

Exploring in-the-Wild Game-Based Gesture Data Collection

Kiyoshi Oka¹(✉), Weiquan Lu², Kasım Özacar¹, Kazuki Takashima¹,
and Yoshifumi Kitamura¹

¹ Research Institute of Electrical Communication,
Tohoku University, Sendai, Japan
oka@riec.tohoku.ac.jp

² National University of Singapore, Singapore, Singapore
icd-office@ml.riec.tohoku.ac.jp

Abstract. This paper presents an automatic 3D gesture collection concept and architecture based on a rhythm game for public displays. The system was implemented using an off-the-shelf gesture controller, was deployed on a public vertical screen, and was used to study the effects of alternative gesture guidance conditions. In the evaluation presented, we examined how alternative gesture guidance conditions affect users' engagement. The study showed that demonstration animation (CDA) and tracking state feedback (TSI) each encourages sustained game engagement. The underlying concept and architecture presented here offer actionable UI design insight to help creating large gesture corpus from diverse populations.

1 Introduction

Gesture elicitation with off-the-shelf controllers such as Microsoft Kinect and Leap Motion has made gesture-based interfaces more accessible. Interfaces leveraging this class of controllers hold the promise for intuitive 3D interaction and provide benefits such as touch-free and remote interactions in public spaces [11]. For example, gesture-based interaction can benefit interaction in areas where touch input is not suitable. However, to develop gesture-based interfaces, UI designers may incorporate gesture recognition systems, often requiring a large corpus of training data [1–3]. In many cases, it is difficult to efficiently collect a large corpus of training data in a short period of time, representing a limitation in the use of gesture elicitation controllers.

This paper proposes and validates a low-cost concept and architecture (Fig. 1) to collect an in-the-wild gesture corpus from a large and potentially diverse user population. The proposed system is based on a rhythm game combined with a Walk-Up-and-Use Display [4, 5]. The system consists of a sensor for detecting 3D hand gestures and a computer with a large display for running a game. The system can be deployed in a public space and store gestures to an online database, and the collected data may then be used to make a large gesture corpus that would help development and evaluation of robust gesture recognition algorithms. In such systems, however, users of the system should be informed of how to use such a display as they approach it, while

also being attracted to use the system, which can be a challenging task for UI designers [4]. For in-the-wild settings, the systems should (a) teach users how to interact with the system and (b) help sustaining user engagement at least of one full game session.

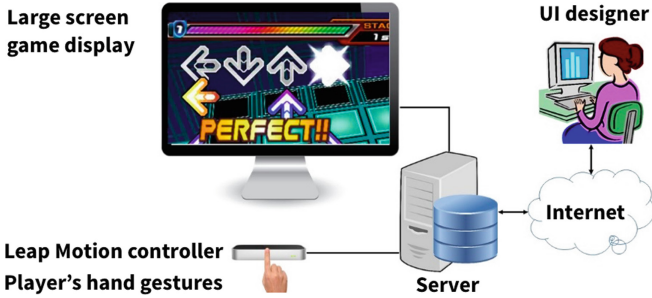


Fig. 1. Proposed concept and system architecture.

To address these challenges, we developed a gesture collection system using a rhythm game that can be deployed in-the-wild, and explored the types of guidance would result in greater sustained user engagement in opportunistic data collection. The game-based gesture collection system, Gesture Revolution (GGR), was designed for a Walk-Up-and-Use Display to serve as a platform for in-the-wild studies. We then conducted an in-the-wild study using GGR to investigate the effects of different guidance conditions on user's total engagement process. The main contributions of this paper are (a) describing design and implementation of an automatic gesture collection system using a simple rhythm game for a Walk-Up-and-Use Display and its fundamental benefits, and (b) describing a three-week user study that found that the guidance conditions with Contextualized Demonstration Animation (CDA) and Tracking State Indicator (TSI) result in more correct and sustained user gesture input.

2 Related Work

There have been successful examples of such in-the-wild gesture studies. For example, Hinrichs [10] carried out a field study at an aquarium to investigate how visitors interact with a large interactive table, and found that users' choice and use of gestures were affected by the interaction and social contexts. Walter [11] compared three strategies in revealing mid-air gestures in interactive public displays, and found that 56% of users were able to perform gestures with spatial division. Marshall [5] used a Walk-Up-and-Use tabletop in-the-wild to study social interactions around such devices, and found that these interactions were very different from those in lab settings. Most recently, Ackad [8] used an in-the-wild study to explore whether their system design supported learning, how their tutorial feedback mechanisms supported learning, and the effectiveness for browsing hierarchical information. Based on this related work, we use the in-the-wild approach as it enabled us to collect large and realistic data from diverse users if proper guidance was provided.

With respect to guidance, there were also many studies that have used the concept of “gesture guidance” [4, 7–9]. Gesture guidance systems are displays that show the gesture commands that can be used by the user to interact with the system. Rovelto et al. [4] compared a dynamic gesture guidance system with a printed traditional static gesture guidance showing snapshots of gesture sequences in a lab-based study. They found that for simple gestures, the dynamic system did not necessarily significantly improve users’ ability to learn and perform the correct gestures, but for complex gestures, the dynamic guide did result in an improvement. While previous work addressed the learning of gestures in a lab-based study, we are more interested in the effects of guidance conditions on the engagement process for an in-the-wild setting.

3 Game-Based Gesture Collection

We implemented an in-house design large-display game called Gesture Revolution (GGR) and designed it as an in-the-wild study platform that enabled passers-by to interact with the game by using simple stroke gestures.

3.1 Creating Gesture Revolution (GGR)

We studied several different game genres that use body or hand movements, and decided to base our game on the rhythm and dance genre, made popular by Konami’s Dance Dance Revolution [6]. This game concept is simple and offers a range of game design possibilities, while providing sufficient methods for conducting experiments by controlling the game and constraining gesture interactions.

We selected four simple hand gestures for collection: swipe up, swipe down, swipe left, and swipe right. These gestures were selected because they appeared simple to perform, but users could perform them in a wide variety of ways, thereby presenting a challenge for gesture recognition systems. Such gestures could be reliably detectable by the Leap Motion sensor, and they could be incorporated into more complex gestures.

We now describe the gameplay of GGR. When the game starts, an arrow appears at the top of the screen, and starts moving downwards (Fig. 2(a)). The arrow indicates which gesture is required for that part of the game. When the arrow reaches the bottom, the player must complete the correct gesture to score points (Fig. 2(b)). The player scores a different amount of points, depending on when s/he completes the gesture. The closer s/he is to the bottom of the screen, the more points s/he receives. However, the player receives no points if s/he performs the gesture too early (too far from the bottom of the screen), too late (the arrow moves out of the screen), or s/he performs wrong gestures (Fig. 2(c)). Depending on the scores, the players would receive different visual feedback (Bad, Good, Great, Perfect and Wrong Gesture). We focused only on visual feedback as it sufficiently shows the user the state of gesture input, and we plan to use different channels (e.g., sound) in a follow-up study.

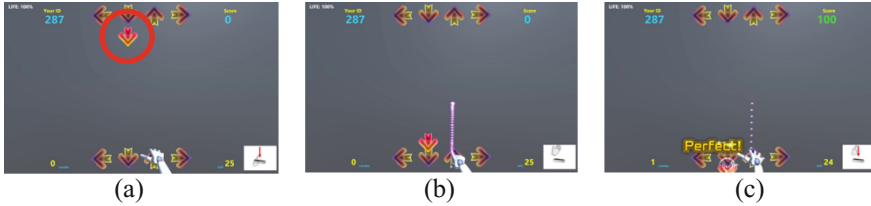


Fig. 2. Sequence of game dynamics (left to right): (a) An arrow appears at the top of the screen. (b) As the arrow moves down, the player must make the required gesture in order to score points. (c) If timed correctly, the player will receive the maximum score.

3.2 System Architecture

Our prototype of the system was implemented using a client-server architecture consisting of a database server backend, and a game client frontend (Fig. 1). A server hosted the game software (created in Unity) and MySQL database (part of the WAMP server package). We used a desktop computer with an Intel i7-6700 CPU and 8 GB RAM, running Windows Server 2012 R2. The 50-in. large plasma TV was used to display the game and visual feedback for players. The server was connected to the Internet, and therefore enabled researchers to remotely access the database, update the parameters/variables of the game, and monitor the gameplay.

4 In-the-Wild Study

We conducted an in-the-wild study to evaluate the potential of game-based opportunistic gesture data collection and the effects of the three traditional guidance types on user engagement. The in-the-wild approach was chosen in order to obtain diverse gesture data from many different users with minimal resources.

4.1 Implemented Guidance Conditions in GGR

We designed three guidance conditions, two levels and one control, based on the concept of Scaffolding Means (Instructing, Modeling and Feeding-back) [19]:

- Looping Introductory Animation (LIA). The Looping Introductory Animation (Fig. 3(a)) included detailed instructions on how, when and where all the gestures were needed in the game. The animation was played in a loop until a user interrupted the loop by placing his/her hand over the Leap Motion sensor. This was the implementation of Instructing [19], and served as a standard guidance that could quickly show all necessary steps of the game.
- Contextualized Demonstration Animation (CDA). As shown in Fig. 3(b), this animation sequence was overlaid on the game view just before a user was required to perform a specific gesture for the first time. This was a form of Modeling [19], and provided dynamic and explicit models that users could imitate when performing required gestures.

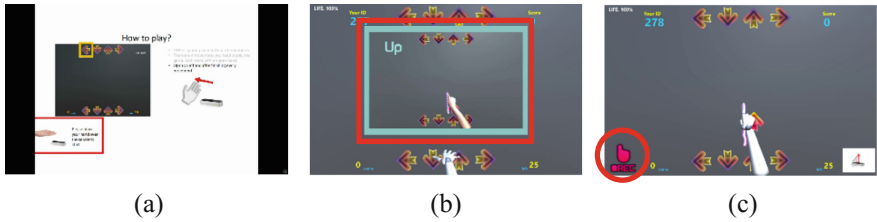


Fig. 3. Implemented guidance types: (a) Looping Introductory Animation (LIA), (b) Contextualized Demonstration Animation (CDA), (c) Tracking State Indicator (TSI).

- Tracking State Indicator (TSI). This indicator was shown at all times, and had two states. One state indicated that the system was detecting the user’s hand gesture successfully (Fig. 3(c)), and another indicated that the system was not detecting the user’s hand gesture successfully. This was a form of Feeding-back [19], and served as an implicit guidance that would give users a fundamental understanding of how fingers are tracked.

The LIA was shown at the start of the game, in every guidance condition, as it was implemented as part of core functionality of the system. The other two guidance conditions were mutually exclusive. As the experiment was conducted over a span of three weeks, Week 1 was LIA alone (LIA), Week 2 was LIA with CDA (LIA + CDA), and Week 3 was LIA with TSI (LIA + TSI).

4.2 Deployment Venue

The system was deployed as a typical Walk-Up-and-Use Display. Figure 4 shows a snapshot of the system deployment with a group of passers-by playing the game. It consisted of a Panasonic 50-in. display mounted on a high display stand. The system was configured to power on at 08:30 am and shut down at 08:30 pm every day, as per requirements from the exhibition venue sponsors. A research institute within a local university was the exhibition venue. About 400 individuals passed through this venue every day, and about 25% were female, according to our visual observation. About 80% of human traffic moving in and out of the building had to pass by the exhibit, as it was placed between the entrance and elevators.



Fig. 4. System deployment.

4.3 Experiment Protocol

We used the following protocol:

1. A potential user walks into the venue.

2. His/her attention is drawn to the display, which shows the LIA explaining how to play the game, along with a disclaimer and statement of research intention. This animation plays on a loop until the game starts.
3. If the user decides to play the game, s/he is instructed to hover his/her hand over the Leap Motion for about 2 s to activate the game.
4. The system asks the user to specify his/her profile number or age and gender.
5. When the user hovers his/her hand over the “Play” button, the game starts.
6. After the game ends, the system asks the user to rate his/her enjoyment of the game on a scale of 1 to 5.
7. Once done, the system displays the user’s score on a leaderboard, generates a user profile number and shows it to the user. The user can then use the profile number when s/he plays the game next time. This allows us to track repeat users, and also allows users to keep track of their own progress.
8. Finally, the system returns to the initial LIA, awaiting the next user.

4.4 Measurements

In order to investigate which guidance conditions would result in greater user engagement, we measured five dependent variables.

1. Correctness was defined as the average number of correct gestures made by each player, over the total number of gestures required in the game session.
2. Percentage of partial quitters (Partial Quitters) was defined as the number of users who managed to perform at least one gesture successfully but decided to quit playing before completing the game, over the total number of users.
3. Percentage of successful task completions (Completions) was defined as the number of users who successfully completed the task (performing 25 gestures correctly), over the total number of users.
4. Score was defined as the total average score obtained by the users, with higher scores awarded for more challenging moves.
5. Average number of sustained successful gesture performances (Sustained) was defined as the average number of chained successful gesture performances.

5 Results

During the three-week experiment period, the system recorded a total of 171 unique users (Mean age = 26.6, SD = 9.57, 38 females). Based on the data, we conducted statistical analysis in terms of Correctness, Partial Quitters, Completions, Score and Sustained. We excluded the data from self-reported repeat user sessions, meaning that all the data was from users who only played for the first time, enabling us to conduct a between-subjects analysis.

Correctness. Figure 5(a) shows the average number of correct gestures made by each user, over the total number of gestures required in the game session. Since the data did not fit the assumption of normality, we conducted the Kruskal-Wallis test and found that different guidance types had a significant effect on correctness ($p < 0.01$).

The post-hoc Steel-Dwass test was carried out to compare the three guidance conditions and it revealed that the correctness was significantly higher in the LIA (Looping Introductory Animation) + CDA (Contextualized Demonstration Animation) condition than in the LIA-only condition ($p < 0.01$). We further conducted the Steel test that focuses on differences between the baseline (LIA) and each of the LIA+CDA and LIA+TSI (Tracking State Indicator) conditions, and found that the correctness was significantly higher in the LIA+CDA and LIA+TSI conditions, than in the LIA-only condition ($p < 0.01$, $p < 0.05$).

Partial Quitters. Figure 5(b) shows the percentage of partial quitters. We conducted a chi-square test in order to statistically compare the proportions across guidance conditions, and did not find any significant effect of different guidance conditions on partial quitters ($\chi^2 = 1.51$, $p \geq 0.05$).

Completions. Figure 5(c) shows the percentage of successful task completions. We used a chi-square test in order to statistically compare the proportions, and found that the different guidance types had a significant effect on completions ($\chi^2 = 7.60$, $p < 0.05$). Chi-square pairwise comparisons with Bonferroni corrections revealed that the number of completions were significantly higher in the LIA+CDA condition than in the LIA-only condition ($p < 0.05$).

Score. Figure 5(d) shows the average score. Kruskal-Wallis test revealed that different guidance types had a significant effect on the score ($p < 0.01$). The post-hoc Steel-Dwass test revealed that the score was significantly higher in the LIA+CDA condition than in the LIA-only condition ($p < 0.01$).

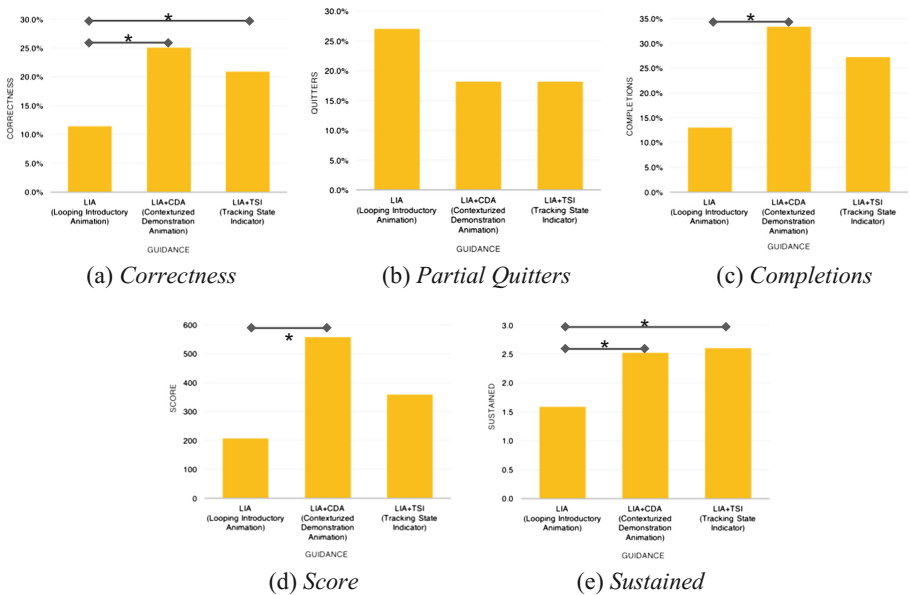


Fig. 5. Graphs of (a) Correctness, (b) Partial quitters, (c) Completions, (d) Score and (e) Sustained. * depicts $p < 0.05$.

Sustained. Figure 5(e) shows the average number of sustained successful gesture performances. Kruskal-Wallis test revealed that different guidance types had a significant effect on the number of sustained successful gesture performances ($p < 0.01$). The post-hoc Steel-Dwass test revealed that the LIA+TSI condition offered significantly more sustained successful gesture performances than the LIA-only condition ($p < 0.01$). We further conducted the Steel test that focused on differences between baseline (LIA) and each of the LIA+CDA and LIA+TSI conditions, and found that the average number of sustained successful gesture performances was significantly higher in the LIA+CDA and LIA+TSI conditions, than in the LIA-only condition ($p < 0.05$, $p < 0.01$).

6 Discussion

First of all, the system successfully encouraged many passers-by to participate in the game and collected data from various users just by deploying the system in a public place (e.g., the game was played over 100 times in the first week of the deployment), thereby suggesting the potential for low-cost and automatic gesture data collection.

In terms of user engagement, Fig. 5(a) and (e) show that LIA+CDA and LIA+TSI had significant effects on Correctness and Sustained. Furthermore, In Fig. 5(c) and (d), we see that LIA+CDA has a significantly higher percentage of Completions and Score (LIA+TSI was also trending in that direction, but the differences were not significant). These suggest that incorporating CDA and TSI resulted in more correct and sustained gesture input.

The reason for LIA+TSI's effectiveness in Fig. 5(a) and (e) could be explained from the perspective of feedback, in which giving people positive feedback on a task increases people's intrinsic motivation to do it [12], and the LIA+TSI certainly provided such feedback. In CDA, in addition to providing detailed and explicit information on what to do, the timeliness of the information seems to have enabled users to associate the information with the required game actions, resulting in more sustained gesture input. Similarly, in TSI, the implicit yet timely feedback helped users understand how best to play the game. While LIA did provide such information, it provided the information at a less contextually relevant time (before the actual gameplay started) therefore this could have affected the association of the information with its use during the game. These suggest that providing timely information, whether it is explicit or implicit, is important to enhance user engagement in rhythmic game-based, in-the-wild gesture collection.

7 Conclusions and Future Work

We presented a gesture data collection system using a rhythm game combined with a Walk-Up-and-Use Display, to create a gesture corpus obtained from diverse users at low cost. The in-the-wild study showed that the system successfully encouraged many passers-by to engage in the game and collected data from various users just by deploying the system in a public place, thereby suggesting the potential of low-cost and

automatic gesture data collection. We also examined the effects of different gesture guidance conditions on user engagement in Walk-Up-and-Use displays and found that the guidance conditions with Contextualized Demonstration Animation and Tracking State Indicator resulted in sustained user engagement.

While this represents early work, the current results offer interesting insights into automatic gesture data collection, and pave the way for several future directions, such as further investigations of the venue, game design (e.g., sound effects, game genres), and conducting laboratory studies. Different gestures, applications and guidance models should also be investigated in order to generalize our proposed concept and results. Furthermore, we plan to analyze and evaluate the obtained gesture dataset by using it to train gesture recognition systems.

Acknowledgment. This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative. This work is also supported by JSPS Bilateral Joint Research Project, and Cooperative Research Project Program of the Research Institute of Electrical Communication, Tohoku University.

References

1. Chen, F.-S., Fu, C.-M., Huang, C.-L.: Hand gesture recognition using a real-time tracking method and hidden Markov models. *Image Vis. Comput.* **21**, 745–758 (2003)
2. Hoffman, M., Varcholik, P., LaViola, J.J.: Breaking the status quo: improving 3D gesture recognition with spatially convenient input devices. In: *Virtual Reality Proceedings IEEE*, pp. 59–66 (2010)
3. Luzhnica, G., Simon, J., Lex, E., Pammer, V.: A sliding window approach to natural hand gesture recognition using a custom data glove. In: *2016 Proceedings of the IEEE Symposium 3D User Interfaces*, pp. 81–90 (2016)
4. Rovelto, G., Degraen, D., Vanacken, D., Luyten, K., Coninx, K.: Gestu-Wan - an intelligible mid-air gesture guidance system for walk-up-and-use displays. In: Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M. (eds.) *INTERACT 2015*. LNCS, vol. 9297, pp. 368–386. Springer, Cham (2015). doi:10.1007/978-3-319-22668-2_28
5. Marshall, P., Morris, R., Rogers, Y., Kreitmayer, S., Davies, M.: Rethinking “multi-user” - an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In: *Proceedings of Conference Human Factors in Computer System*, pp. 3033–3042 (2011)
6. Konami: DanceDanceRevolution. <http://www.konami.jp/bemani/ddr/jp/>
7. Delamare, W., Coutrix, C., Nigay, L.: Designing guiding systems for gesture-based interaction. In: *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computer System - EICS 2015*, pp. 44–53 (2015)
8. Ackad, C., Clayphan, A., Tomitsch, M., Kay, J.: An in-the-wild study of learning mid-air gestures to browse hierarchical information at a large interactive public display. In: *Proceedings of the International Joint Conference on Pervasive Ubiquitous Computer, International Symposium Wearable Computer*, pp. 1227–1238 (2015)
9. Delamare, W., Janssoone, T., Coutrix, C.C., Nigay, L.: Designing 3D gesture guidance: visual feedback and feedforward design options. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI 2016*, pp. 152–159 (2016)

10. Hinrichs, U., Carpendale, S.: Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits. In: Annual Conference on Human Factors Computer System, pp. 3023–3032 (2011)
11. Walter, R., Bailly, G., Müller, J.: StrikeAPose: revealing mid-air gestures on public displays. In: SIGCHI Conference on Human Factors Computer System, pp. 841–850 (2013)
12. Deci, E.E.L.: Effects of externally mediated rewards on intrinsic motivation. *J. Pers. Soc. Psychol.* **18**, 105–115 (1971)