JUSTIN: An Audience Participation Game With A Purpose for Collecting Descriptions for Artwork Images

Ngoc Cuong Nguyen
Graduate School of Information Science and Engineering
Ritsumeikan University
Kusatsu, Shiga, Japan
gr0398pi@ed.ritsumei.ac.jp

Pujana Paliyawan
Research Organization of Science and Technology
Ritsumeikan University
Kusatsu, Shiga, Japan
pujana.p@gmail.com

Ruck Thawonmas
College of Information Science and Engineering
Ritsumeikan University
Kusatsu, Shiga, Japan
ruck@is.ritsumei.ac.jp

Hai V. Pham
School of Information and Communiation Technology
Hanoi University of Science And Technology
Hanoi, Vietnam
haipv@soict.hust.edu.vn

***Abstract*—This paper presents JUSTIN standing for Japanese Ukiyo-e Streaming That Improves Narrative, a game system that is designed for collecting descriptive data for artwork images (Japanese ukiyo-e in this study). The proposed game is the first "Audience Participation Game With a Purpose (APGWAP)," which combines two existing concepts: Audience Participation Game and Game With A Purpose. The APGWAP concept is aimed at addressing problems that computers can hardly solve by utilizing human participation on a live streaming platform. The descriptive data of ukiyo-e images can be beneficial in humanity research and in promoting cultural heritage. However, such data are currently insufficient, sparse or poor in quality. JUSTIN proves to be a powerful tool in exploiting crowd-sourcing for addressing such problems through solid evidence provided from the results of conducted experiments in this work in terms of the quality of collected descriptions as well as player experience.**

***Index Terms*—GWAP, APG, Ukiyo-e, Crowd-sourcing, Game Live Streaming, Game and Interactive Media**

I. Introduction

Image information, such as keywords, titles, and descriptions, has proven to be useful in improving the performance of machine learning systems [1] and the accessibility of visually impaired people [2]. For the former, the text-based metadata information of images can be used to augment training data, while for the latter, such information allows visually impaired people to grasp the content of an image through its description and text-to-speech technologies. However, acquisition of descriptive data for artwork images, which we target in this work, by human experts is time-consuming and costly. As a result, there exist only a few small-scale data sets [3], [4]. Methods in existing studies on caption generation for arbitrary images [5] [6] [7] do not perform well for artwork images unless they are enhanced using a sufficient number of targeted image data and their descriptive data, which are however not available yet.

A concept that can be applied to collect descriptions for artwork images is Game With a Purpose (GWAP), which uses games for humans to address problems that computers cannot solve. GWAPs have been successful in many different domains varying from computer vision, natural language processing to web accessibility. In this work, this concept is implemented on a live streaming platform to

- exploit the chat function of the platform for social collaboration of audience participants (players),
- focus more on developing the game mechanisms instead of building our own multi client server application, which is costly in terms of resources and time, and
- approach a huge group of active audiences on the platform (Twitch in this study).

With the rise of live streaming platforms such as Twitch and YouTube, Audience Participation Games (APGs) have grown in popularity. An APG is a game live streaming that allows audiences to not only watch but also participate in gameplay in part. Although the potential of both APG and GWAP concepts for descriptive data collection of artwork images (ukiyo-e in our study) was recently described in two short papers [8] [9], by the authors, the main contributions of this current work are as follows:

- A new concept called Audience Participation Game With a Purpose (APGWAP),
- A complete design of our APGWAP for collection of ukiyo-e descriptions, which can also be of use as a reference for other purposes, and
- Results as well as discussions on description quality and player experience from conducted experiments about the

978-1-7281-4533-4/20/$31.00 ©2020 IEEE

proposed APGWAP.

II. Related Work

A. Games With A Purpose

GWAPs, the first one being proposed by von Ahn et al. [10], have been successfully used in several applications such as data annotation in natural language processing (NLP) [11] [12] and image labeling in computer vision and web accessibility. The first implementation of GWAPs was the ESP game [13], with the purpose of collecting proper labels for images on the Web, and the game was shown to be successful in terms of player enjoyment and the number of valuable labels obtained, two main goals of any GWAP. However, resulting labels from the ESP are generic and not specific enough to differentiate similar images, as was pointed out by Steinmayr et al. [14]. As a solution, they proposed Karido to collect more specific labels for artwork images. Besides these keyword-collection games, the Phetch game [15] aimed to collect descriptions for arbitrary images for coping with the need to attach descriptive sentences to images on the web, which is beneficial for web accessibility. In all of these GWAPs, however, social interactions among players such as chatting can not be done in the games.

B. Audience Participation Games

APGs are games that empower audiences on live streaming platforms by allowing them to control gameplay indirectly or directly through their messages sent in the chat room area [16]. The most well-known example of APGs is Twitch Plays Pokmon [17], which allows players to directly control gameplay. Another example of APGs is Choice Chamber, a real-time, procedurally generated game where participant audiences constantly send commands to evolve the game in real time, e.g. choosing enemies, changing the game rules, while the streamer or player controls the main character in the game. From these examples, it can be clearly seen that APGs have blurred the line between audiences and players [16]. However, as pointed out by our survey work [9], the purpose of existing APGs was just for fun or to promote social interactions between audiences and streamers or among audiences themselves, not for obtaining valuable data as a side effect through audience participation in gameplay.

III. Proposed Concept and Game

A. Audience Participation Game With a Purpose

Combining both APG and GWAP concepts as APGWAP can not only enhance social interactions in GWAPs but also conduct valuable tasks through playing APGs. The APGWAP concept is straight forward. Namely, an APGWAP is a GWAP on a live-streaming platform where players take part directly or indirectly in gameplay. Participation in such gameplay can be done in many ways, depending on platforms, but in the proposed APGWAP, described in the following subsection, this is done via chat messages. Since chat messages, in general, can be seen by other audiences on a live streaming platform, unlike traditional GWAPS where a player of interest cannot see other players' inputs while making his/her input decision, care must be taken in the design of an APGWAP to ensure the quality of task results.

B. Japanese Ukiyo-e Streaming That Improves Narrative

Based on the APGWAP concept, we propose a novel game named JUSTIN which stands for Japanese Ukiyo-e Streaming That Improves Narrative. The targeted live-streaming platform is Twitch.tv. The game is designed for collecting good quality descriptions for ukiyo-e artwork images by executing chat commands sent by audiences through the chat room on a specific channel in Twitch.tv.

In JUSTIN, players directly control the game by sending messages in predefined formats. The predefined formats are introduced to distinguish with other audiences typical chat messages. A set of a certain number of ukiyo-e images are randomly displayed in each round consisting of three consecutive sessions: $describing$, $voting$ and $result$. Each image is given an image id, imageID. There are two roles in each round that an audience who wants to join the game, or to become a player, can choose: describer and voter. Describers and voters play the game during the $describing$ session and the $voting$ session, respectively.

Ideally, there should be at least one player and four players in the $describing$ and $voting$ sessions, respectively. When necessary, two descriptions will be automatically generated for each image by Pythia [6] [7]; more details on this are given in III-B-2. At least four votes are required to activate a proposed penalty mechanism described in III-B-3. Note that a player cannot take both roles in a given round. The $describing$ session, which is for collecting descriptions for images, and the $voting$ session, which is for contributing in assuring the descriptions quality, are the most important parts of the game. To prevent user distraction, both of these sessions as well as the $result$ session are conducted during live streaming on the same platform.

Each of the three sessions in a round is described as follows:

- In the $describing$ session, a group of describers can describe each of the displayed images by typing a message in a predefined format, "imageID:description." A describer can only send one description per image, and only a certain number of descriptions from describers for each image are accepted on the first-come-first-serve basis. After this session ends, the game turns to the $voting$ session, and audiences who are not the describers in the $describing$ session can vote for the best description as voters.
- In the $voting$ session, the system displays each image's list of descriptions, each associated with an id ("descriptionID"), under the image. Voters must vote for the best description they subjectively consider suitable for the corresponding image in a predefined format, "#descriptionID," in a certain period of time. Each voter can only vote for at most one description in a round.
- In the $result$ session, which is the last one in a given round, the game shows the results of the current round that consist of the wining description for each image,

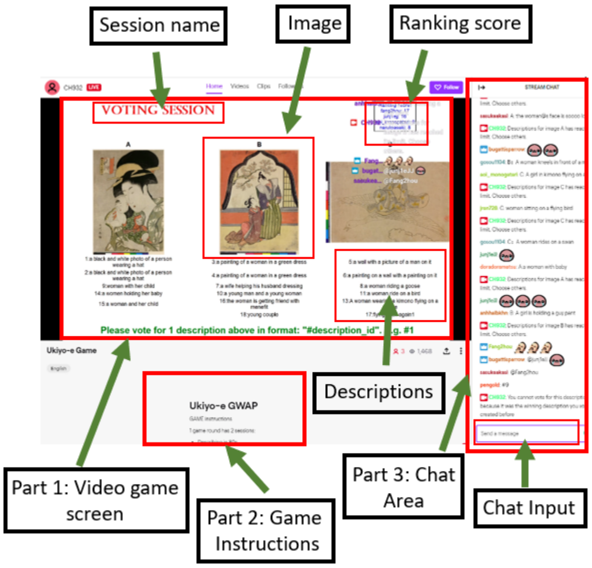

Fig. 1. JUSTIN on Twitch

rewarded players (those obtaining a plus score), and penalized players (those obtaining a minus score). Note that the winning description of an image is the one that receives the highest number of votes among the image's descriptions in the current round.

Since describers can describe all the images while voters can only vote for at most one image, describers have a higher chance to be rewarded more points than voters based on the reward mechanism in III-B-3. Due to this setting, it is expected that creative players would take the describing role. However, this role, of describing images, takes more time and efforts than the voting one. Due to this balance, there should be enough number of players for both roles in each round.

The information is given later on the number of displayed images (III-B-1), the number of accepted descriptions from describers for each image (III-B-1), and the length of each session in a round (III-B-4).

1) GUI Design: Our GUI design consists of three main parts. The first part is the video game screen that shows:

- the session name,
- three ukiyo-e images with temporary ids A, B and C,
- notifications for prompting audiences to describe or vote,
- three lists of up to six descriptions each, one list per image, in the $voting$ session and the results in the $result$ session, and
- the ranking scores of the top four players,

where space availability was considered in the determination of the number of displayed images, the number of descriptions per image, and the number of players whose ranking scores are displayed. The second part is the chat area for showing chat messages, general ones or commands, from audiences who can type and send them at the chat input box in the bottom. The third part is the panel showing game instructions, the links to a demo video clip, a survey page, and a ranking page that shows all the players who have played the game until the current round and their accumulated scores sorted in decreasing order. The layout of each part is shown in Fig. 1.

2) Implementation: The game was implemented in Python and is currently live streaming on Twitch's ch932 channel[1] using StreamLabs OBS [18]. For promoting interactions with audiences, we created a chatbot for this channel with Twitch API. It chats information messages, such as "You cannot vote because you already described.", and moderator messages, such as "You can see the ranking at (url address)."

In the $voting$ session, due to the limit in available space on the game screen, up to six descriptions that are shown in the list for each image consist of:

- two descriptions automatically generated by Pythia [6] [7] when an image of interest is displayed for the first time or when it is repeated but there is still no previous winning description for it,
- all of its previous winning descriptions that meet a certain condition, and
- if there are places left, the earliest descriptions sent by describers in the current round.

Note that automatically generated descriptions are for providing more choices to voters. In addition, the aforementioned condition considers descriptions' score, determined according to reward/penalty mechanisms in III-B-3, and is described in III-B-4.

3) Reward and Penalty Mechanisms: The system uses the reward/penalty mechanisms for giving plus/minus scores to players and descriptions. The reward mechanism is introduced to encourage players to play well, and the penalty mechanism is introduced to ensure the quality of winning descriptions. An image and its winning descriptions are shown in several rounds, during which winning descriptions are accumulated, for collecting good quality descriptions. In each round, the reward mechanism is applied to the round's winning players and winning descriptions, but the penalty mechanism is only applied to the previous winning descriptions of a repeated image that lose in the current round and to their respective describers and previous voters. The details of both mechanisms are given in the following.

In the reward mechanism, player p_i who is the describer or each voter for winning description d_{ij} of image j will accumulate positive points as follows:

$$Score(p_i) \mathrel{+}= \frac{1}{M}\sum_{j=1}^{M}\frac{v_{ij}}{v_j}, \quad (1)$$

where M is the number of images shown in the current round (three in this study), v_{ij} is the number of votes on d_{ij} in the current round, v_j is the number of votes on all the descriptions of image j in the current round. In addition, winning description d_{ij} is rewarded as follows:

$$Score(d_{ij}) \mathrel{+}= \frac{v_{ij}}{v_j}. \quad (2)$$

[1]https://www.twitch.tv/ch932

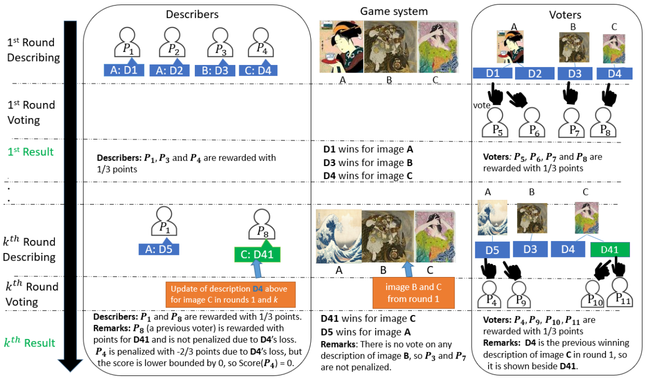

Fig. 2. JUSTIN gameplay example

Note that both $Score(p_i)$ and $Score(d_{ij})$ have an initial value of 0, and image j can be either an image displayed in the current round for the first time or a repeated one. For a repeated image, one of its previous winning descriptions will also be rewarded according to Eqn. (2) if it wins again in the current round.

The penalty mechanism will be applied in a round where there are more than M votes, based on our rule of thumb to ensure the reliability of the mechanism. Let d^*_{kj} denote a previous winning description of image j which was described or voted for by players p_k, representing not only the describer but all of the previous voters of the description. In a round of more than M votes where the image of d^*_{kj} is repeated, if d^*_{kj} loses, the system will go through the following two steps.

1) If there is a player among p_k who creates an another description for image j and the new description wins in this round, a penalty will not be applied to the player, who will instead be rewarded by the reward mechanism for his/her new winning description. This is to encourage players to continue improving the quality of descriptions.
2) The score of description d^*_{kj} and all of its p_k, except for the player in Step 1, will be decreased as follows:

$$Score(d^*_{kj}) \mathrel{-}= u_j(1 - \frac{v_{kj}}{v_{ij}}) \;\; with\; i \neq k \tag{3}$$

$$Score(p_k) \mathrel{-}= u_j(1 - \frac{v_{kj}}{v_{ij}}) \;\; with\; i \neq k \tag{4}$$

Here, $u_j = \frac{v_j}{v}$ represents how popular or preferable image j is among voters (compared to the other images in the current round), v_j is the number of votes on all the descriptions of image j in the current round, v is the number of voters in the current round, v_{kj} is the number of votes on d^*_{kj} in the current round, and v_{ij} is the number of votes on the winning description of image j (d_{ij}) in the current round. Although not shown in Eqn. (4), the score of a player is lower bounded by zero.

Figure 2 shows an illustrative example of JUSTIN's gameplay and its reward and penalty mechanisms.

4) Description Quality Assurance: The penalty on players, described above, is aimed at preventing describers and voters from creating poor-quality descriptions and not seriously making votes, respectively. In addition, to ensure the quality of collected descriptions, we introduce a lower-bound threshold, L, and an upper-bound threshold, U. All the previous winning descriptions of an image with the score less than L will be removed from the game and thereby not shown in the image's list of descriptions during the voting session. This is aimed at removing poor-quality descriptions that somehow won in previous rounds. On the contrary, each previous winning description with the score of U or above will be selected as an official description for its image, and once this is done, their images will no longer be shown in the game. In this study, we empirically set $L = 0.3$ and $U = 2.0$.

The length of the $describing$, $voting$, and $result$ sessions is set, based on feedback from our pilot studies, to 80 s, 30 s, and 10 s, respectively. The length of the $describing$ and $result$ sessions must give enough room for audiences to come up with good descriptions and to grasp the results in a round, respectively. For the $voting$ session, if its length is too long, audiences can see trends in voting and might just follow earlier voters, so the $voting$ session's length must be carefully set. Note that due to network lag in streaming, each audience might experience slightly different session lengths, but solving this issue is beyond the scope of this work.

IV. Experiments and Results

Two types of experiments were conducted: the first one is to evaluate the quality of collected descriptions and the second one is to evaluate the experience of players. All the ukiyo-e images in use are from an online collection of the British Museum[2], which has 1748 images annotated with both descriptions and keywords. In the following, such descriptions and keywords are called expert descriptions and expert keywords, and descriptions collected from audiences through JUSTIN are called non-expert descriptions.

A. Description Quality Evaluation

Here, our objective is to compare non-expert descriptions and expert descriptions as well as expert keywords for 20 images randomly selected from the aforementioned British Museum images. The experiment was divided into three stages: describing, voting and evaluation; the first two stages simulate the $describing$ session and the $voting$ session of JUSTIN. Each stage is described in the following.

In the describing stage, there were 24 participants, all being graduate students in computer science. They were divided into four equal groups, and each group was randomly assigned five images from the selected 20 images. Each participant was asked to describe each of the five images assigned to their group.

In the voting stage, 19 participants, undergraduate students and graduate students in computer science, participated. All

[2]https://research.britishmuseum.org/research/collection_online/search.aspx?searchText=ukiyo-e

Fig. 3. Examples of ukiyo-e images and their expert keywords, expert description, and non-expert description

of them did not participate in the describing stage and were individually asked to select the best description for each image's six descriptions obtained in the describing stage. In particular, each participant was asked to select the description they considered most fit for a given image. For each image, the description that has the highest number of votes was used as the non-expert description of the image in the evaluation stage.

In the evaluation stage, there were 24 participants; most of them were participants in the describing stage plus some new graduate students. They were divided into two equal groups, each group being assigned 10 images that its participants had not previously seen. The participants in each group were individually asked to answer ten questions per condition for three conditions: keywords, non-expert, and expert. For each image in either group, a set of five images were formed consisting of the image itself and four other images, from the set of 1748 images, that were the top four images most similar to it, with respect to expert descriptions where doc2vec [19] was used to calculate image similarity. Each question targeting an image of interest asks a participant to select the image among the respective set of five images that they consider most representative by its keywords (keywords condition), non-expert description (non-expert condition), or expert description (expert condition).

For the non-expert condition, the participants were able to select the correct image, or the targeted image in each question, 92.9% of the time. This is much higher than the expert condition, 75%, and the keywords condition, 50.7%. This indicates that descriptions created by non-expert players, in our study graduate students in computer science, best represent their images to non-expert population, compared to not only expert keywords, whose information is limited, but also expert descriptions, albeit sometimes containing specific information of academic value such as location names. As a result, they can be used for conveying information in images to general people or being used as augmented data in training computer vision tasks. Figure 3 shows two images used in this experiment and their expert keywords, expert descriptions, and non-expert descriptions.

B. Player Experience Evaluation

To evaluate the experience of players playing JUSTIN, we conducted two experiments at two different locations. The first experiment was conducted using 25 students (3 undergraduate students and 22 master's students, out of which 5 females) in Ritsumeikan University, Japan; and the second one, with an improved version of JUSTIN based on feedback from the first experiment, used 26 1st-year undergraduate students in Bangkok University, Thailand. In both experiments, after going through a demo session about how to play the game, participants played JUSTIN on Twitch.tv for approximately 20 minutes. A new set of 12 ukiyo-e images from the British Museum collection was used for the first experiment to ensure that most of the images will be randomly displayed at least three times during the experiment, so we can evaluate the reward and penalty mechanisms as well as the improvement of descriptions' quality over time, while the whole set of 1748 images was used in the second one.

TABLE I
FEEDBACK AND IMPROVEMENT FOR EACH GUESS FACTOR

Factor	Feedback	Improvement
1	The rules are difficult to understand.	Make a demo video with subtitles.
3	None	None
4	It's boring for playing long time without music.	Add background music.
5	More images more interesting.	Add more images to the game.
7	I don't have enough time to read all the descriptions.	Increase the voting time from 15s to 30s.
8	None	None
9	More readable graphic interface.	Increase the font size of descriptions in the voting session.

Factors: (1) Usability/Playability, (3) Play Engrossment, (4) Enjoyment, (5) Creative Freedom, (7) Personal Gratification, (8) Social Connectivity, and (9) Visual Aesthetics

After the playing session, a questionnaire on player experience was conducted using an online survey site. The questionnaire is based on the GUESS questionnaire [20]. Because our game had no narrative and sound/music at the

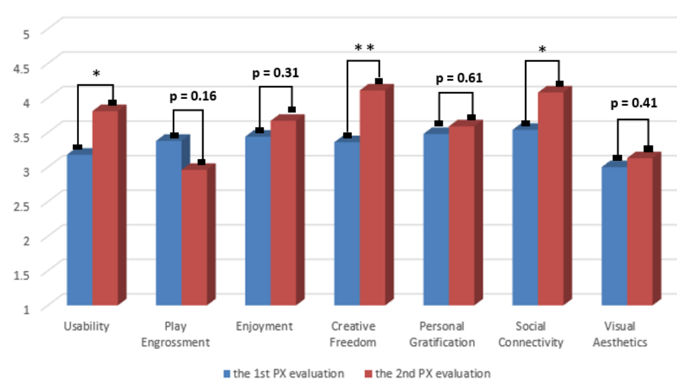

Fig. 4. Result of each satisfaction factor on JUSTIN game in the first and second control experiment with t-test

time of the first experiment, GUESS' factor 2 (Narratives) and 6 (Audio Aesthetic) were excluded. The remaining seven factors are Usability/Playability, Play Engrossment, Enjoyment, Creative Freedom, Personal Gratification, Social Connectivity, and Visual Aesthetics. For each of them, the two top GUESS questions were used in a 5 point Likert scale from 1 (Strongly disagreed) to 5 (Strongly agreed) and displayed to participants in random order.

The results from the first experiment and the second one are shown in the left (blue) bar and the right (red) bar, respectively, for each pair of bar charts in Fig. 4. Main feedback from participants in the first experiment and our improvements of JUSTIN for the second experiment are summarized in Table I. An unpaired t-test was conducted for each of the GUESS factors in use, and the tests reveal that Usability, Creative Freedom, and Social Connectivity factors increase with a statistical significance. It is interesting to see this trend for Social Connectivity since no direct improvement was conducted for it; We conjecture that its increase is due to the effects from improvements of other factors. However, there is room for further improving Play Engrossment and Visual Aesthetics.

Next, we discuss the reward/penalty mechanisms and development of winning descriptions through game rounds. Figure 5 shows the development of winning descriptions and their scores for two images from the first experiment. For the image on the left, the "man and woman in European clothes" description won in two rounds and gained a score of 1.0 in each of the winning rounds. This description became the official description for the left image because of its score reaching $U = 2.0$. Hence, as mentioned earlier, after this point, the image would not be displayed again by the game system. It can be seen that this official description has a better quality than the other previous winning description, "a man and a woman," that has a final score of 0.66.

For the right image in Fig. 5, this example shows that the proposed reward/penalty mechanisms work as designed in helping improve the quality of descriptions by increasing the score of high-quality descriptions and decreasing the score of low-quality descriptions for an image every time it is shown. In this example, the image was displayed five times. The "A woman riding a goose" description won in the first two rounds, but then another description, "A woman wearing a kimono flying on a goose," kept winning in the next two rounds and obtained a total score of 1.0. In the last round, a new description, "Flying duck again1" somehow won but with a low score 0.4. If the game continued to be played, it is highly possible that this new description would not win again as it was not as good as "A woman wearing a kimono flying on a goose." For the other previous winning description, "A woman riding on goose," it kept losing and being penalized and finally had a score of 0.3. As a result, one can see that if the game is being played for a sufficient period of time, a good quality description can be collected for each targeted image.

Time flow	Winning descriptions	Score
1st	A man and a woman	1.0
2nd	Man and woman in European clothes	1.0
	A man and a woman	0.8↓
3rd	Man and woman in European clothes	2.0↑
	A man and a woman	0.66↓

Time flow	Winning descriptions	Score
1st	A woman riding a goose	0.4
2nd	A woman riding a goose	0.7↑
3rd	A woman wearing a kimono flying on a goose	1
	A woman riding a goose	0.5↓
4th	A woman wearing a kimono flying on a goose	1.3↑
	A woman riding a goose	0.4↓
5th	Flying duck again1	0.4
	A woman wearing a kimono flying on a goose	1.0↓
	A woman riding a goose	0.3↓

Fig. 5. Development of winning descriptions of an image through game rounds

V. Conclusions and Future Work

We presented a novel live-streaming game named JUSTIN that implements a proposed game design concept called Audience Participation Games With a Purpose (APGWAP), for collecting descriptive sentences for ukiyo-e images through chatting on a game live streaming platform. Descriptions of ukiyo-e images collected from JUSTIN can be directly used in other applications, such as those for improving accessibility for visually impaired people to artwork images. Our experiments offer evidence that APGWAPs can be effectively exploited for data collection in humanities research, with JUSTIN as the first successful example. Our future work includes supporting more languages, such as Japanese, Thai, Vietnamese so that non-English speaking players can play JUSTIN as well, adding functions to the chat bot to make it more interactive, such as cheering after players win to encourage them, and improving graphics.

Acknowledgement

This research was supported in part by Grant-in-Aid for Scientific Research (C), Number 19K12291, Japan Society for the Promotion of Science, Japan.

References

[1] Hoang, Cong Duy Vu, Trevor Cohn, and Gholamreza Haffari. "Incorporating side information into recurrent neural network language models." In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1250-1255. 2016.

[2] Petrie, Helen, Chandra Harrison, and Sundeep Dev. "Describing images on the web: a survey of current practice and prospects for the future." Proceedings of Human Computer Interaction International (HCII) 71 (2005).

[3] The British Museum, Ukiyo-e collection. https://research.britishmuseum.org/research/collection_online/search.aspx?searchText=ukiyo-e. Last accessed on Febuary 08, 2020.

[4] Ritsumeikan University, Ukiyo-e Portal Database. https://www.dh-jac.net/db/nishikie/search_portal.php?&lang=en. Last accessed on Febuary 08, 2020.

[5] You, Quanzeng, Hailin Jin, Zhaowen Wang, Chen Fang, and Jiebo Luo. "Image captioning with semantic attention." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4651-4659. 2016.

[6] Singh, Amanpreet, Vivek Natarajan, Meet Shah, Yu Jiang, Xinlei Chen, Dhruv Batra, Devi Parikh, and Marcus Rohrbach. “Towards vqa models that can read.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8317-8326. 2019.

[7] Singh, Amanpreet, Vivek Natarajan, Yu Jiang, Xinlei Chen, Meet Shah, Marcus Rohrbach, Dhruv Batra, and Devi Parikh. “Pythia-a platform for vision language research.” In SysML Workshop, NeurIPS 2019. 2018.

[8] Nguyen, Ngoc Cuong, Zhenao Wei, Pujana Paliyawan, Hai V. Pham, Ruck Thawonmas, and Tomohiro Harada. “Using GWAP to Generate Informative Descriptions for Artwork Images on a Live Streaming Platform.” In 2019 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), pp. 43-44. IEEE, 2019.

[9] Nguyen, Ngoc Cuong, Pujana Paliyawan, Ruck Thawonmas, Hai V. Pham, Harada Tomohiro, Keiko Suzuki, and Masaaki Kidachi, “Potentials of Games With a Purpose and Audience Participation Games for Descriptive Data Collection in Humanities Research.” In Japanese Association for Digital Humanities Conference 2019 (JADH2019) Osaka, Japan, pp. 104-106, Aug. 29-31, 2019.

[10] Von Ahn, Luis, and Laura Dabbish. “Designing games with a purpose.” Communications of the ACM 51, no. 8 (2008): 58-67.

[11] Dziedzic, Dagmara. “Use of the Free to Play model in games with a purpose: the RoboCorp game case study.” Bio-Algorithms and Med-Systems 12, no. 4 (2016): 187-197.

[12] Fort, Karn, Bruno Guillaume, and Hadrien Chastant. “Creating Zombilingo, a Game With A Purpose for dependency syntax annotation.” In Proceedings of the First International Workshop on Gamification for Information Retrieval, pp. 2-6. 2014.

[13] Von Ahn, Luis, and Laura Dabbish. “Labeling images with a computer game.” In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 319-326. ACM, 2004.

[14] Steinmayr, Bartholomus, Christoph Wieser, Fabian Kneil, and Fracois Bry. “Karido: A GWAP for telling artworks apart.” In 2011 16th International Conference on Computer Games (CGAMES), pp. 193-200. IEEE, 2011.

[15] Von Ahn, Luis, Shiry Ginosar, Mihir Kedia, Ruoran Liu, and Manuel Blum. ”Improving accessibility of the web with a computer game.” In Proceedings of the SIGCHI conference on Human Factors in computing systems, pp.79-82. ACM, 2006.

[16] Seering, Joseph, Saiph Savage, Michael Eagle, Joshua Churchin, Rachel Moeller, Jeffrey P. Bigham, and JessicaHammer. “Audience Participation Games: Blurring the Line Between Player and Spectator.” In Proceedings ofthe 2017 Conference on Designing Interactive Systems, pp. 429-440. ACM, 2017.

[17] Ramirez, Dennis, Jenny Saucerman, and Jeremy Dietmeier. “Twitch plays pokemon: a case study in big ggames.” In Proceedings of DiGRA, pp. 3-6. 2014.

[18] Streamlab OBS. https://streamlabs.com. Last accessed on March 5, 2020.

[19] Le, Quoc, and Tomas Mikolov. “Distributed representations of sentences and documents.” In International conference on machine learning, pp. 1188-1196. 2014.

[20] Phan, Mikki H., Joseph R. Keebler, and Barbara S. Chaparro. “The development and validation of the game user experience satisfaction scale (GUESS).” Human factors 58, no. 8 (2016): 1217-1247.