Regular Paper

A Method for Determining the Moment of Touching a Card Using Wrist-worn Sensor in Competitive Karuta

Hiroshi Yamada[1,a]   Kazuya Murao[2,b]   Tsutomu Terada[1,3,c]   Masahiko Tsukamoto[1,d]

Abstract: Competitive karuta is an official Japanese card game and is described as "martial art on the tatami." Recently, competitive karuta has attracted a great deal of attention among young people. One of characteristic rules of competitive karuta is that there is no referee; therefore players must judge themselves even if the difficult situation arises. Consequently, the players sometimes get into an argument over their judgement, which disrupts the other matches in the room because all the matches proceed in parallel. In this paper, we propose a system that judges the player who took a card first in a competitive karuta match. Our system measures motion data when players take a card by using a wrist-worn accelerometer and gyroscope, and estimates the times when the players touched the card. From the evaluation experiments, 69.2% of rounds were estimated without error and 99.0% of rounds were estimated within 20-ms error. When our system was introduced on the close game, the accuracy of deciding the player taking a card was 75%.

Keywords: karuta (Japanese playing cards), accelerometer, officiating system

1. Introduction

Competitive karuta is an official Japanese card game within the format and rules set by the All Japan Karuta Association. Competitive karuta has been around since the start of the 19th century and now it is played widely in Japan [1]. In particular, it has become popular among young people. The number of participants in the All Japan Senior High School Ogura Hyakunin Isshu Karuta Championship has increased three-fold over the last 10 years. Moreover, it has begun gaining international players as well. The first international tournament was held in September 2012, with participating players from the U.S., China, South Korea, New Zealand, and Thailand.

Competitive karuta is a one-on-one game that uses the "Ogura Hyakunin Isshu" karuta cards. Ogura Hyakunin Isshu is 100 Japanese poems collected in a classical Japanese anthology. Hyakunin Isshu can be translated as "100 people, one poem each." There are several different compilations of Hyakunin Isshu, the most famous of which is Ogura Hyakunin Isshu, called simply Hyakunin Isshu. It was compiled by Fujiwara No Teika, one of the great Japanese poets, while he lived in the Ogura district of Kyoto, Japan. All the poems in Hyakunin Isshu are written in the Tanka style, consisting of 31 (5–7–5–7–7) syllables, with the first 17 (5–7–5) syllables called kamino-ku and the final 14 (7–7) syllables called shimono-ku. Two types of karuta cards are used for the game: yomifuda (reading card) and torifuda (playing card), consisting of 100 cards each. A picture of the poet and the complete poem (31 syllables) are shown on yomifuda, while only the final 14 syllables are displayed on torifuda in Hiragana, the Japanese syllabary.

Competitive karuta is facilitated by a reciter (card reader). Torifuda are aligned between two players facing each other and yomifuda are held by the reciter. The concept of competitive karuta is that two players find and take a torifuda that matches the yomifuda read by the reciter. The trickiest part of the game is that the reciter reads only kamino-ku and the players take cards on which only shimono-ku is written. Therefore, the players have to memorize all 100 poems. **Figure 1** shows the basic rules of competitive karuta. First, 50 cards are randomly chosen from 100 torifuda and then each player randomly chooses 25 cards from the 50 and places them face-up in their respective territories. The territory is the space in front of the player, measuring 3-card height by 16-card width. The territories must be 3 cm apart from each other. After aligning the cards, the players are given 15 minutes to memorize all the cards in both territories, and then the game starts. The reciter slowly reads aloud a card randomly chosen out of 100 yomifuda, and the players touch the torifuda on the field corresponding to the yomifuda that the reciter is reading. The player who touches the card first is the winner of that round and removes the card from the field. When a card is removed from the loser's territory, an arbitrary card in the winner's territory is transferred to the loser's territory. Therefore, one card drops away from the winner's territory in one round. The player who gets rid of all their cards first is the winner of the game.

1   Graduate School of Engineering, Kobe University, Kobe, Hyogo 657–8501, Japan
2   College of Information Science and Engineering, Ritsumeikan University, Kusatsu, Shiga 525–8577, Japan
3   PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332–0012, Japan
a)   yamadahiroshi@stu.kobe-u.ac.jp
b)   murao@cs.ritsumei.ac.jp
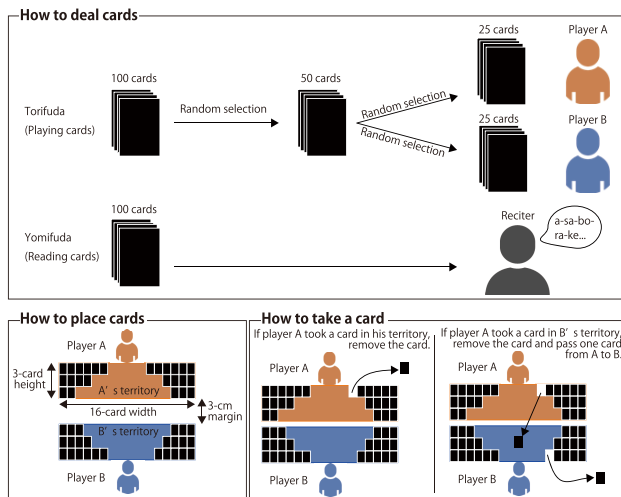c)   tsutomu@eedept.kobe-u.ac.jp
d)   tuka@kobe-u.ac.jp
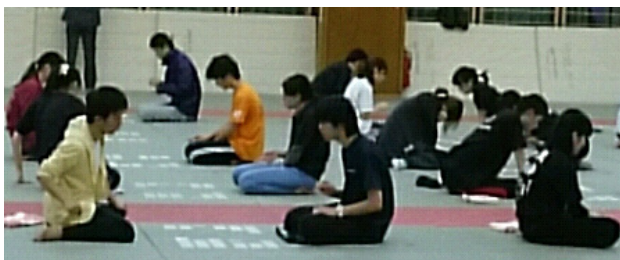
Fig. 1   Rules of competitive karuta.

Fig. 2   Competitive karuta tournament.

There are fouls in competitive karuta. If a player touches any card on the floor when karafuda is read, it is a foul. Karafuda means a card that is not in either territory. Karafuda is read by the reciter since the game starts with 50 torifuda in their territories but the reciter reads a card out of 100 yomifuda at random. Half of the poems recited are karafuda. If this foul is committed, the player who committed the foul has to take one card from the opponent's territory. In addition to this foul, if a player touches any cards that are in a different territory from the territory in which the card corresponding to the recited card is, it is a foul; i.e., when the card to be taken is in A's territory, and player A touches B's territory, it is a foul. If this foul is committed, the player who committed the foul has to take one card from the opponent's territory.

In a competitive karuta tournament, multiple matches are simultaneously played in parallel with one reciter in a large room, as shown in **Fig. 2**. Generally, competitive karuta is played without a referee, so players must judge themselves (self-judgement) even if a difficult situation arises. Most rounds are not controversial, but sometimes players get into an argument over their judgement of who touched a card first, which disrupts the other matches in the room because all matches must proceed in parallel. The players in question must bring the argument to an end, which means the other matches are interrupted for a while. To solve similar problems in sports such as tennis and soccer, sensing technologies such as Hawk-Eye [3] and GoalControl-4D [*1] have been used in recent years. These technologies would be applicable to competitive karuta, but it is difficult to install the system from the viewpoints of cost and physical problems so that all matches in

the room can be monitored. Recording the matches with a simple 30-fps camera would be possible, but it is hard to judge which player touches a card first, since the time difference is less than 40 msec. The reason the time difference is quite small is that the card to be taken can be identified at a specific time. There are several poems whose first few syllables are the same. For example, if the first three syllables of two poems are the same, the torifuda to be taken is identified when the fourth syllable is read by a reciter; therefore, the players rush to take the card at almost the same time.

In this paper, we propose a system that judges which player took a card first in a competitive karuta match. Our system measures the motion data when players take a card by using a wrist-worn accelerometer and gyroscope and estimates the times when the players touched the card. Our two main contributions in this paper are:

- **The first officiating system for competitive karuta**
  Recent advances in microelectronics, sensor integration, and data analysis techniques have opened up new possibilities for wearable technology [2]. Using sensor technology in sports has helped promote fairness and enhanced spectator enjoyment. However, competitive karuta is played without a referee attending each game, so players must judge themselves, even if a difficult situation arises. To the best of our knowledge, this is the first paper that addresses a wearable system for supporting competitive karuta.

- **Establishment of a time estimation method for the performance of activities**
  Although previous studies have shown ways of detecting an event when performing activities, the detected timing of the event is not evaluated with millisecond accuracy. In competitive karuta, the time difference of touching a card is extremely short, and so our proposed method distinguishes the time difference in milliseconds. We also confirm that our proposed method can be applied to other activities in addition to competitive karuta.

2.   Related Work

2.1   Sports Officiating

The introduction of technology to the officiation of sporting matches has helped improve the accuracy of decision making and can potentially reduce the number of controversial incidents. Officiating technologies such as multiple camera-based ball trajectory analysis, electronic field manipulation, and video replays have already been used in some major sports. For example, Hawk-Eye [3] is used in many ball-games officiating as a support for the umpire, where several high-speed cameras are situated around the court to determine where the ball has bounced and whether it is in or out within five seconds of the bounce. GoalControl-4D is a vision-based system used in soccer featuring 14 high-speed cameras located around the pitch and directed at both goals. GoalRef [*2] is a goal-line technology that uses a microchip embedded inside a football. When the ball crosses the goal line, it interrupts a magnetic field in the goal mouth, and a

---

*1   http://goalcontrol.de/en/

*2   http://www.iis.fraunhofer.de/en/ff/kom/proj/goalref.html

signal is then sent to the referee to indicate a goal. The Goal-Ref system is licensed by the Fédération Internationale de Football Association (FIFA) but is not currently installed in any stadium. FIFA's Web site [*3] lists 106 stadiums with licensed goal-line technology installations, all of which use either the Hawk-Eye or Goal Control-4D systems. However, due to the expense of goal-line technology systems, the technology is currently used only at the very top levels of the game. Moreover, when we consider that several games are simultaneously played in competitive karuta, these systems are not appropriate. Chi et al. [4] proposed a system that assists the umpires in Taekwondo matches by attaching piezoelectric sensors to the body protectors of the players. Helmer et al. [5] proposed an automated scoring system for amateur boxing by attaching an array of piezoelectric sensors to the players' vests. A similar system can be created by attaching piezoelectric sensors to players' fingertips, but this is not acceptable in competitive karuta because there is a rule that bans the players from attaching anything to their hands.

2.2   Activity Recognition

There have been plenty of studies on activity recognition using wearable sensors, some of which have been applied to sports. Lapinski et al. [6] evaluated professional baseball pitchers and batters by using wearable sensor systems, and Ladha et al. [7] proposed a climbing performance analysis system by using a watch-like sensing platform that measures acceleration. Kosmalla et al. [8] also proposed a system for climbing by using wrist-worn inertia muasurement units. The system can automatically recognize the route which a user climbed during a climbing session. Bächlin et al. [9] built a system consisting of sensing and feedback hardware for swim analysis. The system opens up exciting new possibilities in the field of swim training, as objective values can be provided at all times for complete training. None of these systems, however, can estimate the instant an action is performed. Zhou et al. [10] constructed a system that uses textile pressure sensing matrices. The system can distinguish different ways in which the players' foot strikes the ball. Connaghan et al. [11] investigated tennis stroke recognition using a single inertial measuring unit (IMU) attached to a player's forearm. This research classifies the tennis strokes into serve, forehand, and backhand. However, these researches do not measure the timing of the ball being struck. Blank et al. presented an approach for a ball impact localization on table tennis rackets using piezo-electric sensors [12]. However, they do not examine whether a ball impact timing is precise or not. Blank et al. also proposed a system that uses inertial sensors attached to a table tennis racket [13]. The system detects table tennis strokes by using an event detection method. This method detects strokes with an accelerometer installed on the racket grip and achieves a precision of 95.7 and a recall of 98.2. In these sports, significant changes in the data are linked to events such as the striking of a ball; however, such changes are not necessarily linked to the timing of touching a card in competitive karuta.

3.   Proposed System

3.1   System Requirements

Our system decides which player touched a card first in a competitive karuta match at each round. Considering the characteristics of competitive karuta, the following factors are required.

- **Sensor placement**
  There is a rule that bans the players from attaching anything to their hands. However, they are allowed to attach ornamentation on parts of their body other than their hands. In addition, nothing is allowed to be attached to the cards.
- **Sampling rate**
  If the time difference of two players' touches is about 50 ms, players have to discuss it, as the time resolution of the human eye is about 50 ms [15]. Therefore, we need a sensor that can sample at 40 Hz or more by using the sampling theorem.
- **Infrastructure**
  It is difficult to install the system somewhere in a building (e.g., on the ceiling) because the players in the room and the positions of the cards are subject to the environment.

3.2   System Structure

With the above requirements in mind, we propose a system that uses inertial sensors attached to the wrist of the player's dominant hand, as shown in **Fig. 3**. Attaching an object to the wrist is allowed within the rules of competitive karuta. The sensor used in the system contains a wireless accelerometer and a gyroscope (WAA-010 by Wireless Technologies, Inc. [*4]). The size of the sensor is W39 × H44 × D12 mm and it weighs just 20 g. In other words, the sensor is small and light and does not interfere with the game. If we placed sensors on the floor next to the players' territories the sensors would be able to sense touch events but would have a harder time identifying the player. Recent cameras sold as consumer products can video-record with a high frame rate – for example, the SONY Cyber-shot RX100 IV (DSC-RX100M4) can record a video with a 1,920 × 1,080 resolution at 960 fps – but it is difficult to place a camera next to the players' territories with a tripod from the safety point of view, since the players move substantially when touching a card. We set the sampling rate to 50 Hz, as we feel that our system does not accurately judge like the Hawk-Eye system in tennis but rather officiates in the place of a human. In addition, higher sampling frequency consumes
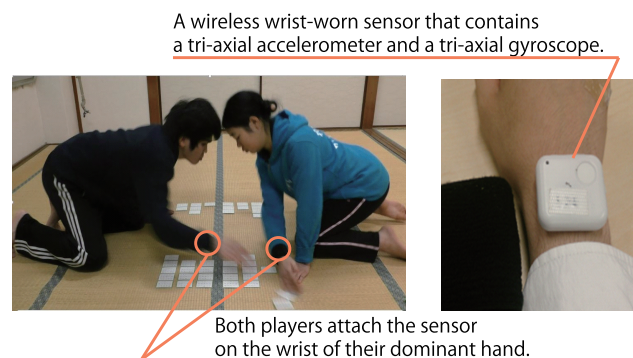
A wireless wrist-worn sensor that contains a tri-axial accelerometer and a tri-axial gyroscope.

Both players attach the sensor on the wrist of their dominant hand.

Fig. 3   The 3-axis accelerometer and 3-axis gyroscope used in the system.
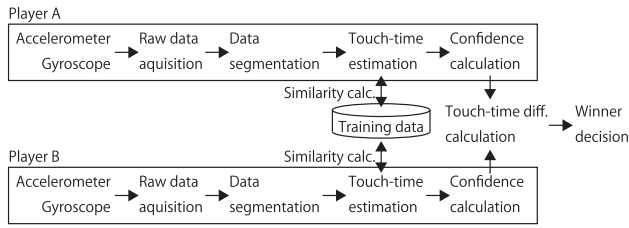
Fig. 4 Algorithm for determining the player taking a card.

Fig. 5 Extraction of swipe motion.

much power, therefore we set the sampling rate to 50 Hz on the safe margin. Our proposed system judges which player took a card first, i.e., the winner of the round, by estimating the timings of players touching cards.

Figure 4 shows the flow of the system. A player's movement is captured through the small wrist-worn sensing device, which is configured to record 3-axis acceleration data and 3-axis angular velocity data. The sensor data is sent to a tablet next to the players via Bluetooth and the system, which is installed on the tablet, judges whether the data includes a touch movement and extracts it. Then, the time at which the player touched the card is estimated by comparing the extracted data with training data. Exact time of touching a card is labeled with the training data, which is collected in advance. The confidence of the time estimation is then calculated. Lastly, by comparing both players' card touch times considering the confidence, our system judges the winner of the round and displays the winner on the screen of the tablet.

3.3 Assumed Environment

As stated previously, owing to the absence of a referee in competitive karuta matches, players have to judge themselves. When it is clear which player took the card first, there is no controversy. However, when the time difference is so small that the players cannot determine which of them took the card first, disagreements often arise. In tennis, players can claim the line technology system if a player complains about the referee's judgement. In competitive karuta, our system is assumed to be used when both players are in disagreement. The proposed system outputs three types of judgement: "player A won," "player B won," and "draw." "player A won" and "player B won" refer to respective winners of the round, and "draw" means that both players touched a card at almost exactly the same time. The reason the proposed system includes a "draw" option is that the system cannot distinguish very small time differences due to the sampling rate, and this is also hard for humans to distinguish. Our system makes a "draw" judgement when the estimated time difference is shorter than the threshold. The system performance on the threshold is evaluated in the experiment in the later section. In the rules of competitive karuta, the round is considered a draw and the card taken is removed from the territory if the players do not reach the conclusion; i.e., in the case that the card taken is in player A's territory, the card is just removed, therefore the result is the same as if player A had won. Applying the rule to this case, when the system makes the "draw" decision, the card taken is removed from the territory.

In order to prevent indiscriminate claims, if the estimated time difference is longer than 100 ms, the late player is considered to
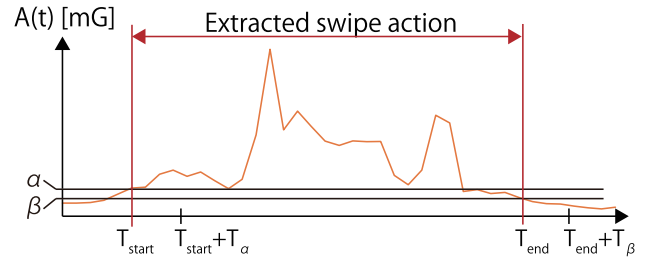
have committed a foul. In such a case, the player who committed the foul has to take one card from the opponent's territory in addition to the judgement of the system. For example, if the card in player A's territory was taken and player A took the card much faster than player B, the system makes a judgement of "player A won" and "foul of player B." As a result, the card is removed from player A's territory and one additional card is passed from player A to player B, resulting in player A getting rid of two cards. In general, since a player can perceive a 100-ms time difference in touching a card, this parameter is reasonable. This penalty is defined in the regulation of competitive karuta and applied for foul such as karafuda introduced in Section 1. Along with the installation of the system, it can be considered to apply the penalty accordingly to prevent the indiscriminate claims. Please note that this penalty is not officially defined in the regulation of competitive karuta and does not affect any experimental results in the paper.

If either player touches an incorrect card, this is typically clear and does not become controversial. Almost all controversy is related to two players touching the same correct card, so the proposed system is assumed to be used when both players touch the same card. The system calculates the time difference and makes a judgement at every round for each point at which cards can be placed, but this judgement is only displayed by inputting the point at which players touch a card when a judgement is required; therefore, the calculation for all points is performed during the controversy and the judgement is displayed immediately after the requirement.

There are some cases where the system detects an undesirable swipe action, such as a practice swing for image training between the events of reciting. To remove this action, the system does not always operate; rather, it operates when the system is sent a trigger for the start of sensing the data. In addition, it is sent when the reciter starts to read a poem. Therefore, we assume the algorithm that detects the card touch time is conducted at the beginning of the data sensing.

3.4 Data Segmentation

Sensor data is captured while the reciter is reading the card. The system detects the movement of swiping a card within the streams of acceleration data and angular velocity data and extracts the data. **Figure 5** shows an example of data while swiping a card. The system calculates the composite value of the 3-axis accelerometer $A(t) = \sqrt{a_x^2(t) + a_y^2(t) + a_z^2(t)}$, where $a_x(t)$, $a_y(t)$, and $a_z(t)$ are the acceleration values of the x-axis, y-axis, and z-axis, respectively. If condition $A(t) > \alpha$ is first satisfied for

T_α ms, the proposed system judges that the swiping action began at time T_{start}. Then, if condition $A(t) < \beta$ is satisfied for T_β ms, the system judges that the swiping action finished at time T_{end}. α and β are thresholds of the start and end of a swiping action set to 1,300 mG and 1,100 mG from a pilot study, respectively. The following segmented sensor data G is obtained through the data segmentation, where $g_x(t)$, $g_y(t)$, and $g_z(t)$ are angular values around the x-axis, y-axis, and z-axis, respectively:

$$
\mathbf{G} = \begin{pmatrix}
a_x(T_{start}) & a_x(T_{start}+1) & \cdots & a_x(T_{end}) \\
a_y(T_{start}) & a_y(T_{start}+1) & \cdots & a_y(T_{end}) \\
a_z(T_{start}) & a_z(T_{start}+1) & \cdots & a_z(T_{end}) \\
g_x(T_{start}) & g_x(T_{start}+1) & \cdots & g_x(T_{end}) \\
g_y(T_{start}) & g_y(T_{start}+1) & \cdots & g_y(T_{end}) \\
g_z(T_{start}) & g_z(T_{start}+1) & \cdots & g_z(T_{end})
\end{pmatrix}
\tag{1}
$$

3.5 Acquisition of Training Data

The proposed system compares the segmented data and training data to estimate the card touch time. One might think that the card touch time in the segmented data could easily be estimated just by seeing the waveform, without training data. However, this is difficult because sensor data differs depending on the placement of a card and the starting position of a swiping action. Moreover, the card touch time in the segmented data does not always correspond with the peak of sensor data attached to the wrist because players only touch the card with their fingertip at the time. Therefore, the proposed system collects swipe action data for all 96 points at which cards can be placed. At the same time, acceleration data of the card is collected by setting a 3-axis accelerometer under the card for labeling the exact card touch time. The time at which the 3-axis composite value of the accelerometer under card $A'(t)$ first exceeds a threshold is defined as an exact card touch time. The exact card touch time and the placement of the card are added to the data of the wrist-worn sensor. Note that the accelerometer under the cards is needed only for the training data collection and is not used during the match.

3.6 Estimation of Card Touch Time

The proposed system utilizes two methods for estimating card touch time: a feature value-based method (method 1) and a waveform similarity-based method (method 2). By combining these two methods, the proposed system estimates card touch time. The detailed algorithms of the methods are described below.

3.6.1 Method 1: Feature Value-based Method

The feature value-based method uses a sliding-window approach. Given the segmented data of swipe action G in Eq. (1), feature values are extracted over a 3-sample window that is slid in steps of one sample. The feature values used are max, min, and variance for the six axes and angle of the wrist for the three axes, for 21 dimensions in total. The angle of the wrist is calculated by integrating angular velocity. These feature values $F(t) = (f_1(t), f_2(t), \ldots, f_{21}(t))$ are calculated over a window ranging $[t-1, t+1]$ from $t = T_{start} + 1$ to $T_{end} - 1$. Feature values are also calculated from the training data at card touch time T_{touch} only; i.e., $F(T_{touch})$ is calculated. Feature vector $F(t)$ is standardized using $Z(t) = \frac{F(t)-M}{S}$, since scales of the feature values are

different, where M and S are respectively the mean and standard deviation of $F(t)$ over the training data. After this conversion, the mean and variance of $Z(t)$ become 0 and 1, respectively.

Euclidean distance between i-th training data $Z_{trn}^{(i)} = (z_{trn,1}, z_{trn,2}, \ldots, z_{trn,21})$ and input data $Z_{inp}(t) = (z_{inp,1}(t), z_{inp,2}(t), \ldots, z_{inp,21}(t))$ is calculated by

$$
Euclid(Z_{trn}^{(i)}, Z_{inp}(t)) = \sqrt{\sum_{j=1}^{21}(z_{trn,j}^{(i)} - z_{inp,j}(t))^2}
\tag{2}
$$

The system calculates $Euclid(Z_{trn}^{(i)}, Z_{inp}(t))$ from $t = T_{start} + 1$ to $t = T_{end} - 1$ and from $j = 1$ to N for all the training data collected at each point and finds the time T_{min} when $Euclid(Z_{trn}^{(i)}, Z_{inp}(t))$ shows minimal value, where N is the number of training data at each point. T_{min} is estimated as a card touch time of input data, since the waveform of input data around T_{min} is similar to the waveform at the card touch time of training data.

3.6.2 Method 2: Waveform Similarity-based Method

The waveform similarity-based method calculates the similarity between training data and input data with the dynamic time warping (DTW) algorithm [14], which measures the similarity of two time-series data. Advantages of DTW include the ability to calculate temporal nonlinear elastic distance, the ability to measure similarity between two sequences that may vary in time or speed and the fact that the numbers of both samples need not be equal. The details of the algorithm are explained as follows. For the sake of simplicity, we explain it for one-dimension data. When training data $X = (x_1, x_2, \ldots, x_m)$ and input data $Y = (y_1, y_2, \ldots, y_n)$ with length m and n are compared, an $m \times n$ matrix is defined as $d(i, j) = |x_i - y_j|$. Next, warping path $W = (w_1, \ldots, w_k)$, which is the path of the pairs of X and Y indices, is found. W meets three conditions.

- Boundary:
  $w_1 = (1, 1), w_k = (m, n)$
- Continuity:
  $w_k = (a, b), w_k - 1 = (a', b') \Rightarrow (a - a' \le 1) \land (b - b' \le 1)$
- Monotony:
  $w_k = (a, b), w_k - 1 = (a', b') \Rightarrow (a - a' \le 0) \land (b - b' \le 0)$

The following steps are used to find the path with the lowest cost that satisfies the above conditions.

Initialization:
$f(0, 0) = 0$
$f(i, 0) = \infty$ for $i = 1, \ldots, m$
$f(0, j) = \infty$ for $j = 1, \ldots, n$
Do for $i = 1, 2, \ldots, m$
    Do for $j = 1, 2, \ldots, n$

$$
f(i, j) = d(i, j) + \min \begin{cases} f(i-1, j-1) \\ f(i-1, j) \\ f(i, j-1) \end{cases}
\tag{3}
$$

Output:
    Return $D(X, Y) = f(m, n)/(m + n)$

The obtained cost $D(X, Y)$ is the distance between X and Y. The returned $D(X, Y)$ is divided by the sum of the length of the input and training data, since the DTW distance increases with the length of the sequences.
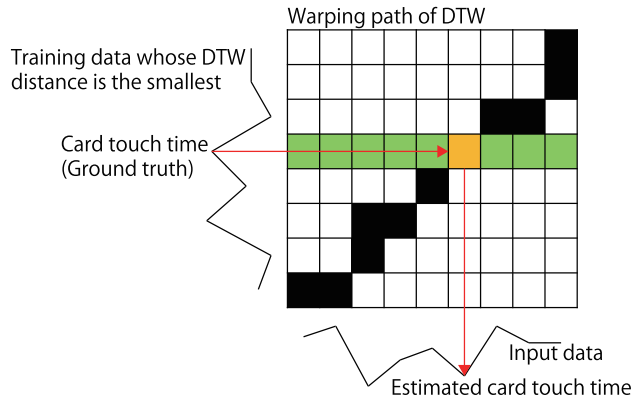
Fig. 6 Estimation of card touch time by using DTW.

Fig. 7 Algorithm for calculating the confidence.

In our system, input data and training data have six axes. It is necessary to calculate the distance with the six axes to accurately measure the similarity between input data and training data. However, the data on several axes have low reproducibility in performing gestures, which has an unwanted effect on the DTW calculation. Therefore, the proposed system selects appropriate axes out of the six axes for each person in advance. Suppose that ten items of training data are collected for each of the 96 points at which cards can be placed. Then, the card touch time of the training data at each point is estimated on a leave-one-sample-out cross validation basis for all combinations of the axes, $2^6 - 1 = 63$ patterns. Error of the estimated time is obtained by calculating the difference of the exact touch time (ground truth) and the estimated card touch time, and lastly the system finds the best combination of the axes that shows the largest number of samples whose error is equal to or less than 20 ms. The threshold is set to 20 ms since the sampling interval is 20 ms (50 Hz) and the labeled card touch time can include an error up to 20 ms. Method 2 calculates the DTW distances between input data and all training data at each point for the axes obtained above and finds the training data whose distance is the smallest.

The card touch time for the input data is estimated by finding the index of input data corresponding to the index of the card touch in the training data on the warping pass, as shown in **Fig. 6**. If multiple indices of input data correspond to the index of the card touch time in the training data, the estimated card touch time is set to the earliest index.

3.7 Estimation of Confidence

Since the motions to swipe a card are not always similar even on the same point, the proposed system has to consider the strange input data. Even if the input data is completely different from the training data, the card touch time is estimated anyway, which would be inaccurate. To address this problem, our system utilizes a confidence value that represents whether the estimated card touch time is reliable or not. If it is reliable, the confidence becomes "OK," and otherwise, "NG."

Figure 7 shows the algorithm to calculate the confidence. Suppose that several samples of training data were collected in advance for all of the 96 points at which the cards can be placed. The actual card touch time is labeled to the training data and the card touch time is estimated for the training data with leave-one-
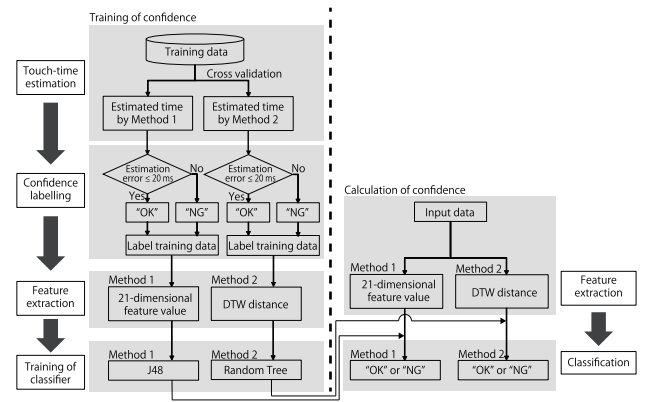
sample-out cross validation. The difference between the labeled card touch time (ground truth) and the estimated card touch time becomes an error. If the error is equal to or smaller than 20 ms, the confidence "OK" is given to the training data, and otherwise, "NG." The threshold is set to 20 ms since the sampling interval is 20 ms (50 Hz) and the labeled card touch time can include an error up to 20 ms. It can be said that the training data with "NG" label is strange data. The proposed system classifies the input data into "OK" and "NG" with the model that has learned training data in parallel with estimating the card touch time. We use the WEKA data mining software [*5] to classify confidence. The confidence is calculated for each method. For method 1, 21-dimensional feature values over the window at $t = T_{min}$ extracted in method 1 are trained with J48, which is a C4.5 algorithm implemented on WEKA. For method 2, DTW distances between the input data and the best matching training data calculated in method 2 are trained with RandomTree, which is a decision tree algorithm implemented on WEKA. The reason that RandomTree is employed is J48 cannot classify scalar explanatory variable. The confidence of methods 1 and 2 can be obtained by classifying the input data.

3.8 Calculation of Difference of Touch Time

Since preliminary experiments showed that card touch times estimated with the methods are not always accurate, the proposed method calculates the difference of the card touch time of both players by combining both methods and considering the confidence. There are four combinations: two kinds of confidence by two methods. The conclusive estimated touch card time is adopted according to the following rules.

- If the confidence of method 2 is "OK," the card touch time estimated by method 2 is adopted regardless of the confidence of method 1. This is because a preliminary experiment showed that the accuracy of estimating the card touch time by method 2 was superior to that of method 1.
- Otherwise, if the confidence of method 1 is "OK," the card touch time estimated by method 1 is adopted.
- If the confidences of both methods are "NG," the card touch time estimated by method 1 is adopted. This is because confidence "NG" of method 2 means that the waveform of the input data is not similar to that of training data. Moreover, DTW is susceptible to the difference in the number of peaks

in the waveform, while the method based on feature value is more robust against the difference in the number of peaks in the waveform than DTW.

After estimating the card touch time of both players, our system calculates the time difference between both players' card touch time by $T_{diff} = T_A - T_B$, where T_A and T_B are estimated card touch time of player A and B, respectively.

3.9 Judgement of the Winner

Lastly, our system judges the winner of the round. Basically, the player whose card touch time is earlier is the winner. However, if the time difference T_{diff} is quite short, it would be outside of the performance limit and our system makes a "draw" judgement. The "draw" judgement is not a problem because there is a rule that regulates a draw: the card taken is removed from the territory which the card is in.

3.10 Implementation

We constructed an application that judges the player taking a card based on the proposed method. ThinkPad X240 by Lenovo Corporation is used as a computer to receive and analyze data. The application is developed with Visual C#. **Figure 8** shows the screen shot of the application. How to use the application is as follows. First, the players set their training data and make their models of J48 and random tree to calculate the confidence. Then, the players exchange high-five to synchronize the clock of the two sensors. The time at which the 3-axis composite value of the accelerometer $A(t)$ first exceeds a threshold is defined as time 0. During the match, two sensors measure the players' movement, and both players can refer to the judgement of our system if a problem arises. When our system judges the winner, the system estimates the place of the card taken at the same time. If the place of the card is incorrect, the players can manually correct it and the system judges the winner again based on the correct place of card. Lines of rushy stitch can be seen on a tatami mat. Players basically place their card aligning with the lines from the edge in order to make a foul clear, therefore cards are placed as if there are grid and card positions on the display matching the card in the field. Moreover, our system has a function that reads cards automatically, which enables the users to play a game by using our system only. In addition, the players can easily review their
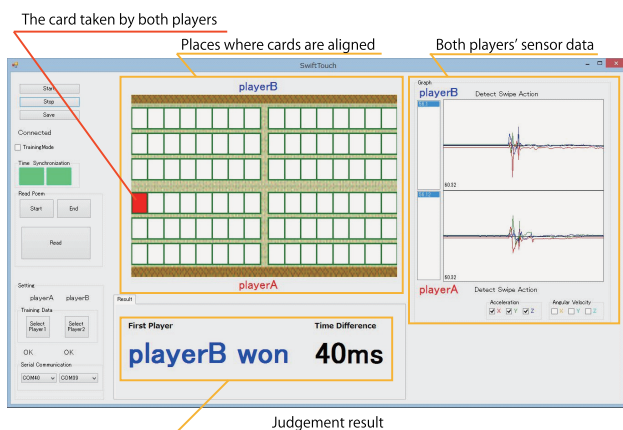

Fig. 8   Screen shot of the application.

match because our system saves the results of all the judgements.

4. Evaluation

4.1 Evaluation on Estimation of Card Touch Time

4.1.1 Experiment Environment

We evaluated the accuracy of estimation of card touch time. Data of swipe action for all 96 points at which cards could be placed were captured ten times from three players (two males and one female, aged 20–22 years) through the proposed system. The players were right-handed and had more than one year's experience playing competitive karuta. As an indicator of the performance, we measured the error of the estimated card touch time, which is the difference between estimated card touch time and exact card touch time. The exact card touch time is obtained with an accelerometer placed under the cards. The card touch time of each card is estimated using ten trials at each point with leave-one-out cross validation; nine trials are used for training and one trial is estimated. We conducted this procedure for all 96 points and 960 estimated card touch times are obtained for each player, 2,880 results in total.

4.1.2 Results

Figure 9 shows the histogram of the estimation error of card touch times. The horizontal axis indicates the error and the vertical axis indicates relative frequency. From the results, the largest error is −60 ms and the 69.2% of the estimated card touch times are exact (±0-ms error). Considering that the sampling interval of the system is 20 ms, the error within ±20 ms is 99.0%. We would say that there is no problem to allow the ±20-ms error since it produces 40-ms error in the difference of the card touch times of two players at the most, which is within the limit of human perception. We assume that the system officiates the game instead of a human. We also evaluated the error of card touch times by changing the proportion of training data and test data. For the case that uses five samples for training and the remaining five samples for testing and that uses one sample for training and nine samples for testing, the errors within ±20 ms are 98.4% and 96.7%, respectively. It can be said that card touch times are estimated accurately even with one training data. In the data collection, it took 105 minutes for a player to take a card 960 times, therefore it takes 95 minutes to take a card 864 times which were used as training data. Assuming one training data for each point, it takes approximately 10 minutes in simple calculation.

The frequency of −20-ms error is larger than that of 20-ms er-
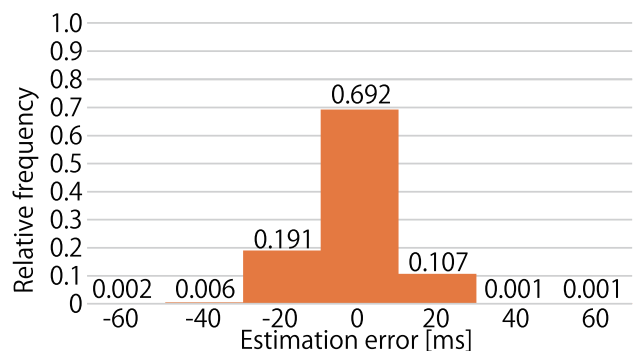

Fig. 9   A histogram of the error of estimated card touch time.

ror. The labeled card touch time is the time when the value of the accelerometer under the card exceeded a threshold, however, the sensor value moves after the player touches the card due to discrete sampling, therefore a couple of data is labeled with "late" card touch time. If both test data and the best-matched training data are with either accurate or late label, the error is also almost zero. If the test data is with late label and the best-matched training data is with accurate label, the error is negative, while if the test data is with accurate label and the best-matched training data is with late label, the error is positive, however, the latter case hardly occurs since the training data with late label will be the best-matched one due to the small amount of training data with late label. As a result, it can be said that the distribution of the error in Fig. 9 leaned on the negative side.

We also evaluated the accuracy of position estimation for cards taken by a player. The proposed system finds the best matching training data, position which is the estimated card position. This evaluation is conducted for 2,880 samples in 10 fold cross validation, resulting in 48.9% accuracy at exact match and in 80.4% accuracy at accepting right and left cards. For this result, the position estimation error is not large and the accuracy of two-once is an acceptable degree.

4.2 Evaluation on Judgement in the Match
4.2.1 Experiment Environment

We evaluated the accuracy of the judgements of the proposed system in the competitive karuta match. In competitive karuta, the time difference of card touch time of both players becomes short as the number of cards on the territories decreases since the candidate cards are limited and the players can easily guess the card to be read. Therefore, we conducted an experiment on the condition that four cards are remaining on each corner of the territories, i.e., cards are placed on the right edge and the left edge nearest the front for player A and B. In competitive karuta, it is standard tactics to keep cards at a distance from the opponent player, therefore the four cards are necessarily placed in this way. Two players out of the three players who joined the aforementioned experiment joined the experiment. The reciter read a card corresponding to the four cards 20 times at random and the players took the card. Training data of both players are given to the proposed system. The match was video-recorded with a high-speed camera (SONY Cyber-shot RX100 IV (DSC-RX100M4)) at 960 fps and the exact time of difference is manually obtained afterward.

4.2.2 Results

Figure 10 shows the scatter plot of the results. The horizontal axis indicates the time difference by camera and the vertical axis indicates the time difference by our system. If the time difference is exactly estimated, the results are plotted on the line of $y = x$. "○" marks mean the error is less than 40 ms and "×" marks mean the error is equal and more than 40 ms. From the result, the mean absolute error is 22.6 ms and the maximum error is 68 ms. There is no case when both of confidence calculated by the two time estimation methods are "NG" in this experiment. In other words, when we calculate the time difference between both players' card touch time, the result that the error is equal
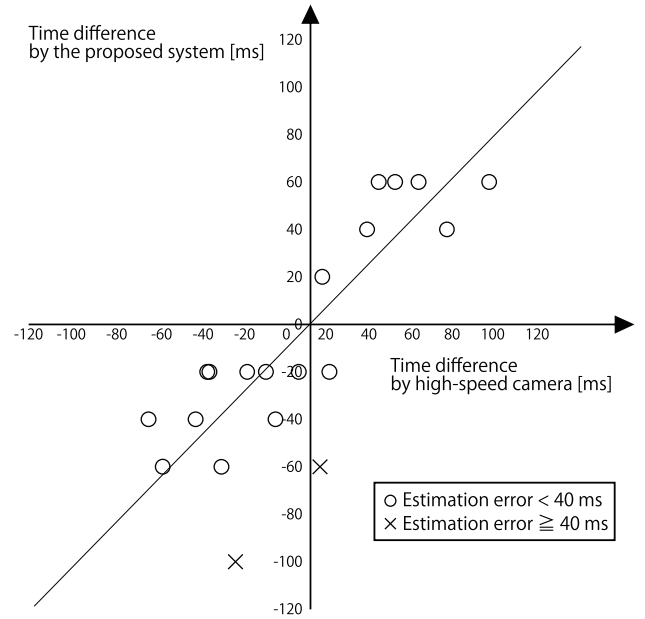

Fig. 10   Scatter plot of time difference in the game, illustrating the correlation between time difference by system and the ground truth.

Table 1   Judgements of the proposed system.

Actual time difference t ms by high-speed camera	System judgements								
	$Th = 0$			$Th \leq 20$			$Th \leq 40$		
	A	draw	B	A	draw	B	A	draw	B
$40 < t$ (A is faster)	3	0	0	3	0	0	2	1	0
$20 < t \leq 40$ (A is faster)	3	0	0	3	0	0	2	1	0
$0 < t \leq 20$ (A is faster)	1	0	2	0	2	1	0	2	1
$t = 0$	0	0	0	0	0	0	0	0	0
$-20 \leq t < 0$ (B is faster)	0	0	3	0	2	1	0	3	0
$-40 \leq t < -20$ (B is faster)	0	0	3	0	1	2	0	1	2
$t < -40$ (B is faster)	0	0	5	0	2	3	0	4	1

and more than 40 ms should not occur. The average of DTW distances using the estimation of players' card touch time 20 times for player A is 1613.2 and the average for player B is 1631.2 while the average of it when we estimate the card touch time of the training data with leave-one-out cross validation for player A is 800.5 and the average for player B is 707.4. This is because the sensor data of the follow through action while taking a card is different from the training data collected in advance. Moreover, our decision tree outputs "OK" when the DTW distances exceed a threshold. The DTW distances in this experiment are all satisfied with the threshold. In addition, player A's confidence calculated by method 1 is "NG" in two samples of "×" and the card touch time estimated by method 2 is earlier than method 1. From previous results, there is a high possibility of mistaking the card touch time estimated by method 2 because we know that the distribution of the error in Fig. 9 leaned on the negative side. Therefore, we need to reconsider the way of calculating the confidence for method 2.

Table 1 shows the judgements of the proposed system according to the actual time difference of card touch time obtained by the high-speed camera, where Th is the threshold of the estimated time of difference to be judged as "draw" which is set to 0 ms, 20 ms, and 40 ms, respectively. The shaded cells mean true positives. The accuracies of the system decision for $Th = 0$, 20, and 40 are 90%, 75%, and 50%, respectively. As the threshold

of draw decision increases, accuracy dropped. When $Th = 0$, the system is the most accurate, but the two false positives are mistaking wins of player A as player B, which is a serious problem. This is because the system makes draw judgements only when both players touched at exactly same time even though the sampling interval is 20 ms. When $Th \leq 20$, the accuracy dropped to 75%. All the false positives are mistaking draw as either player's win and either player's win as draw. When $Th \leq 40$, the cases that the actual time difference is around 60 ms are incorrectly estimated as 40 ms and less, resulting in draw judgements. With this setting, the system gives up judgement. From these results, the system with $Th \leq 20$ setting is the most appropriate and can be used as a system to support competitive karuta matches.

This section lastly describes the limitation of the proposed system. Someone may think that the form of touching a card changes as the number of cards remaining in the field decreases, however the touching form is basically unchanged. When there are many cards in the field, the players try to touch a card from a neutral position. There are two basic strategies when there are a couple of cards remaining in the field, two or three of which have the same first syllable: placing them distantly or placing them next to each other. If the cards are placed distantly, the form of touching a card would not change as the players try to touch the target card from a neutral position. If the cards are placed next to each other, the situation is more complicated since there is a rule that allows the players to touch any cards in the territory where the card read by the reciter is placed. Therefore, the players often use tactics that brush off all cards including the target card when placing the cards next to each other. Assuming that three cards are remaining in the field and the opponent has two cards with the same syllable and you have the rest with a different syllable, the opponent placed two cards next to each other whose first syllables are the same, the last card is placed away from the two cards from a tactical point of view, resulting in the neutral hand position to start.

5. Conclusion

We proposed in this paper a system that judges the player who took a card first in a competitive karuta match. The proposed system measures data of players taking a card with a wrist-worn accelerometer and gyroscope, and estimates the card touch time of the players, then judges the player who took the card first. We proposed two methods for estimating card touch time: feature-value based method and waveform-similarity based method, and a way of calculating the confidence of the time estimation. The system judged the winner of the round by combining the two methods based on the confidence. From the evaluation experiments, 69.2% of rounds were estimated without error and 99.0% of rounds were estimated within 20-ms error. The maximum error was 60 ms. From the results of actual matches, when the threshold of the estimated time of difference to be judged as "draw" is within 20 ms, the accuracy is 75%.

In the future, we plan to extract the other feature values that can handle the variety of motions involved in taking a card because there are a lot of ways to take a card in the game. In addition, we will scale the system to tournaments. Furthermore, we will apply the proposed algorithm to other activities.

References

[1] David, B.: Karuta: Sport or culture?, *Japan Quarterly*, Vol.43, No.1, pp.67–77 (1996).

[2] Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R. and Havinga, P.: Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey, *Proc. 23rd International Conference on Architecture of Computing Systems* (*ARCS 2010*), pp.1–10 (Feb. 2010).

[3] Bal, B. and Dureja, G.: Hawk Eye: A Logical Innovative Technology Use in Sports for Effective Decision Making, *Sport Science Review*, Vol.21, No.1-2, pp.107–119 (Apr. 2012).

[4] Chi, E.H., Song, J. and Corbin, G.: "Killer App" of Wearable Computing: Wireless Force Sensing Body Protectors for Martial Arts, *Proc. 17th Annual ACM Symposium on User Interface Software and Technology* (*UIST 2004*), pp.277–285 (Oct. 2004).

[5] Helmer, R.J.N., Hahnb, A.G., Staynesa, L.M., Denninga, R.J., Krajewskia, A. and Blanchonetta, I.: Design and Development of Interactive Textiles for Impact Detection and Use with an Automated Boxing Scoring System, *Procedia Engineering*, Vol.2, No.2, pp.3065–3070 (Mar. 2010).

[6] Lapinski, M., Berkson, E., Gill, T., Reinold, M. and Paradiso, J.A.: A distributed wearable, wireless sensor system for evaluating professional baseball pitchers and batters, *Proc. 13th International Symposium on Wearable Computers* (*ISWC 2009*), pp.131–138 (Sep. 2009).

[7] Ladha, C., Hammerla, N.Y., Olivier, P. and Plötz, T.: ClimbAX: Skill Assessment for Climbing Enthusiasts, *Proc. 15th International Conference on Ubiquitous Computing* (*UbiComp 2013*), pp.235–244 (Sep. 2013).

[8] Kosmalla, F., Daiber, F. and Krüger, A.: ClimbSense: Automatic Climbing Route Recognition using Wrist-worn Inertia Measurement Units, *Proc. 33rd Annual ACM Conference on Human Factors in Computing Systems* (*CHI 2015*), pp.2033–2042 (Apr. 2015).

[9] Bächlin, M., Förster, K. and Tröster, G.: Swimmaster: A wearable assistant for swimmer, *Proc. 11th International Conference on Ubiquitous Computing* (*UbiComp 2009*), pp.215–224 (Oct. 2009).

[10] Zhou, B., Wirth, M., Martindale, C., Koerger, H., Zwick, C., Cruz, H., Eskofier, B. and Lukowicz, P.: Smart Soccer Shoe: Monitoring Foot-Ball Interaction with Shoe Integrated Textile Pressure Sensor Matrix, *Proc. 20th International Symposium on Wearable Computers* (*ISWC 2016*), pp.64–71 (Sep. 2016).

[11] Connaghan, D., Kelly, P., O'Connor, N.E., Gaffney, M., Walsh, M. and O'Mathuna, C.: Multi-sensor classification of tennis strokes, *Proc. IEEE Sensors 2011 Conference*, pp.1437–1440 (Oct. 2011).

[12] Blank, P., Kautz, T. and Eskofier, B.M.: Ball Impact Localization on Table Tennis Rackets using Piezo-electric Sensors, *Proc. 20th International Symposium on Wearable Computers* (*ISWC 2016*), pp.72–79 (Sep. 2016).

[13] Blank, P., Hoßbach, J., Schuldhaus, D. and Eskofier, B.M.: Sensor-based Stroke Detection and Stroke Type Classification in Table Tennis, *Proc. 19th International Symposium on Wearable Computers* (*ISWC 2015*), pp.93–100 (Sep. 2015).

[14] Myers, C. and Rabiner, L.: A Comparative Study of Several Dynamic Time Warping Algorithms for Connected Word Recognition, *The Bell System Technical Journal*, Vol.60, pp.1389–1409 (Sep. 1981).

[15] Farzin, F., Rivera, S.M. and Whitney, D.: Time Crawls: The Temporal Resolution of Infants' Visual Attention, *Psychological Science*, Vol.22, pp.1004–1010 (Aug. 2011).

Hiroshi Yamada received his B.Eng. and M.Eng. degrees from Kobe University in 2015 and 2017, respectively. His research interests are wearable computing and ubiquitous computing.

Kazuya Murao is an Associate Professor at the College of Information Science and Engineering, Ritsumeikan University, Japan. He received his B.Eng., M.Info.Sci., and Ph.D. degrees from Osaka University in 2006, 2008, and 2010, respectively. Prof. Murao is working on wearable computing, ubiquitous computing, and context aware systems. He is a member of IEEE and ACM.

Tsutomu Terada is an Associate Professor at the Graduate School of Engineering, Kobe University. He received his B.Eng., M.Eng., and Ph.D. degrees from Osaka University in 1997, 1999, and 2003, respectively. He has been an Assistant Professor at Cybermedia Center of Osaka University and a Lecturer in 2000 and 2005, respectively. He is currently investigating wearable computing, ubiquitous computing, and entertainment computing. He is a member of IEEE, ACM, and IEICE.

Masahiko Tsukamoto is a Professor at the Graduate School of Engineering, Kobe University. He received his B.Eng., M.Eng., and Ph.D. degrees from Kyoto University in 1987, 1989, and 1994, respectively. From 1989 to 1995, he was a research engineer of Sharp Corporation. From 1995 to 1996, he has been an Assistant Professor at the Department of Information Systems Engineering, Osaka University and since 1996, he has been an Associate Professor at the same department. He is currently investigating wearable computing and ubiquitous computing. He is a member of eight learned societies, including ACM and IEEE.