Challenges in Measuring Partner Dancing Skills via Wearable Accelerometers

Lorenzo A. Rossi
Department of Computer Science
University of Houston

Shengrong Yin
Department of Computer Science
University of Houston

Omprakash Gnawali
Department of Computer Science
University of Houston

ABSTRACT

Social partner dancing is a fun but challenging activity requiring different motion related skills. Common criteria used by professionals to assess the quality of this type of dancing fall in the categories of timing, technique and teamwork (often referred to as "the 3 Ts") and variety of motion (*i.e.* "moves"). We focus on the teamwork and variety skills for practitioners of a type of Swing dancing called Balboa. Our dataset consists of the wearable accelerometer data collected from the participants to 3 different Balboa social dance contests. Panels of professional dancers judged the contests. Later, some of those professional dancers evaluated the skills of each participant by watching video recordings of the contests. We propose four novel measures for teamwork and motion variety and we evaluate them versus the expert assessments and also activity based labels. Our preliminary results show that the measures can be useful for activity recognition and somehow useful for teamwork assessment.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; • **Computer systems organization** → *Sensors and actuators*.

KEYWORDS

accelerometry; wearable sensors; arts; dance; skill level assessment; activity recognition.

ACM Reference Format:
Lorenzo A. Rossi, Shengrong Yin, and Omprakash Gnawali. 2020. Challenges in Measuring Partner Dancing Skills via Wearable Accelerometers. In *The 6th ACM Workshop on Wearable Systems and Applications (WearSys '20), June 19, 2020, Toronto, ON, Canada.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3396870.3400013

1 INTRODUCTION

Social dancing is a major recreational activity practiced around the world. By social dancing we refer to partnered dances (*i.e.* that require a leader and a follower) and that are practiced as a social activity, generally without choreography. Example of such dances are Swing dancing, Tango, Salsa, and Ballroom. Generally, dance practitioners may need hundreds of hours of lessons and finalized activity to achieve advanced proficiency level. The social importance of dancing and the desire of many dancers to improve their dancing provides a tremendous opportunity (possibly impacting billions of people) to make the world a happier place using technology.

The broad goal of this work is the application of wearable computing to social dancing in order to assess the skill levels of the dancers and provide them feedback. In particular, we perform our experiments on a type of Swing dancing called Balboa (or Bal Swing)[1] but the approach should be applicable to other types of dancing. We focus on 2 types of dancing skills: teamwork and motion variety. Partner dancing has many analogies with a conversation between two people. Within this analogy, teamwork is about how well the speakers can dialog with each other. The equivalent of motion variety is the richness of the vocabulary of a speaker. This suggests that designing automated methods to assess proficiency level of dancers is a very challenging problem. Similarly, the automatic evaluation of the quality of a dialogue between two subjects would be a very challenging artificial intelligence problem.

Existing work in accelerometer-based quantification and recognition of common daily activities such as walking, running and standing [1, 9, 12, 15], analyzes activities performed by a single person. Social dancing is done with a partner. Thus, individual-oriented analytics, such as those in the activity recognition literature, do not give full insight about the dynamics during a dance. We need to analyze the accelerometer data of both the partners to get a complete picture of the skill level of a dancer. Prior works in accelerometry-based analysis of dance (e.g. [11], [5]) also focus on individual performance and in controlled environment.

In this study, we instrument a dancer with a single device usually on the back in a packaging design that does not interfere with dancing motion. We collected data via 3 public dance competitions with a total of 41 instrumented participants, judged by professional dancers. Later, skills of each individual participant such as teamwork and motion variety were assessed by a panel of expert dancers by watching video recordings of the contests. Based on the accelerometer data collected from the competitors, we design measures for teamwork and variety for a dance segment (song). We then evaluate the measures versus aggregated scores from the experts and lower level semantic labels depending on specific subtype of dance performed during the contests.

We make these contributions:

- Collected the largest set of accelerometer data of social partner dancing and its quality as judged by expert professional dancers.
- Proposed novel measures to represent teamwork and motion variety of the dancers.
- Evaluated above measures versus expert assessments.

WearSys '20, June 19, 2020, Toronto, ON, Canada

ACM ISBN 978-1-4503-8013-3/20/06…$15.00
https://doi.org/10.1145/3396870.3400013

[1] en.wikipedia.org/wiki/Balboa_(dance).

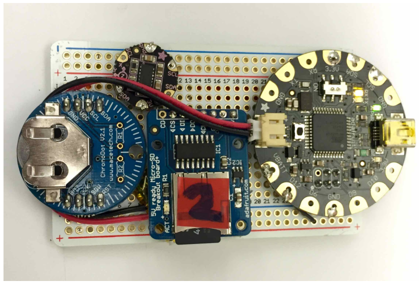

Figure 1: The wearable device: board, sensor, clock, micro-SD board.

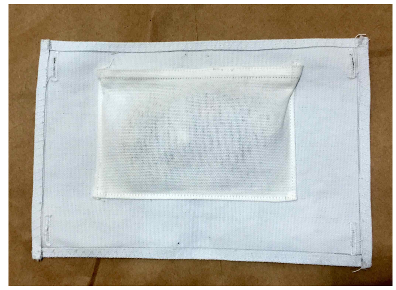

Figure 2: The back side of the fabric packaging with the wearable sensor inside the pocket.

2 RELATED WORK

Several works on the use of accelerometers to recognize different common physical activities [14]. Acceleration can be captured either using accelerometers in a custom device or through smart phones. Accelerometers (including those on smartphones) have also been used to detect activities, activity levels, and physical world sensing, e.g., potholes on the roads or traffic conditions. Using Gyroscopes with accelerometers could lead to better human activity recognition performance [7]. We use accelerometer as a minimal and simple 3 dimensional IMU-based sensing system. There are other related efforts to use a single accelerometer to detect quality of gestures in expressive body movements [22]. In a similar spirit, our work utilizes single accelerometer to capture data that can be used to assess the quality of dancing.

Wearable sensor data analysis has also been used recently in sports to give athletes feedback to improve the quality of their performance (*e.g.*, accelerometery data for traditional Japanese sword skills [2] or EMG data for pedaling skills [20]) and in the medical domain to assist the recovery of patients [10].

There is a lot of work in analyzing the dance motion, particularly the rhythm. Some works focused on recognizing and analyzing dance steps and gestures [6], [16]. Some research analyze the motions of a single dancer, even when a group of dancers are instrumented [17], or with a limited number of participants [21]. The latter work proposes an approach to measure rhythm via accelerometers for students of a Brazilian partner dance performing specific sets of exercises. A study that instrumented 20 non-professional adult dancers with accelerometers and heart rate telemetry tried to understand individual performance in a lab setting [5]. Compared to this body of research, we instrument dancers in real social dancing public competitions and obtain labels from dance professionals who judged those competitions. We focus on estimating the dancing skill levels, also via features designed to understand the quality of the interaction between the two partners involved in a dance *i.e.* the lead and the follow, and we rely on the largest in-depth accelerometer dataset for a social dance.

3 SENSOR DEVICE

We design a wearable device to collect acceleration data from dancers. The accelerometer packaging was designed to be as unobtrusive as possible during social dancing. The device is based on a Flora platform from Adafruit. It weighs 34 grams and measures 102mm x 51mm x 12mm. We connect LSM303DLHC, a 3D linear accelerometer, to the Flora board. The sensor has a programmable acceleration scales of ±2g, ±4g, ±8g, ±16g. We use a sampling rate of 100 Hz, then reduced to 50 Hz to limit the inter-sample jitter from about 1.5ms to 0.5ms. We use Chronodot, which has a DS3231 real time clock with an independent power supply. We use a 4GB micro-SD board to store sensor data and a 3.7v 150mA Lithium Ion Polymer battery to power the device. This battery can power the device for more than 3 hours. The device is shown in Figure 1.

Packaging. During competitions, the dancers pin a paper or fabric bib number on their back to be identified by the judges. We designed a bib number made with canvas fabric. The bib has a small pocket on the back as shown in Figure 2 where we insert the sensor device. The dancers pin this bib number just like any other bib number they normally pin during a competition. This sensor package design is unobtrusive and does not interfere with dance moves or the aesthetics of the dance clothes.

Overall, the sensor devices meet the needs and aesthetics of social dancing accelerometer data capture.

4 THE DATASET

We collected accelerometer data from the competitors in 3 dance contests. The data collected from each session is called a collection. During each session, each competitor wore a triaxial accelerometer device and danced 4 or 5 songs switching partner after each songs (Table 1). The duration of each segment of dancing varied from 1 to about 3 minutes. Besides, for each participant we have the following types of evaluations performed by panel of experts (dance professionals):

- rankings produced during the contests
- skill assessments produced via videos after the contests.

For each song danced by a participant, the devices recorded 3 accelerometer readings across the x, y, z relative directions, relative and global time stamps at a sampling rate of either 100Hz or 50Hz. The rationale for using only one accelerometer per dancer is given by the nature of our deployments "in the wild" and not in a lab type of settings. We have IRB approval for this research. The entire dataset was acquired well before the beginning of the social distancing regulations due to the Covid-19 pandemics.

Generally for the first one or two songs of each contest, participants were requested to perform a particular style of the dance called

Collection	Samples	Partnered	Leads	Follows	Songs	Judges	Kappa
1	56	48	6	6	4	(3,3)	(0.67, 0.47)
2	65	60	6	7	5	(3,3)	(0.3, 1.0)
3	64	64	8	8	4	(3,3)	(0.66, 0.64)
Totals	185	172	20	21	13		

Table 1: Summary of the contests organized to collect the data for this work. We specify the counts of judges for leads and follows in pairs: *e.g.* $(3,3)$ means 3 judges for the leads and 3 judges for the follows. The agreement scores (Fleiss Kappa) among the judges are also included. In some cases, competitors declined to wear the devices or their data collection was faulty: hence the difference between (total) samples and partnered ones.

Pure Balboa that allows only close position with chest to chest connection between a lead and a follow preventing them to spin. We use the indicator variable for Pure Balboa songs as a binary activity label. During the remaining songs in contests, the participants were free to break away from close position and to spin (*i.e.* dancing standard Balboa, also known as Bal-Swing).

The readers can check out a short video clip and other supplementary material via a **GitHub repository**[2] showing a song in one of the competitions in our data collections. The accelerometers are inside the competition numbers worn by the participants. A red LED light, indicating that the devices are recording data, can be noticed underneath the competition numbers closest to the camera. Some of the judges are visible holding clipboards around the dance floor. One can realize also how challenging it would be to track all the couples on the packed dance floor via video cameras.

4.1 Partner dancing evaluation

Each dancer was evaluated separately by 3 judges (dance professionals) for his/her overall performance in a contest consisting of 4 to 5 short songs with either the same or different partners. Each judge ranks the dancers. The overall aggregated rankings are computed via the Relative Placement method (which is a sort of majority voting). Relative Placement is preferred to simpler scoring methods, because it is robust against outliers (*e.g.* due to biases of judges). We binarized the overall scores to divide up the top from the bottom of the rankings (*i.e.* the top ranked dancers were labeled as '1'). Besides, we assigned the binary score of each dancer to each of his/her dances in the session. E.g. if dancer A has a binary score of 0, all A's dances is scored 0 and evaluated as a different sample in the training/classification stage.

On a second stage, some of dance professionals who acted as judges were also shown the videos of the contests and rated each participant in their timing, technique, teamwork and variety of motion. In particular, 3 evaluators rated the leads and 3 more rated the follows.

5 SKILL MEASURES

We segmented the collected accelerometer sequences to remove the intervals without dancing motion preceding and following the start and the end times of the songs in the contests. Then we preprocessed the sequences to remove gravitational and very low frequency (< 20 BPM) components of the acceleration.

We use the terms **dominant beat** and **main beat** to indicate respectively the beat of the dancer over a particular time segment and the beat of the harmonic of the dominant beat closest to the beat of the music. For example, Figure 3 shows main and dominant beats for a dancer moving at 118 BPM to a 244 BPM song and the module of the associated discrete Fourier transform. The extraction of the beats of dancers is not new in accelerometry [13].

The absolute vertical component acc_z of the acceleration[3] turned out to be the most effective way to estimate the rhythm of the dancers, although also the horizontal components, A_x and A_y reflect the main dancing rhythm more or less remarkably depending on the individual style of the dancers.

Given a sequence of accelerometer samples $\{acc_x(t), acc_y(t), acc_z(t)\}$, we first compute the module of the 3 components of the acceleration, *i.e.*:

$$acc_r(t) = \sqrt{\sum_t (acc_x(t)^2 + acc_y(t)^2 + acc_y(t)^2)}. \qquad (1)$$

The jerk is the first derivative of the acceleration and for the radial case it is computed as:

$$j_r(t) = acc_r(t) - acc_r(t-1). \qquad (2)$$

Finally the normalized radial jerk feature is given by:

$$j_r^{(n)} = \frac{\|j_r(t)\|_2}{\|acc_r(t)\|_2}. \qquad (3)$$

Intuitively good dancing motion is characterized by low jerk levels: hence the choice of jerk for the above metric.

5.1 Teamwork

It has been shown that determining if two accelerometers are on the same body is a hard task [3]. Our problem of extracting some useful information from a pair of accelerometers placed on two *different* bodies, although dancing together, is even more challenging, because in partner dancing there are intentional breaks from the synchronicity between leader and follower. We attempt to measure quality of the interaction between dance partners in term of synchronicity to the musical rhythm and smoothness of their joint motion.

[2] github.com/elleros/dance-accelerometry.

[3] acc_z is the component parallel to the gravity acceleration and so perpendicular to the ground. For the sake of the dancer tempo extraction, it can be approximated by the accelerometer component with the largest median value.

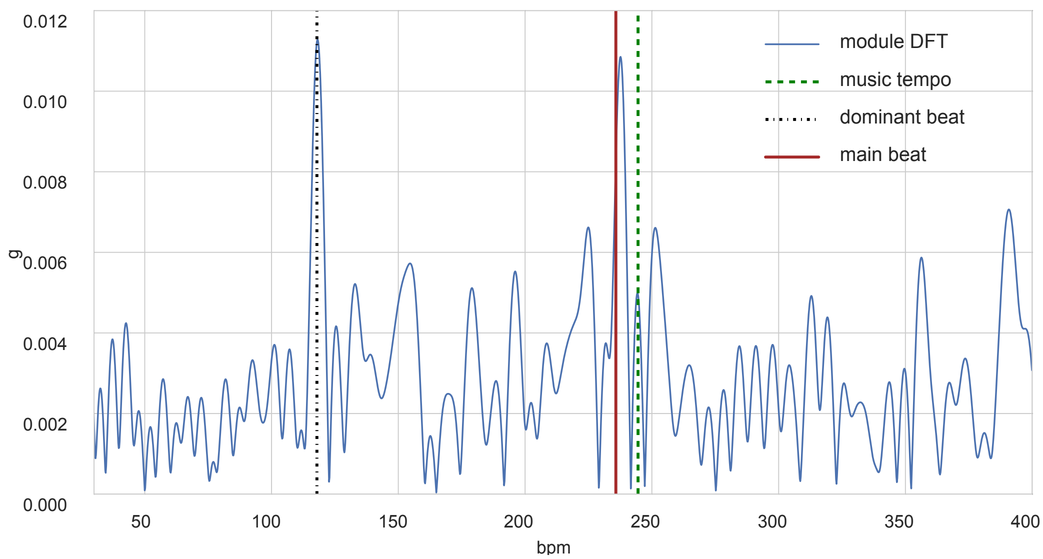

Figure 3: The module of the discrete Fourier transform (DFT) of 12 seconds of the vertical component of the acceleration sequence for a Balboa dancer. The gravitational component of the acceleration has been removed. The *music tempo* is 244 BPM. The dancer is moving half tempo at about 118 BPM (*dominant beat*).

We extract a sequence of dominant beats of a dancer over a sliding window on the accelerometer sequences. The length of the window is adjusted according to the tempo of the underlying music, to take into account that some dancers may be dancing half tempo, especially if the music is too fast:

$$\mathbf{v} = \{b_1, b_2, ...\}, \tag{4}$$

where b_i are the main beats for each windowed accelerometer sequence. We then express the **Tempo Similarity** measure of a pair of dancers on a certain song as the cosine similarity of their vectors of main beats, *i.e.*:

$$S_b = \frac{\mathbf{v}_l \cdot \mathbf{v}_f}{\|\mathbf{v}_l\| \|\mathbf{v}_f\|}, \tag{5}$$

where $\mathbf{v}_l$ and $\mathbf{v}_f$ are the sequences of main beat differences for lead and follow.

The radial **Jerk Similarity** is expressed as the peak of the cross correlation of the radial jerks (2) of the two partners divided by product of the L_2 norms of each radial jerk. This is intended to express the smoothness of lead and follow motion. Let $j_r^{(l)}$ and $j_r^{(f)}$ be respectively the radial jerks of a lead and a follow dancing together. The radial **Jerk Similarity** $r_{\max}(j_r^{(l)}, j_r^{(f)})$ is given by:

$$r_{\max}(j_r^{(l)}, j_r^{(f)}) = \max_{\tau} \frac{R_{(j_r^{(l)}, j_r^{(f)})}(\tau)}{\|j_r(t)\|_2 \|j_r(t)\|_2}, \tag{6}$$

where $R(\tau)$ is the cross correlation.

We compute the above features for a couple and assign the same value to both leader and follower.

5.2 Motion variety

For this measure, our goal is to find an approximate measure of the "number of moves" performed by a dancer during a song. To this end we propose two measures: ZCR Phase and Motion Diversity.

The **ZCR Phase** is defined as the median of the zero crossing rate (ZCR) of the phases of the DFT of windowed accelerometer sequences. Typical patterns in Balboa involve rotations of the partners in close position on the dance floor. The *ZCR Phase* is intended to measure variations of such rotations (*e.g.* from clockwise to counter-clockwise), as those indicate diversification of the dance.

For the **Motion Diversity**, we compute the module of the 3 components of the acceleration, eq. (1), over segment of dancing. We partition the module over n_w temporal non overlapping windows $\{w_1, w_2, ...\}$. We then compute the module of the discrete Fourier Transform for each window and extract the values at the main beat and its harmonics and fractional harmonics. Each temporal subsegment w_i is represented by a vector $\mathbf{v}_i$. We aggregate such vectors into a matrix $\mathbf{V}$. We define the motion diversity μ_{md} as the number of singular values of $\mathbf{V}$ that are greater than 1:

$$\mu_{md} = |s_i(\mathbf{V}) : s_i(\mathbf{V}) > 1|, \tag{7}$$

where $s(\mathbf{V})$ is a singular value of $\mathbf{V}$. Similarly to the case of *ZCR Phase*, the intuition behind *Motion Diversity* lies in domain knowledge: advanced Balboa practitioners have the ability to switch motion *e.g.* from full to half tempo: slower motions serve to highlight certain segments of the dance. Hence we assume a relationship between the harmonic components of the motion and the skill level.

6 RESULTS

As ground truth to evaluate the proposed measures, we considered activity and skill based sets of labels. The latter types of labels require expert assessments. The activity related labels take a value of 1 if the style requested for a song in the contest was Pure Balboa and 0 otherwise (Sec. 4). The skill level labels are obtained by aggregating binarized competition rankings with expert skill assessments for teamwork and variety respectively. Since during Pure Balboa, the dancers are constrained in close position for the whole song, one should expect relatively lower variation of motion and higher level of teamwork. Therefore the Pure Balboa activity labels can be seen as low semantic level proxies for the skill based labels. To evaluate the previously defined measures for teamwork and motion variety, we compute the Pearson correlation statistics.

The results are in Table 2. Overall the proposed measures are moderately correlated with the activity labels, weakly correlated with the teamwork skill assessments and very weakly correlated with the variety of motion skill assessments. For instance, the *jerk similarity* measure can capture whether a leader and a follower are dancing in close position throughout a song (0.63 correlation). The correlation statistics obtained for the Pure Balboa recognition task seem reasonable considering also that usually standard Balboa dance includes segments of Pure Balboa anytime dancers are in close position. The rhythm variety measure shows some capacity to measuring teamwork skill level (0.26 correlation). The assessment of variety skill level from accelerometer data seems the hardest among the three tasks, possibly because *e.g.* a lot of physical motion by unskilled dancers may still appear as lack of variety to experts, while subtle motions performed by advanced dancers may be considered rich in variety. Moreover, it is worth noting that in Balboa dance, there are multiple dance moves that exhibit similar movements of the core of the body (e.g., switching lead and follow roles) and hence could have similar accelerometer trace but could be determined by the experts to be different moves, thus adding further challenge to the variety skill evaluation tasks.

We computed the Fleiss Kappa to measure the agreement among the dance experts who judge the contests and evaluated the skills from the videos, obtaining values between 0.1 (slight agreement) and 0.25 (low end of fair agreement). The range of Fleiss Kappa for the live judging of the contests was usually between 0.3 and 0.66 (Table 1), better than for the video based skill assessments, but still below substantial level of agreement. The generally low values of Kappa suggest that dancing skill assessment tasks are hard for human experts and partially explain the very low correlation values between our measures and the assessments of the judges.

Measure	Pure Bal Activity	Teamwork Skill	Variety Skill
jerk similarity	0.63	0.10	
tempo similarity	-0.27	0.26	
motion diversity	-0.22		0.03
ZCR phase	0.39		0.02

Table 2: Correlation coefficients between skill measures and activity (Pure Balboa) and skill level labels.

7 DISCUSSION

Judging is a hard task: In social dancing competitions, judges may not have uniform criteria or could miss relevant segments for having to watch multiple couples during the same short time frame, or they may have biases. The phenomenon of inconsistencies among judges have been studied for Ballroom Dancing [19], which is in a more controlled setting than the type of dance considered in our project. Similar challenges have been reported in the literature in the context of pair figure skating [18].

Social dancing and conversation: One can see an analogy between social (non choreographed) partner dancing and a conversation between two individuals [8]. The type of dance (*e.g.* Swing, Salsa, Tango etc.) represents the language of the conversation. The underlying music provides the topic. We can assume the verbal skill of an individual to be more or less constant across different chunks of conversations and similarly somebody's dancing ability to be uniform across different songs. The evaluation of somebody's conversational ability is clearly much easier if we have the conversation transcripts of both actors of the conversation. Similarly the evaluation of someone's dancing ability is easier if we have access also to the data of the partner. Due to the rise of conversational agents, there has been interest in creating a framework for evaluating interactions aspects of dialog using metrics such as engagement, coherence, and conversational depth [23]. It would be interesting to seek analogies of such metrics in the context of evaluation of partner dancing.

Sensor placement rationale: There are many choices in sensor placement when instrumenting a dancer. We should consider efficacy in terms of capturing the relevant motion characteristics as well as potential interference in dancing. Generally, in partner dancing torso dictates the motion, although arms and legs may look more active. As a matter of fact, instructors recommend to lead with the whole body (rather than with the arms) creating *connection* [4], which allows the dancers to communicate and synchronize their dance movements. Therefore, more relevant sensing data relevant to partner dynamics will be obtained by a sensor on a dancer's back compared to the one on his/her limbs. Placement of the device in on body parts such as shoulders, belly on chest would interfere with the motions of the dancers (*e.g* in close position). In addition, in the type of dancing we consider in this study, the dancers typically do not wear loose clothing. Thus, the sensor can remain close to the body and capture body movements rather than the movements of loose clothing which may not be correlated to dance movements.

8 CONCLUSION AND FUTURE WORK

Teamwork and motion variety are important skills in social partner dancing. We evaluated four novel measures of teamwork and motion variety obtained from accelerometry data for a number of dance contest competitors. As ground truth, we computed scores from the aggregation of expert evaluations on the dancers (skill level labels). We also obtained activity based labels from a specific style of dancing performed during portions of the contests. The results show that the proposed measures are moderately correlated with the activity labels, weakly correlated with the teamwork skill assessments and very weakly correlated with the variety of motion skill assessments. This, along with low Kappa measures from the ratings of our experts, suggests that the problem of measuring social dancing skills from accelerometer data is very challenging. This is a preliminary partial building block towards the long term goal of developing intelligent assistants to dancers that can analyze accelerometer data, metadata of the background music and, if available, also the partner's data in order to provide an evaluation of the user's dancing and insights on weaknesses and strengths.

ACKNOWLEDGMENTS

We would like to thank Charlotte Scovill for making the canvas pockets for the volunteers to wear the accelerometer devices and all the dancers and judges who participated in our data collections.

REFERENCES

[1] Akram Bayat, Marc Pomplun, and Duc A Tran. 2014. A study on human activity recognition using accelerometer data from smartphones. *Procedia Computer Science* 34 (2014), 450–457.

[2] Woong Choi, Sho Mukaida, Hiroyuki Sekiguchi, and Kozaburo Hachimura. 2008. Quantitative analysis of Iaido proficiency by using motion data. In *2008 19th International Conference on Pattern Recognition*. IEEE, 1–4.

[3] Cory T Cornelius and David F Kotz. 2012. Recognizing whether sensors are on the same body. *Pervasive and Mobile Computing* 8, 6 (2012), 822–836.

[4] Joseph Daniel DeMers. 2013. Frame matching and δP t eD: a framework for teaching swing and blues dance partner connection. *Research in Dance Education* 14, 1 (2013), 71–80.

[5] Pablo A Domene and Chris Easton. 2014. Combined triaxial accelerometry and heart rate telemetry for the physiological characterization of Latin dance in non-professional adults. *Journal of Dance Medicine & Science* 18, 1 (2014), 29–36.

[6] Abu Zaher Md Faridee, Sreenivasan Ramasamy Ramamurthy, H M Sajjad Hossain, and Nirmalya Roy. 2018. HappyFeet: recognizing and assessing dance on the floor. In *Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications (HotMobile '18)*. ACM, New York, NY, USA, 49–54.

[7] Anna Ferrari, Daniela Micucci, Marco Mobilio, and Paolo Napoletano. 2019. Human Activities Recognition Using Accelerometer and Gyroscope. In *Ambient Intelligence*. Springer, Cham, 357–362.

[8] Sommer Elizabeth Gentry. 2005. *Dancing cheek to cheek: Haptic communication between partner dancers and swing as a finite state machine*. Ph.D. Dissertation. Massachusetts Institute of Technology.

[9] Piyush Gupta and Tim Dallas. 2014. Feature selection and activity recognition system using a single triaxial accelerometer. *IEEE Transactions on Biomedical Engineering* 61, 6 (2014), 1780–1786.

[10] Shane Halloran, Lin Tang, Yu Guan, Jian Qing Shi, and Janet Eyre. 2019. Remote monitoring of stroke patients' rehabilitation using wearable accelerometers. In *Proceedings of the 23rd International Symposium on Wearable Computers*. ACM, 72–77.

[11] E. A. Heinz, K. S. Kunze, M. Gruber, D. Bannach, and P. Lukowicz. 2006. Using Wearable Sensors for Real-Time Recognition Tasks in Games of Martial Arts - An Initial Experiment. In *Symposium on Computational Intelligence and Games*. IEEE, 98–102.

[12] J Ingebrigtsen, I Stemland, CS Christiansen, S Jorgen, C Hanisch, P Krustrup, and A Holtermann. 2013. Validation of a commercial and custom made accelerometer-based software for step count and frequency during walking and running. *J Ergon* 3, 119 (2013), 1–6.

[13] M. Kuhn, R. Wattenhofer, M. Wirz, M. Fluckiger, and G. Troster. 2011. Sensing Dance Engagement for Collaborative Music Control. In *International Symposium on Wearable Computers*. IEEE, 51–54.

[14] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. 2011. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter* 12, 2 (2011), 74–82.

[15] Oscar D Lara and Miguel A Labrador. 2013. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials* 15, 3 (2013), 1192–1209.

[16] Sohaib Laraba and Joëlle Tilmanne. 2016. Dance performance evaluation using hidden Markov models. *Computer Animation and Virtual Worlds* 27, 3-4 (2016), 321–329.

[17] Eric Lee, Urs Enke, Jan Borchers, and Leo de Jong. 2007. Towards Rhythmic Analysis of Human Motion Using Acceleration-Onset Times. In *International Conference on New Interfaces for Musical Expression*. ACM, New York, NY, USA, 136–141.

[18] Marilyn A Looney. 2004. Evaluating judge performance in sport. *Journal of Applied Measurement* 5, 1 (2004), 31–47.

[19] Jerneja Premelč, Goran Vuckovic, Nic James, and Bojan Leskošek. 2019. Reliability of judging in DanceSport. *Frontiers in psychology* 10 (2019), 1001.

[20] Osamu Saisho, Makoto Nakayama, Kentaro Tanaka, and Daigoro Yokozeki. 2018. Practical Pedaling Skill Items Extraction for Efficient Pedaling Training with Surface EMG Wear. In *International Symposium on Wearable Computers (ISWC '18)*. ACM, New York, NY, USA, 76–79.

[21] Augusto Dias Pereira dos Santos, Lie Ming Tang, Lian Loke, and Roberto Martinez-Maldonado. 2018. You Are Off The Beat! Is Accelerometer Data Enough for Measuring Dance Rhythm?. In *Proceedings of the 5th International Conference on Movement and Computing (MOCO '18)*. ACM, New York, NY, USA, Article 12, 8 pages.

[22] Diego Silang Maranan, Sarah Fdili Alaoui, Thecla Schiphorst, Philippe Pasquier, Pattarawut Subyen, and Lyn Bartram. 2014. Designing for Movement: Evaluating Computational Models Using LMA Effort Qualities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 991–1000.

[23] Anu Venkatesh, Chandra Khatri, Ashwin Ram, Fenfei Guo, Raefer Gabriel, Ashish Nagar, Rohit Prasad, Ming Cheng, Behnam Hedayatnia, Angeliki Metallinou, et al. 2018. On evaluating and comparing conversational agents. *arXiv preprint arXiv:1801.03625* 4 (2018), 60–68.