Crowdsourcing Image-Word Associations to Generate Affective Space Maps for Affective Words

Andreas SEEG[†1] Kohei FUJIMORI[†1] Masahiro TOYOURA[†1] Xiaoyang MAO[†1]

Abstract: There has been a growing demand to search for content that matches personal preferences, all while the amount of available media increases constantly. Such demand calls for the necessity to label media with affective words, which describes the emotional impression the images might evoke. On the other hand, Affective Space has been gaining more and more attentions for the quantitative representation of emotion. In this study, we propose to use a crowdsourcing approach to collect all affective words that correspond to a given image from IAPS, a database consisting of images with known Affective Space coordinate. An online game is designed to get users involved in associating affective words with images. As the result, we can not only label IAPS images with affective words, but also consequentially obtain the Affective Space coordinates for affective words through their associated images.

Keywords: Affective words, affective space, crowdsourcing, gamification

1. Introduction

There has been a growing demand to search for content that matches personal preferences, all while the amount of available media increases constantly. Such demand creates the necessity to label media with affective words, which describes the emotional impression the images might evoke[1,2]. Accordingly, Affective Space research[3] has been gaining more and more attention for the quantitative representation of emotion. Researchers have conducted various experiments to obtain the Affective Space coordinates for different media, ranging from images[4] to sounds[5] and video clips[6].

Affective words known to represent emotions[2] are collected in various databases. Approaches that use a simple text analysis are not able to give us an exact understanding of the words. Conversely, a consensus among native speakers using the same affective words seems to exist. Exploring this consensus between humans using a non-artificial way is our goal.

In this study, we try to crowdsource finding the affective coordinate for emotional terms. To do so, we collect affective words that correspond to given images from IAPS[3], a database consisting of images with a known Affective Space coordinate. A network game is designed where users associate affective words with images. As a result, we can not only label IAPS images with affective words, but also consequentially obtain the Affective Space coordinates for affective words through their associated images.

We decided against a paper based method for getting the required data. To begin with, the environment of a research lab could affect the mood of the volunteers and thus taint the acquired data. Additionally, accurate results require much data to be collected from many individuals, but the number of individuals would be limited by the amount of survey time available.

More volunteers will be willing to participate in a study if they do not have to invest time to go to a specific laboratory, as is the case with the web-based approach. This also keeps them in a familiar place and should increase the accuracy of the acquired data. Lastly, expanding our approach to get even more test subjects from outside universities seems to be doable without large modifications.

As an application of this research, better media search seems a likely candidate, especially considering the large amount of research put into finding the affective space for various media. As a concrete example, the task of finding fitting images or music for advertisement, where evoking specific emotions is seen as important, should be much easier if entering emotional terms yields media that evokes the entered feeling.

2. Online Game to Generate Affective Space Maps for Affective Words

Previous research used a program called ESP Game [7] to correctly assign labels to a predefined set of images. In this online game, two people playing on different computers have to enter a word that they think represents the image best. If the two players enter the same word, it is stored as a label candidate for the image. Should the word already be listed in the database, a counter is increased. Later, a threshold for this counter can be used to filter out labels that only a small portion of players deemed to be fitting.

As the players do not get any cues on what to enter, this matching usually results in the most prominent item (and thus a noun) being entered. Words that express a feeling or mood were usually not entered by players, which makes this exact approach not feasible for our research.

In order to find emotional labels for images, Moriwaki et al. proposed to show an image and present a list of 20 emotion words [8]. Again, two players selected the most appropriate word on both computers, and matching words were stored.

While this approach gave usable results, limiting the list to a fixed set of 20 words might force the player to select a word that only expresses something similar to what they are feeling. The optimal word for an image might not be presented at all. Furthermore, only the most prominent feeling is likely to be selected. If there are multiple fitting words because an image evokes multiple feelings to a different degree, the resulting data is not able to represent this.

†1 University of Yamanashi

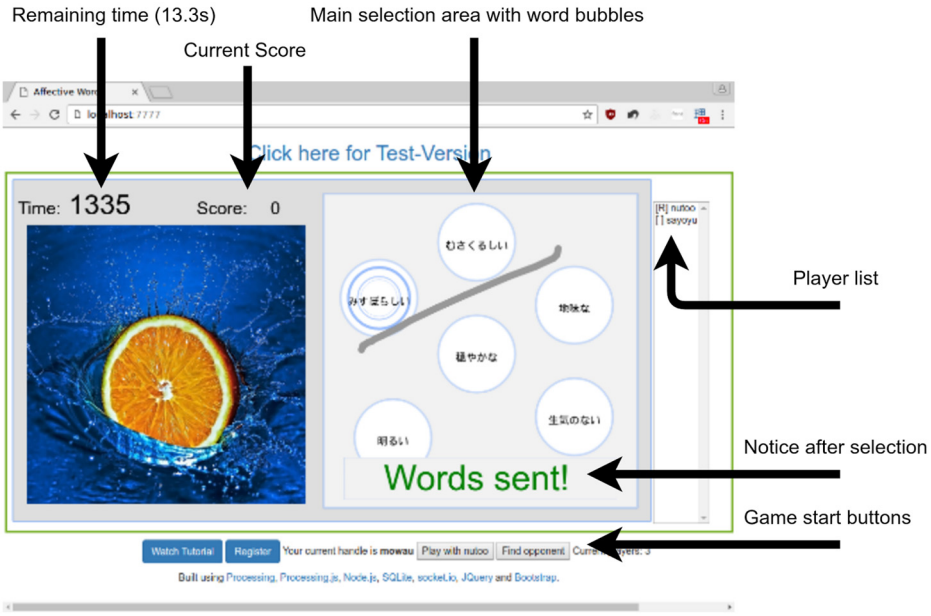


Figure 1. Screenshot of the network game for collecting affective words.

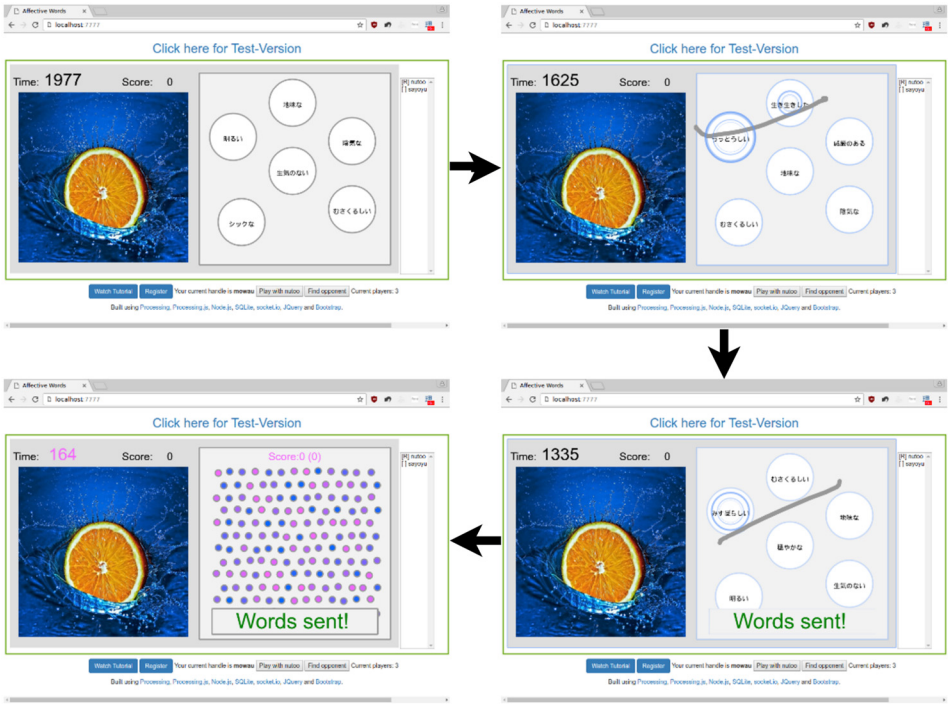


Figure 2. Example of a user input for the online game.

2.1 Game design

To address these problems, we propose a different method to collect fitting words for images as shown in Figure 1: We reduce the amount of selectable words to six, and allow for multiple selections. Furthermore, we ask the player to select the best fitting words prior to words that might not fit as well. If no words match the image, an empty selection can be made.

By showing the image to multiple people and selecting the six words from a larger pool of expressions of emotion, more finely grained data can be collected. On top of that, the method of choosing the words follows a later described system that should allow us to determine the most fitting words from the order of selections.

We hope that the reduction of the number of words will make the selections more intuitive, as feelings might be rationalized if the image is not freshly in one's mind. For similar reasons, a timer limits the amount of time for selecting the proper words to 20 seconds. This should allow a more fluid gameplay and cause less frustration for players that were faster at selecting all fitting words. After each round, the score is shown for 5 seconds, allowing the player to refocus on the upcoming task. The size of

the score also allows the player to determine how well he matched up with the other player: Extra points are awarded for selecting the words in the same order, especially for the first or second word, which should be the words matching the image most. Selecting or not selecting the same words as the other player also gives a minor score increase. Besides being direct feedback after each round, the score also encourages the player to choose the right words, and to anticipate what another person might feel when viewing the image.

To increase the dataset of image-word associations easily, the labeling program was implemented as a web page that can be accessed with any browser within a university network, or theoretically even over the internet (if the license for the image data set allows it). By getting many players to label images according to their emotional reaction, the subjectiveness of emotions and image impressions can be accounted for. By using a selection scheme and presentation that more closely resembles games, we hope to reduce the frustration or any other negative mood influences a paper based evaluation method might evoke.

2.2 Affective word likelihood

Every time the program chooses an image for the next round at random, six words need to be chosen and shown to the player. As a large number of emotional terms might not fit for a given image, many game rounds could consist of the players making no selection at all. To combat this problem, we do not select all words randomly. What we do instead is querying the database for words that have been selected more often than 65% of the time they were shown for the specific picture. Then, we select two random words from that dataset (if possible) and choose the words for the remaining slots from those who have been shown the least so far. The same word is only chosen once per round, even if the two groups we choose our words from might overlap.

2.3 Affective image selection

Each game, 20 of the total 64 affective images are selected at random and in no particular order. The pictures used come from the IAPS database and were chosen to cover a uniform space within the affective space. To rule out familiarization effects from seeing the same picture multiple times, no picture is shown twice during one game, and the players are advised to only play once.

One problem when using the IAPS is the limited availability of pictures in certain regions of the affective space as shown in Figure 3. To keep the pictures uniformly distributed, only images around the center were used. This could result in some extreme words never being selected because they match none of the less extreme images used in the game. It should not affect the correctness of data acquired, though, as players intuitively decide the fitting words and as they are free to not select any word at all.

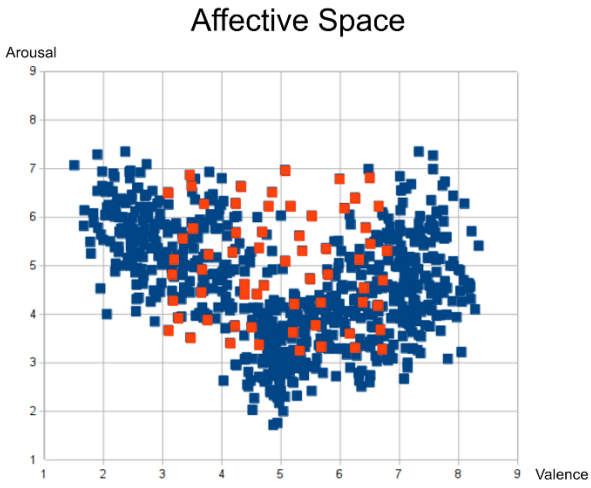


Figure 3. Biased distribution of IAPS Pictures.

2.4 System overview

Our web application follows a standard client-server architecture. The browser of the client connects to our HTTP-/webapp server to download the application written in HTML/JavaScript. All communication is done through the server, there is no direct connection between clients at any point. Furthermore, the design of the game makes communication between clients unnecessary, as the main logic resides in the server side application and the clients are only used for receiving, showing, selecting and sending information to and from the server.

To implement the game, multiple technologies were used. For the server, Node.js, which is a JavaScript based server environment with an included HTTP-Server, and socket.io, a project that simplifies communication to- and from web clients, were used. Because of that, the resulting code is mostly focused on writing the actual game and interface logic, and storing or extracting data from the database.

The database used is called SQLite, which is an easy to add, file based database. It simplifies the setup of the server, which now only consists of copying all project files and installing Node.js through a packet manager or installation file.

On the client side, standard HTML5 and JavaScript libraries were used. Of note is processing.js, a library that converts code in the processing programming language into JavaScript. It enabled us to port the core of an earlier version of the game, and simplified the implementation of graphical features like visual cues on word selection.

2.5 User interface design

While implementing and testing the program, some usage problems surfaced: For one, explaining the controls of the game, i.e. crossing out spheres with affective words in them if they are found to be matching, seemed to be problematic. As a solution, a test version of the game was provided, allowing players to learn how to use the controls without seeing the affective images used for the real game. On top of that, a video explaining the game can be viewed before playing the tutorial.

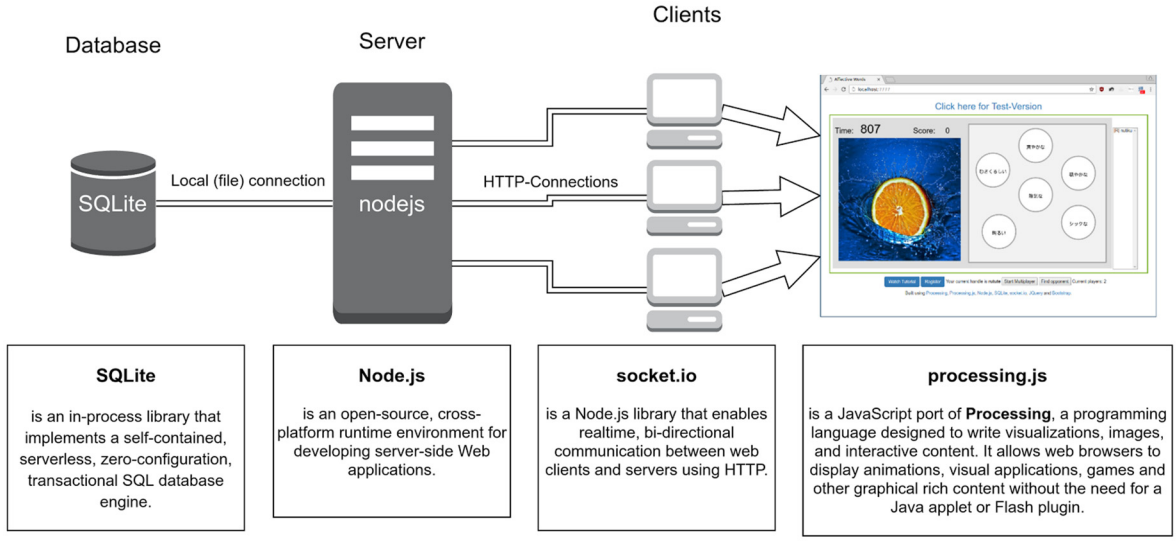


Figure 4. System overview.

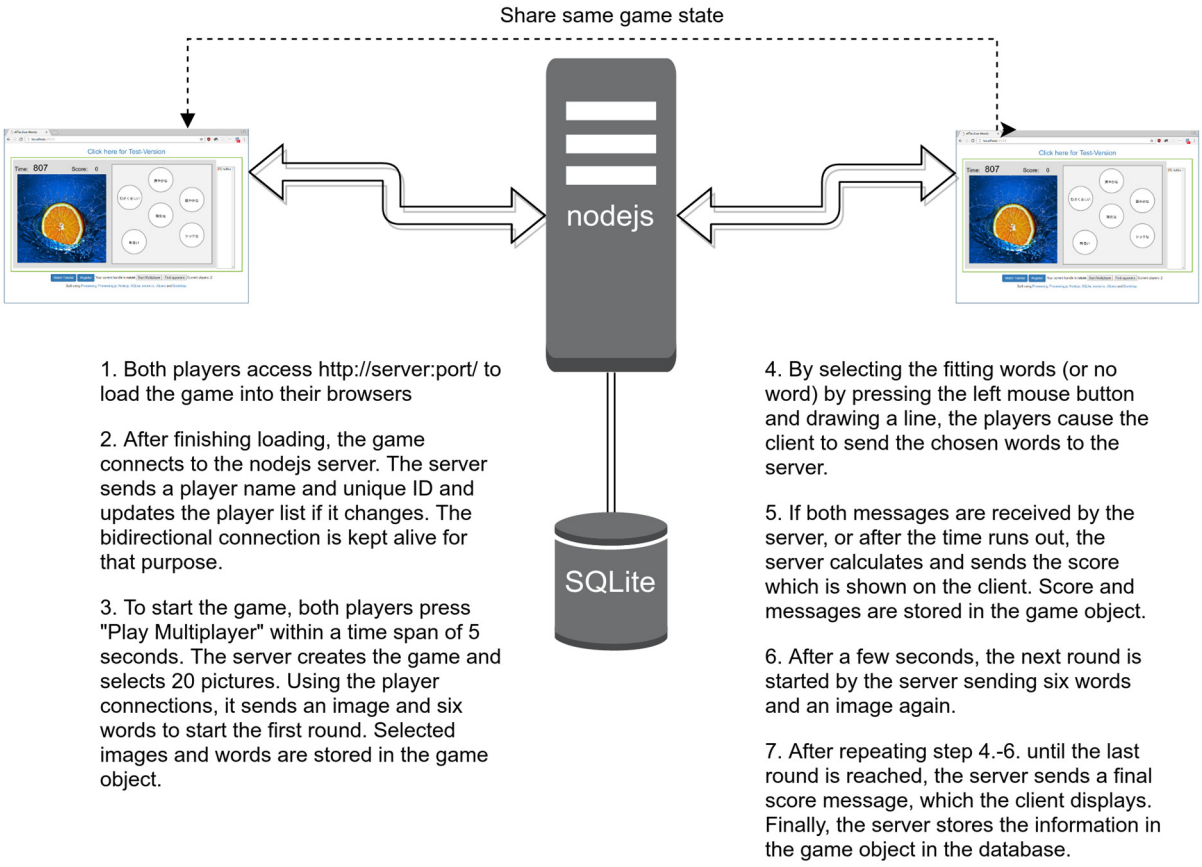


Figure 5. Communication between clients and server.

Another problem was finding the right interface for starting the game. We need two players to sit at a computer at the same time, and they should be paying attention to their browser window when the game is started. The first idea was to allow for selecting a player's name and to start the game by pressing a button. This method proved to be ok in a lab setting with guided use, but as a negative side effect further complicated the explanations. It also needed a final confirmation step so both players were confirmed to be looking at the screen when the game finally started.

Because of that we finally settled on a single button that requests an immediate game start. If another player clicks the same button within 5 seconds, the game will start on both websites without further confirmation. This solution simplified the process for new players, required less explanation, and makes sure that all players are actually focusing on the website when the game starts, resulting in a more robust game start for all

participants.

In previous tests, some players were unsure if their word selection had been properly recognized by the game. As a solution, we added a visual ripple effect that is triggered immediately around the word when the player has made a selection.

2.6 Network communication

The game communication is completely centralized: The server sends out messages to the clients, which then reply with an answer. There is no communication between the clients, though if needed, the server could easily relay messages between them.

To allow the server to communicate with the clients directly and at any time, the game client immediately opens a connection to the server upon finishing loading, and keeps this connection open for the remainder of the session.

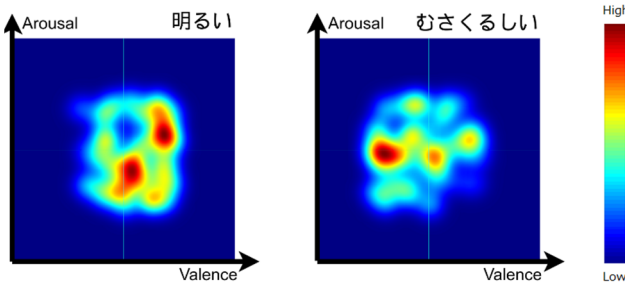
A complete rundown of all communication involved in a single game can be seen in Figure 5.

3. Results and Discussion

So far, 32 players have played 16 games. The game data has been combined into three tables: The first one lists how often a certain word-image-combination has been shown to players, while the second one lists the number of times a word was selected by a player when a specific image was shown. For the third table, the ratio of the data of the first two tables is calculated to tell how often a word has been selected for a given image if it was shown.

There is not enough data to accurately determine arousal or valence for certain emotional words. For some words, some images have not been shown at all. Some image-word combinations have been seen 8 to 12 times, though.

Using the current data, we created a valence-arousal map for each word. This was done by using the known affective space coordinates of the images for a word if people playing the game selected the word. If the likelihood of selection was high, this is denoted by a red spot on the map. A non-existing or seldom selection is denoted by a dark blue or close to dark blue color.



(a) Map for “明るい” (b) Map for “むさくるしい”

Figure 6. Examples of generated maps.

Some of the gaps in the image could be the result of not showing pictures correspondent to that region of the affective space. Likewise, hot spots (red to brown) might be relativized after more players have seen the same word-image combination. Despite these issues, the general selection trend seems fitting for the words we chose.

As the chosen 64 pictures have valence and arousal values between 3 and 7, the resulting maps for words lie in the same part of the affective space. Two potential words that might not get any coordinates because of this are "みすぼらしい" (seedy; shabby) and "うっとうしい" (annoying) as only a minority selected those words for any of the 64 pictures at all. This might suggest that our approach does not easily create false positives, though.

Some current problems limit the usefulness of our game. For one, the IAPS license forbids us from making the game accessible freely over the internet to protect the image files from being widely known and thus useless for further affective research. Another problem is the current requirement for a minimum of two players to play it at the same time. Even if multiple people were asked to play the game, only one player might access the game through their browser during a certain timeframe. As the game cannot start in this case, no data could be acquired, and the person was unlikely to try again later.

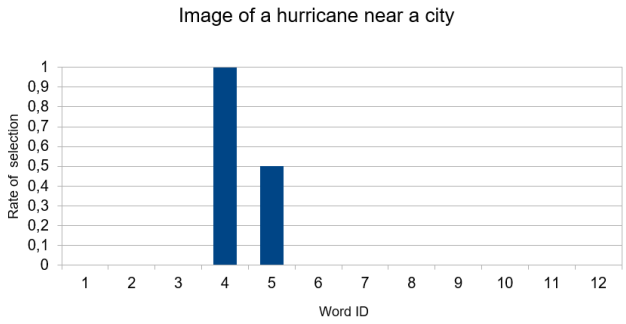


Figure 7. Inkina(dismal) and uttoushii(annoying) were the only fitting words for this image.

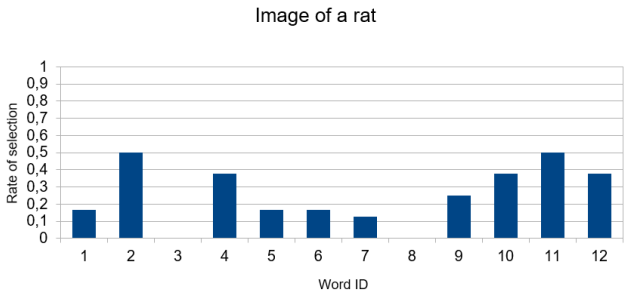


Figure 8. No accurate words in list of 12 words.

4. Conclusion

In this study, we propose to use a crowdsourcing approach to collect all affective words that correspond to a given image from IAPS. An online game is designed to get users involved in associating affective words with images. The resulting data looks promising, but as of yet cannot show the correctness of the approach.

Despite work on making the game more accessible, more interface improvements should be implemented to make it easier and more interesting for the player to enter data. The biggest problem, namely the requirement for two players, should be eliminated by providing game modes that allow for only one

player.

To increase the number of potential players and to allow for outsourcing the word selection process to people on the internet (e.g. crowdsourcing), a different image data set with a more permissive license would provide most useful. Some image datasets exist that seem to be available for this purpose, as long as the game is not made completely public. Future work should consider and implement them.

At the moment, the game can only be played by people who know Japanese quite well. A version of the game that uses emotion words from other languages can easily be implemented, and might allow research that compares the similarity of Japanese words and their translation candidate. Considering the implementation of this game as a web-based game, acquiring data from players of different nationalities should be easier than with traditional approaches.

Preliminary work for some of these features has been done. A single player game mode that replays randomly chosen game rounds from earlier multiplayer games can already be used. To improve this, instead of using a random game round to compare the single player's selection, the consensus found in the database could be used to determine if the answer receives more or less points. This mode would allow for validation and improvement of the data that has been collected so far. It cannot be used for completely new images because of lack of previous game data, though.

4.1 Crowdsourcing

Other game mode ideas include the coupling of the image selection game with a completely different browser game. Playing the affective word game could provide bonuses to the main game if the selection is fitting.

As our game requires the player's full attention for a few seconds, experienced players could also see it as an additional challenge. This mechanism, generally known under the term “mini game”, is usually seen as fun by the player because it forces them to focus on a different problem than the main game presents, increasing the variety of the whole game. Implementing this would provide a much higher incentive to play our game and should increase the number of users.

Another game mode idea increasing incentive for the players focuses on the word selections themselves. By allowing players to replay the same image-word-combinations as their friends, we could provide a similarity rating for friend groups. Furthermore, ratings similar to self-evaluation and polling websites could be provided to tell the player more about himself or his friends. This mechanism would probably also encourage players to invite their friends to play the word association game, similar to how online personality tests are spread among friends.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Number JP25280037 and JP16K12459.

Reference

[1] P. Lewis, H. Critchley, P. Rotshtein, and R. Dolan, "Neural correlates of processing valence and arousal in affective words," Cerebral cortex, vol. 17, pp. 742-748, 2007.

[2] C. Strapparava and A. Valitutti, "WordNet Affect: an Affective Extension of WordNet," in LREC, pp. 1083-1086, 2004.

[3] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, “International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual”, Technical Report A-6, University of Florida, 2005.

[4] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," Psychophysiology, vol. 30, pp. 261-273, 1993.

[5] M. M. Bradley and P. J. Lang, "Affective reactions to acoustic stimuli," Psychophysiology, vol. 37, pp. 204-215, 2000.

[6] S. Zhang, Q. Tian, Q. Huang, W. Gao, and S. Li, "Utilizing affective analysis for efficient movie browsing," in 2009 16th IEEE International Conference on Image Processing (ICIP), 2009, pp. 1853-1856.

[7] L. V. Ahn, and L. Dabbish, “Labeling Image with a Computer Game”, Proceeding of ACM Conference on Human Factors in Computing Systems (CHI), pp. 319-326, 2004.

[8] K. Moriwaki, X. Mao, “Image Labeling by Affective Words with WEB Game”, Visual Computing and Graphics and CAD Symposium, Article 26, 2007. (in Japanese)