perform tasks together with humans to suit the latter's purposes. Finally, the harmony stage is a state in which machines understand human intentions and assist and allow humans to perform more creative activities.

We have developed FORPHEUS, a ping-pong robot able to continue rallying with a human player, to spread the concept of the future man-machine relationship or harmony that OMRON envisions[1)-3)]. Following the start of its development in 2013, our first-generation ping-pong robot made its debut at the open public exposition for the 2014 CEATEC Japan. Since then, this robot has been evolving every year through development efforts to add new functions and improve its performance.

We consider that the following two types of technologies are necessary to build a ping-pong robot embodying man-machine harmony:

(1) Technologies for robots to perform ping-pong tasks
   E.g., high-speed, high-accuracy ping-pong ball measurement technology; and
   high-speed, high-accuracy robot control technology
(2) Technologies for understanding humans and intervening with them
   E.g., motion analysis technology,
   emotion estimation technology, and
   human-machine interaction technology

The technologies listed in (1) have long been research subjects in computer vision, robotics, and other related fields. For the ones listed in (2), studies are underway in a wide range of disciplines, including computer vision, cognitive psychology, and entertainment. What is considered of particular importance to cause human behavioral changes is to induce motivation. As a method of achieving this purpose, gamification is attracting attention.

For people to give full play to their performance and accelerate their growth, we consider it necessary to keep them highly motivated. For conventional ping-pong robots, technology developments have been promoted, focusing on the technologies given in (1), to enable robots to exchange rallies with human players varied in skill level from beginner to advanced. In 2016, our robot was equipped with a function for determining human players' skill levels based on their motions and for exchanging rallies accordingly. As a result, however, monotonous rallies often occurred and reduced the motivation of human players.

Square Enix implements a technology called Meta AI in games to offer the fun of games designed to change opponent characters' behaviors depending on the player's emotion[4],[5]. For example, the Meta AI implemented in Final Fantasy® XV can monitor the player's status and those of the associates and dispatch an associate optimal for assistance from those nearby not engaged in a battle when the player is in a fix[6]. In this way, the Meta AI can recognize and change the whole game situation to stimulate the player's emotion. Thus, it can keep the player motivated without making the person bored.

We considered the possibility whether a ping-pong robot equipped with Meta AI could have motivation control (Moti-Ctrl) over the player during a ping-pong rally. The Meta AI is an algorithm built for an ideal environment of a game, which is relatively free from external noise. On the other hand, ping-pong players' motions, which are vigorous and cause strong external disturbances, posed a challenge to the Meta AI's direct implementation.

To provide a Moti-Ctrl function through the Meta AI's implementation[7], we developed a technology for determining the detailed human skill level of human players based on their motions and another technology for obtaining human players' vital data without making contact and estimating their emotions. This paper presents a report on these technologies.

In what follows, Section 2 describes the configuration of our ping-pong robot, Section 3 presents the verification experiment we performed to evaluate this Moti-Ctrl function, and finally, Section 4 presents the conclusions and future prospects.

2. Technologies required for ping-pong rallies

Our ping-pong robot consists mainly of an industrial robot built in-house and general-purpose equipment, such as our proprietary cameras, to promote our products and technologies (Fig. 1):
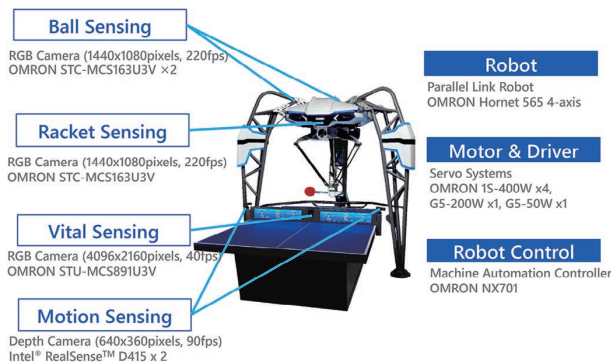

Ball Sensing
RGB Camera (1440x1080pixels, 220fps)
OMRON STC-MCS163U3V ×2

Racket Sensing
RGB Camera (1440x1080pixels, 220fps)
OMRON STC-MCS163U3V

Vital Sensing
RGB Camera (4096x2160pixels, 40fps)
OMRON STU-MCS891U3V

Motion Sensing
Depth Camera (640x360pixels, 90fps)
Intel® RealSense™ D415 x 2

Robot
Parallel Link Robot
OMRON Hornet 565 4-axis

Motor & Driver
Servo Systems
OMRON 1S-400W x4,
G5-200W x1, G5-50W x1

Robot Control
Machine Automation Controller
OMRON NX701

Fig. 1 Configuration of the ping-pong robot

The following subsections present a measurement technology for ping-pong ball (hereafter "ball") positions, human player's motions, and racket positions and attitudes, future trajectory prediction technology for the ball, and a robot's return shot-related motion generation technology as the technologies required for the ping-pong robot to play ping-pong with human players.

2.1 Measurement of ball positions, human motions, and racket positions and attitudes

For ball position measurement, the robot uses two industrial RGB cameras (STC-MCS163U3V, OMRON) with a resolution of 1440×1080 pixels and a frame rate of 220 fps. The cameras are mounted on the left and right sides of the robot's head to include the ping-pong table's whole area in the field of view. RGB images captured by the two cameras undergo conversion into HSV color space images robust to lighting variations followed by binarization and noise removal. Then, the cameras' internal parameters obtained beforehand are used to correct lens distortions. After then, the epipolar constraint, which is a relation holding between the same points in images taken of a subject from different angles, is used to narrow down the ball's candidate points in the left and right images and identify its center of gravity. Finally, the ball's 3D position is calculated based on the stereo method using the cameras' external parameters[8].

For human motion measurement, the robot is equipped with two depth cameras (RealSense D415, Intel) with a resolution of 640×360 pixels and a frame rate of 90 fps. The human player's motions in ping-pong are large and tend to cause self-occlusion, or the partial concealment of one of one's body parts behind another, to occur in images taken by a single camera. Therefore, a camera is installed at either end of the ping-pong table net to capture images of the human ping-pong player's upper body from two different points of view. To each depth image captured by the two cameras, skeleton estimation middleware (Nuitrack, 3DiVi) is applied to obtain the 3D positions and the reliability of 12 positions in the upper-body skeleton[9]. Then, based on each skeletal position's reliability, the human player's dynamic model constraints are added to calculate the skeletal position with high accuracy[10].

For the human-held racket's position and attitude measurement, the robot relies on an industrial RGB camera (STC-MCS163U3V, OMRON) with a resolution of 1440×1080 pixels and a frame rate of 220 fps. This camera is mounted at the center of the robot's head to capture images from diagonally above the human player. The racket is affixed with nine marker seals to calculate the center-of-gravity position using the same

method as in measuring that of the ball. With a small number of markers, planar attitude estimation becomes ambiguous. Hence, a perspective-n-point (PnP) problem with the degree of freedom reduced by removing the rotational components around the normal to the racket surface is solved to obtain a unique racket's position and attitude[11].

2.2 Ball trajectory prediction

In ping-pong, the ball moves so fast that a robot controlled adaptively to the ball's movement cannot respond in time. Therefore, after ball position measurement, the ball's future trajectory must be determined to predict the robot's ball hitting position, and the robot must move to that position. The ball motion during a rally can be approximated using a model of aerodynamic forces acting around a spherical body traveling in the air. Our previously developed technique is used to calculate the ball speed and spin speed from the change in the ball's time-series position measured. Besides, a collision model for describing the collision between the ball and the ping-pong table is used to predict the post-collision ball trajectory[3].

2.3 Robot's return-shot planning

The ball's future trajectory serves as the basis for determining the ball hitting point with the robot's operational range taken into consideration. To the preset target ball return position, speed, and spin speed for the human player's side of the ping-pong table, the aerodynamic model used in Subsection 2.2 is applied to simulate the backward evolution of time and obtain the speed and spin to be achieved by the ball immediately after being hit by the robot. Then, a ball-racket collision model based on these data and the ball speed and spin immediately before getting hit is used to calculate the racket speed and attitude at the time of ball hitting[3].

3. Technologies required to provide Moti-Ctrl

We developed the Moti-Ctrl function, aiming for the robot to make the human player comfortable and more concentrated through interactions to keep the person motivated to continue rallying. This function first estimates the human player's skill level from the ball and skeleton data. The function calculates the limit speeds of the shots that the human player can return with forehand or backhand. The Moti-Ctrl function also performs emotion estimation from the vital data available from the ball's video images and those of the player's face. Based on Russell's circumplex model of affect[12], the pleasure/displeasure scale score and the arousal/non-arousal scale score are calculated. Finally, the values calculated through the skill level estimation and emotion estimation are used to determine the

position and speed of the ball return from the ping-pong robot based on the concept of the Meta AI (Fig. 2):
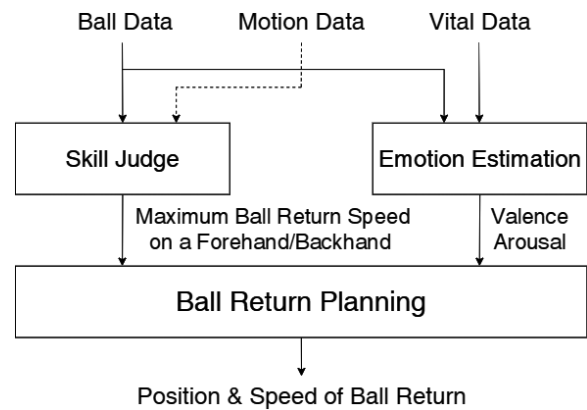

Fig. 2 System flow of the Moti-Ctrl function

The following subsections present a technology for estimating human players' ping-pong skill levels, another technology for contactless obtaining human players' vital data, yet another technology for estimating human players' emotions, and a return-shot planning technique for keeping human players motivated as the technologies required to provide the Moti-Ctrl function.

3.1 Skill level estimation

To enhance human ping-pong players' motivation to continue rallying, the robot must first quickly adjust the degree of rally difficulty (ball speed, course, spin speed, randomness, etc.) to suit each trial player's ping-pong skill level. In 2016, a deep learning-based rally level adjustment function was implemented in our ping-pong robot. This function, however, depended on the designer's subjective view for annotations and posed the problem of failing to set an appropriate rally difficulty level for each trial player. Hence, for the skill level estimation function developed this time, the degree of difficulty of the robot's return shot bordering on whether a trial player can manage to return a shot is defined as the objective reference index. Our aim this time is to estimate this degree of difficulty from a small number of ping-pong rallies. Considering that many of the trial players are beginners this time, we pay attention to the basic difficulty parameters of ball return speed and ball return course to estimate the robot's maximum ball return speed that allows more than a certain degree of probability for the trial player to return a shot. This estimation is to be performed for both forehand and backhand shots.

The trial play time available in a demonstration space at an exposition is short. Accordingly, skill level estimation must be performed based on a small number of ping-pong rallies.

Attention is then paid to the motion data, especially the skeleton data obtained in Subsection 2.1, to perform this estimation successfully based on a prior learning method. Fig. 3 shows the flow of the skill level estimation process. First, the trial player's swing motion is segmented to determine by the motion classification method whether it is a forehand or backhand swing. Next, for each swing type, the maximum ball return speed is estimated based on the learning method.
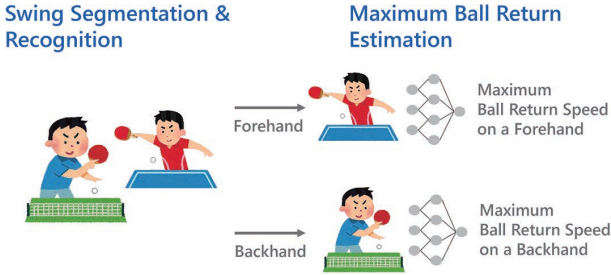

Fig. 3 System flow of skill level estimation

The ball position time-series data obtained in Subsection 2.1 are used for swing motion segmentation. The duration from the time of the robot's return shot reaching the human player's side of the ping-pong table until that of the human player's return shot reaching the robot's side of the ping-pong table is defined as the duration of a single swing from which to extract a series of human player's skeletal positions. For motion classification, a hidden Markov model (hereafter "HMM")-based learning method is used based on its past track record[13] and its upgradability in future developments. From among the skeleton data obtained from the swing motion segmented this time, attention is paid to the time-series data x of the three 3D position vectors in the waist-to-shoulder, waist-to-elbow, and waist-to-wrist directions where major motions occur during a rally. The skeleton data of various trial players' forehand and backhand swings are obtained beforehand to retain an HMM for each swing type (M_{Fore} and M_{Back}). The HMM to be used is the left-to-right type HMM effective at recognizing unidirectionally changing time-series data. When the swing time-series skeleton data x for a new trial player are obtained, whether the swing is forehand or backhand is determined using the following equation where $s = \{Fore, Back\}$:

$$\text{Swing} = \arg \max_{i \in s} P(x \,|\, M_i) \qquad (1)$$

Next, an estimation is performed of the robot's maximum ball return speed that allows more than a certain degree of probability for the trial player to return a shot onto the table. This estimation is performed for each of the forehand and backhand types of swings made by the trial player. As the first step, the maximum ball return speed is defined. Fig. 4 shows the

data of a ping-pong rally in which the shots returned by the robot at various speeds in various courses are hit back by a trial player aiming at the target at the center of the opponent's side of the table. The maximum ball return speed is calculated based on the relative error between the target and the point of impact. More specifically, as shown in Fig. 4, the pass/fail of each return shot is determined based on the error threshold set as 0.75 m from the target. From the data for approximately 20 points near the target, the success probability for the trial player's return shot for each of the robot's ball return speeds is calculated every 0.5 m/s to obtain a graph for the success rate for the trial player's return shots, such as the one in Fig. 5. The robot's maximum ball return speed for when the success rate for the trial player's return shots is below 85% is defined as the maximum ball return speed (Fig. 5).
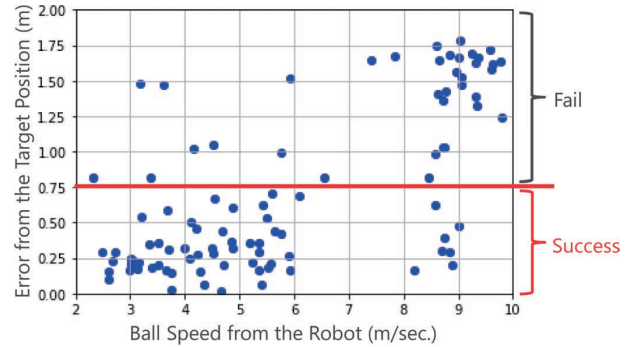

Fig. 4 Data collection results and return-shot pass/fail determination
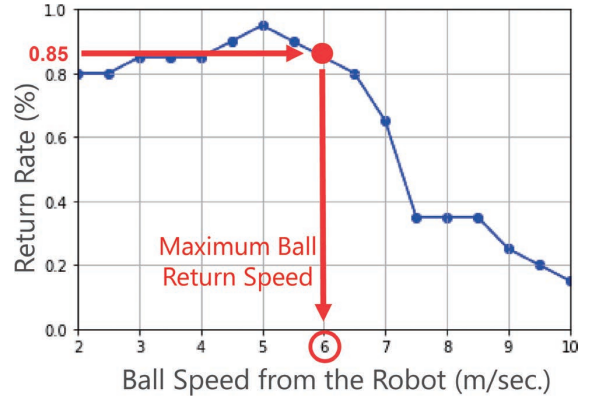

Fig. 5 Calculation method for the maximum ball return speed

As shown above, the maximum ball return speed can be determined correctly only when actual rally results are available from many shots of various speeds. Expositions, however, cannot be expected to afford sufficient time for a trial player to return so many shots. Accordingly, this time, we aim for the robot to learn beforehand the relationship between a single swing motion and the maximum ball return speed and estimate the maximum ball return speed from a single swing motion

without requiring a new trial player to hit many shots.

The prior learning is performed according to the following steps:

The first step is to obtain data from many participants. This time, we collected prior data from 30 healthy individuals aged 18 to 50 years old (including five women). This group can be broken down by ping-pong skill level into five super-advanced-level participants currently active as table tennis club members, five advanced-level participants formerly active as table tennis club members, 12 intermediate-level participants occasionally playing ping-pong as a hobby or for other purposes without belonging to a table tennis club, and eight beginner-level participants with almost no experience in table tennis. Each participant was asked to perform a ping-pong rally task to allow ball and skeleton data collection. In this task, the participants hit back the robot's return shots at various speeds in various courses, aiming at the target at the center of the opponent's side of the table. The second step is to build a regression learning model producing the maximum ball return speed as the output from the input data consisting of the time-series skeleton data of each of the trial players' swings; this model is built using a neural network for each of the forehand and backhand types of swings. The input data to be used are the time-series skeleton data likely to better reflect skill levels, i.e., the 3D time-series data on the shoulder-to-elbow and elbow-to-wrist positions and the time-series data on the shoulder and elbow joint angles. Besides, these data are subjected to low-pass filtering for noise removal, skeleton normalization for smoothing physique differences among trial players, standardization for removing the effects of swing differences due to the differences in the position of the ball return from the robot, and time normalization for entering inputs into the neural network. Assuming that the maximum ball return speed remained unchanged during data collection, labeling was performed so that all the output data were same for the intrapersonal multiple input data (Table 1). These input and output data were used for the prior learning of the regression model.

Table 1 Relationship between the input and output data

Person	A			B			...
Data Pair	1	...	10	11	...	20	...
Ball Speed from the Robot	3.0m/sec.	...	8.5m/sec.	4.5m/sec.	...	7.0m/sec.	...
Swing Skeleton	Swing A		Swing J	Swing K		Swing T	...
Maximum Ball Return Speed	8.0m/sec.			3.0m/sec.			...

Considering its use in a neural network, the dataset obtained from 30 participants is small. Hence, an approximately 16-node, three-layer, fully connected model with a small number of parameters is used so that the input layer takes a form that allows parallel entry of all time-series data while the output layer takes a form capable of producing single-value outputs as the regression results for the maximum ball return speed. For the actual estimation, segmented time-series skeleton data provide the input to produce the output as the maximum ball return speed during each rally. The average value of the outputs from the most recent ten rallies is calculated for each of the forehand and backhand types of swings as the estimated value of the maximum ball return speed. Note that the function presented above estimates the maximum ball return speed based on the data that assume that the trial player returns shots, aiming at the center of the ping-pong table. Hence, this function cannot apply to trial players that hit smash shots for no reason or intentionally return shots at the table's corners.

3.2 Contactless vital data measurement

For vital data measurement, our robot relies on an industrial RGB camera (STC-MCS891U3V, OMRON) with a resolution of 4096×2160 pixels and a frame rate of 40 fps. This camera is mounted at the center of the ping-pong table net to capture frontal images of the human player's face. The video images of the human player's face are used to estimate facial expressions, blink rate, and heart rate contactlessly. Fig. 6 shows the whole flow:
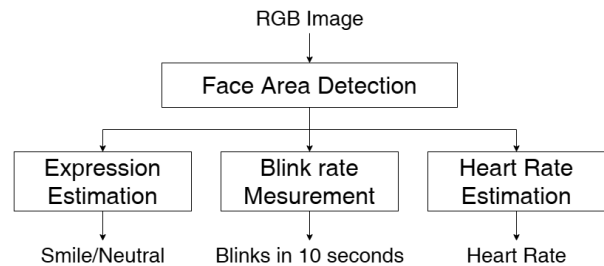

Fig. 6 System flow of vital data measurement

The OKAO Vision[14] middleware for image-based face recognition is used to obtain the face region, the smile degree, the seriousness degree of the face, and the left and right eye blink rates. The number of times the maximum value of the average values of left and right eye blink rates during a certain length of time exceeds a preset threshold is defined as the number of eye blinks.

Besides, remote-photoplethysmography (rPPG) is used for heart rate estimation from facial skin region images. Hemoglobin contained in blood has its absorbance peak in the wavelength range of 500 to 600 nm. Hence, the analysis of the luminance change in this wavelength range allows heart rate estimation. The present study considers the skin diffuse-reflection model in Fig. 7 as the method of heart rate estimation[15].
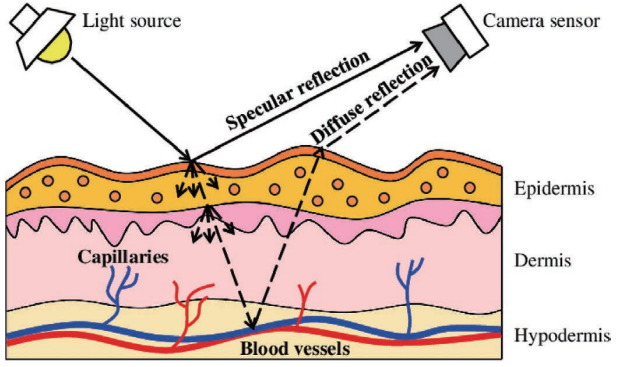
Source: Reference 15)

Fig. 7 Skin diffuse-reflection model[15]

Human skin consists of two layers, epidermis and dermis thereunder. Blood vessels run all over the hypodermal tissue underlying the dermis, and some capillary vessels reach the dermis. When light reaches the epidermis, some of the light penetrates inwards, and the rest is reflected. Some of the light penetrating inwards penetrates the dermis and reaches the blood vessels in the hypodermal tissue. The light reflected there passes through the dermis and the epidermis and is diffused outside the epidermis. The two types of light, one reflected from the epidermis and the other diffused from the epidermis, are captured by the camera. Therefore, the latter's selective extraction from the obtained data allows the detection of heart rate-dependent luminance changes. Fig. 8 shows the flow of heart rate estimation:
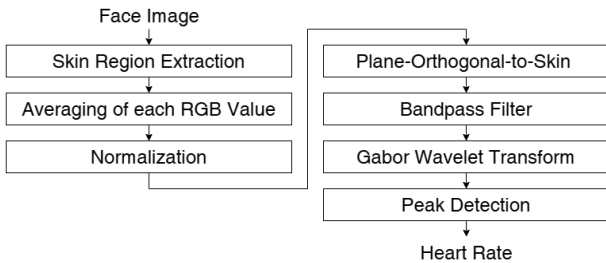

Fig. 8 System flow of heart rate estimation

From the face region trimmed by the OKAO Vision, the skin region is extracted by setting thresholds in the YCrCb color space, which is robust to lighting variations and widely used for skin color determination. Then, the RGB values in the skin region are averaged respectively, followed by the normalization of the time-series data covering a certain length of time, to remove the trends present in the data. Because the situation of our interest is ping-pong, which involves vigorous motions, a plane-orthogonal-to-skin (POS) algorithm robust to the subject's motions is employed to extract heart rate-dependent signals from the time-series data[15]. After that, a band-pass filter for the 1-to-3 Hz range corresponding to the normal heart rate range of

60 to 180 bpm is applied to remove noise from pulse signals. Next, to perform a frequency analysis without losing the signal's time-domain data, the Gabor wavelet transform, which uses a Gaussian function as the window function, is performed to calculate the spectrogram. Finally, from among the extremal values present in the range of 1 to 3 Hz, the highest value is extracted as the heart rate frequency.

3.3 Emotion estimation

Studies on human emotion estimation are widely pursued in the field of human-machine interaction. The methods used fall mainly into two types: methods of the first type are used to estimate emotions from surface layer data, such as facial expressions and pupil size changes, while those of the second type are employed to estimate emotion based on deep layer data, such as heart rates and brain waves. In sports, performance is considered strongly correlated with emotions. Therefore, the present study combines surface layer data, deep layer data, and ping-pong performance data to estimate emotions with high accuracy. The present study adopts as its emotion model a Russell's circumplex model expressed by two scales—one called a pleasure/displeasure scale and the other called an arousal/non-arousal scale[12]. The pleasure/displeasure scale score x and the arousal/non-arousal scale score y are calculated using the number of continued rallies c and the human player-hit ball velocity v, which are obtained as in Subsection 2.1, and the human player's smile degree s, seriousness of the face degree n, heart rate variation value h, and the number of eye blinks b, which are obtained as in Subsection 3.2. For easy implementation and to guarantee real-timeness, we assume this time that the input and output data are in a simple linear relation, which is expressed by Equation (2):

$$\begin{bmatrix} x \\ y \end{bmatrix} = A\begin{bmatrix} s & n & h & b & c & v \end{bmatrix}^T \qquad (2)$$

where A is a 2 × 6 matrix and is composed of the elements in Equation (3):

$$A = \begin{bmatrix} a_1 & 0 & a_3 & 0 & a_5 & a_6 \\ 0 & a_8 & 0 & a_{10} & a_{11} & a_{12} \end{bmatrix} \qquad (3)$$

It is known that the smile degree and the heart rate variation correlate strongly with pleasure/displeasure and have only minor effects on arousal/non-arousal and that the seriousness of the face degree and the number of eye blinks correlate strongly with arousal/non-arousal and do not have significant effects on pleasure/displeasure[16]. Therefore, the values of the elements with minor effects are set to 0. Presented below is how to obtain the values of the other elements. Thirteen beginner players are

asked to keep rallying with the ping-pong robot for about five minutes so that the robot can obtain their rally data and vital data. The ping-pong robot is programmed to randomize the speed and course of its return shots regardless of the participants' skill level. Immediately after their rally breaks, the participants answer a questionnaire as a rally-by-rally record of changes in the pleasure/displeasure and arousal/non-arousal valences. Emotions are difficult to make an absolute evaluation of and hence are evaluated relative to the emotions during an immediately preceding rally. The expected values of changes in pleasure/displeasure and arousal/non-arousal in response to changes in the rally data and vital data are calculated for use as the values of the elements of matrix A. As a result, elements a_1, a_3, and a_8 take a positive value, respectively, while element a_{10} takes a negative value. These results are qualitatively equivalent to the results reported by a preceding study[16].

3.4 Return-shot planning technique

The target ball return position, speed, and spin speed for the human player's side of the ping-pong table need to be preset to determine the robot's ball return action as in Subsection 2.3. This time, a fixed ball spin speed, a variable ball return position, and a variable ball return speed are used to improve the human player's motivation to keep rallying. The human player's emotions are continuously monitored, along with reference to the Meta AI[4] designed to affect emotions, for the robot to change its return-shot plan to suit the changes in the human player's emotions. The ball return speed is made to change from the initial value, which is the output value in Subsection 3.3, in response to the changes in the human player's emotions. Fig. 9 shows the algorithm of this function:
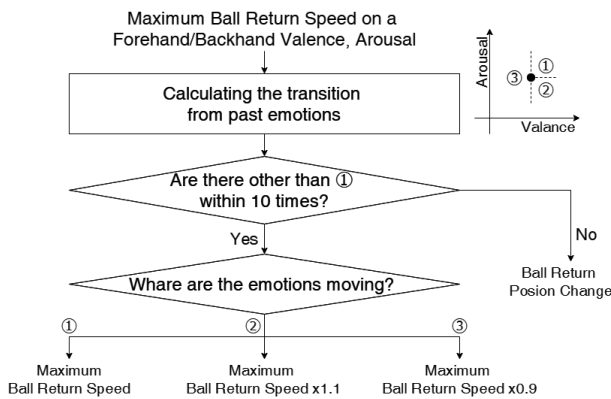

Fig. 9 System flow of return-shot planning

Changes in the human player's emotions from an immediately preceding swing mainly fall into the following three patterns:

(1) Both the pleasure scale score and the arousal scale score increase;

(2) The pleasure scale score increases while the arousal scale score decreases; and

(3) The pleasure scale score decreases.

For pattern (1), the current return-shot plan continues without changing because the current plan is just right for the human player and has improved the motivation to continue rallying. For Pattern (2), the ball return speed needs to be increased because the current return-shot plan feels easy and boring. Conversely, for Pattern (3), the ball return speed needs to be decreased because the current task feels difficult and unpleasant. When Pattern (2) or (3) continues a certain number of times, the ball return course is switched to improve the human player's motivation to keep rallying.

4. Demonstration experiment

This section describes the effectiveness evaluation for the Moti-Ctrl function proposed above.

4.1 Experiment method

At our request, 27 beginner players kept rallying for five minutes (with the target ball return position and speed fixed) for both when the Moti-Ctrl function was enabled and disabled. After then, they were asked to answer and submit the following three-item questionnaire:

(1) In which of the cases below did you rally more pleasantly?

(2) In which of the cases below did you rally with higher concentration?

(3) In which of the cases below do you want to have another rally?

Items (1) and (2) relate to the pleasure/displeasure scale and the arousal/non-arousal scale, respectively. Meanwhile, Item (3) is set as a question on the motivation to have a rally. We asked the participants to have a sufficient number of rallies before the experiment so that their familiarity with ping-pong would not affect the questionnaire results. Besides, we conducted the experiment, with the order of the rallies with and without the Moti-Ctrl function being randomized without providing the participants with prior knowledge about the function.

4.2 Experiment results and discussions

Fig. 10 shows the three-item questionnaire results:

(1) Which mode did you rally comfortably?
(2) Which mode did you rally with focus?
(3) Which mode do you want to do again?

37%
63%

4%
96%

19%
81%

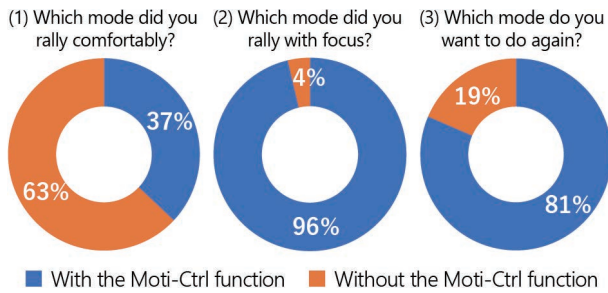■ With the Moti-Ctrl function ■ Without the Moti-Ctrl function

Fig. 10 Questionnaire results for with and without the Moti-Ctrl function

The participants showed a tendency to have a rally more pleasantly with the Moti-Ctrl function disabled (63%). With the Moti-Ctrl function enabled, they showed a tendency to rally with higher concentration (96%) and have higher motivation to continue rallying (81%).

With the Moti-Ctrl function enabled, the participants did not show a tendency to rally pleasantly, probably because of a problem with the reflection of feedback in the return-shot plan. This function upsets the rhythm of the rally, probably making its continuation difficult. As a result, the Moti-Ctrl function may have reduced the proportion of participants that enjoyed rallying pleasantly. Besides, the participants may have differed in their intention to rally with the ping-pong robot: for example, some may have wanted to win, some others may have wanted to continue rallying, and some additional others may have wanted to wait and see the robot's behavior. Consequently, the pleasure/displeasure scale may have become less correlated with the motivation to continue rallying. We consider the following three challenges necessary to be addressed for the future improvement of the motivation of a broader spectrum of human players:

(1) An ability to understand human intentions;
(2) Enhanced accuracy of obtained data; and
(3) Interaction techniques other than return-shot planning.

Regarding Challenge (1), time-series data on human players' motions and rally patterns will provide a useful basis for estimating human intentions. A robot able to understand human players' intentions will contribute to maintaining their motivation in line with their intentions. Regarding Challenge (2), we expect more accurate emotion estimation to be achievable by improving the accuracy of vital data measurement by contact-type devices, obtaining breath or perspiration data by a new sensor, and estimating emotions from obtained time-series data. Moreover, the skill level estimation accuracy will be improved by obtaining and taking into consideration vital data, such as players' line-of-vision data or muscle activity data, besides the

ball data and players' motion data. Regarding Challenge (3), we will explore new interactions based on visual, auditory, and tactile senses, taking into account not only the robot's ball return action but also the conditions of the rally and the human player.

5. Conclusions

For the present study, we developed an interaction system able to improve human players' motivation to continue a ping-pong rally. We implemented a Meta AI-based return-shot planning method for estimating skill levels and emotions from contactlessly obtained human motion data and vital data. As a result, we successfully improved the motivation of more than 80 percent of trial players.

Our aim for the future is to overcome the challenges presented in Subsection 4.2 to improve the motivation of a broader spectrum of human players. Besides, bearing in mind the characteristics of man-man and man-machine relationships, we will develop interactions for improving these relationships through ping-pong rallies to promote further man-machine harmony.

References

1) K. Yamada, "Robot table tennis tutor as an example of 'Harmony between Human and Robot'," (in Japanese), *J. Inst. Electr. Eng. Jpn.*, vol. 137, no. 2, pp. 81–84, 2017.

2) Y. Nishina, M. Suwa, and M. Kawade, "Application of image sensing and AI technologies to table tennis robots," (in Japanese), *O plus E*, vol. 39, no. 12, pp. 1195–1200, 2017.

3) K. Asai, M. Nakayama, and S. Yase, "The ping-pong robot to return a ball precisely ~ Trajectory prediction and racket control for spinning balls ~," (in Japanese), *OMRON TECHNICS*, vol. 51, no. 1, pp. 174–179, 2019.

4) Y. Miyake and Y. Mizuno, "Game design revolution using artificial intelligence (Meta AI)," Computer Entertainment Developers Conf., 2017, https://cedil.cesa.or.jp/cedil_sessions/view/1757 (accessed Jan. 28, 2021).

5) D. Satoi and Y. Mizuno, "Changing the game: Measuring and influencing player emotions through Meta AI," Game Developers Conference, 2019, https://schedule2019.gdconf.com/session/changing-the-game-measuring-and-influencing-player-emotions-through-meta-ai/861775 (accessed Jan. 28, 2021).

6) Square Enix Co., Ltd., "FFXV" AI Team, in *Artificial Intelligence of FINAL FANTASY® XV - Future as Seen from Game AI* (in Japanese), Tokyo: Born Digital, 2019, pp. 116–117.

7) Y. Miyake, M. Nakayama, K. Fujita, and Y. Mizuno, "Application of Game AI Technology to Ping-Pong Robot FORPHEUS," (in Japanese), Computer Entertainment Developers Conf., 2020, https://cedec.cesa.or.jp/2020/session/detail/s5e83300d562e1 (accessed Jan. 28, 2021).

8) M. Nakayama, "Compilation of OMRON's Core Technologies Sensing & Control + THINK, the Ping-Pong Robot FORPHEUS

— Toward the Achievement of Man-Machine Harmony," (in Japanese), *O plus E*, vol. 469, 2019, https://www.adcom-media.co.jp/report-iss/2019/09/25/32422/ (accessed Jan. 28, 2021).

9) NUITRACK, "Nuitrack full body skeletal tracking software," https://nuitrack.com/ (accessed Dec. 21, 2020).

10) K. Yeung, T. Kwok, and C. Wang, "Improved skeleton tracking by duplex kinects: A practical approach for real-time applications," *J. Comput. Inf. Sci. Eng.*, vol. 13, no. 4, p. 041107, 2013.

11) G. Schweighofer and A. Pinz, "Robust pose estimation from a planar target," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2024–2030, 2005.

12) J. Russell, "A circumplex model of affect," *J. Personal. Soc. Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980.

13) J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden markov model," *Proc. 1992 IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Champaign, IL, USA, 1992, pp. 379–385.

14) OMRON, "OKAO Vision," (in Japanese), https://plus-sensing.omron.co.jp/technology/ (accessed Dec. 21, 2020).

15) W. Wang, D. Brinker, S. Sander, and D. Gerard "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, 2017.

16) Y. Ikeda and M. Sugaya, "Estimate emotion method to use biological symbolic information preliminary experiment," *Lect. Comput. Sci.* (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9743, pp. 332–340, 2016.

About the Authors

NAKAYAMA Masamune
Technology Research Center
Technology And Intellectual Property H.Q.
Specialty: Image Processing, Mechanical engineering, Biomecanics

KURISU Takanori
Technology Research Center
Technology And Intellectual Property H.Q.
Specialty: Motion analysis, Human Skill Augmentation

MIZUNO Yuta
AI Unit
Advanced Technology Division
SQUARE ENIX CO., LTD.
Specialty: Meta AI

MIYAKE Youichiro
AI Unit
Advanced Technology Division
SQUARE ENIX CO., LTD.
Specialty: AI in Digital Games

YASE Satoshi
Technology Produce Center,
Technology And Intellectual Property H.Q.
Specialty: Electrical and electronic engineering

The names of products in the text may be the trademarks of each company.