# Rhythmic Unicorn: Building a Music Suggestion Database Through a Computer Game

Duy Nguyen and David Schabelski
May 2018

*Abstract*— **Rhythmic Unicorn is a crowdsourcing application that builds up a music suggestion database through various gamified elements. Through the application, we aim to measure the effectiveness of gamification strategies when combined with crowdsourcing, and whether or not these strategies provide a good incentive for people to contribute to crowdsourcing projects in general. When people use the application, or in other words, play the game, they are also helping to build a database of related songs. This database can then be used by online music services like Google Play Music, Youtube Music or Spotify to help make better suggestions. In addition to implementing the application, we also conducted an experiment as well as analyzed the experiment results.**

## I. INTRODUCTION

During recent years, there are two phenomena that has become popular in the field of Human-Computer Interaction: Gamification and Crowdsourcing. These rising technological trend can be seen both in practice and in academia [4]. As of December 2015, almost 3000 crowdsourcing-related examples are listed at crowdsourcing.org, a leading crowdsourcing industry portal. A simple search on Google Scholar with the keywords "crowdsource gamification" also gives approximately 6000 results, as of May 2018.

Crowdsourcing means outsourcing data-gathering or problem-solving tasks to people on the internet rather than to employees or traditional suppliers [4]. The primary goals of crowdsourcing is either to reduce computing cost or to handle challenging tasks that requires human input on a large scale. However, since crowdsourcing relies on a large amount of human contributions, usually some kind of incentives are needed, like a small amount of money. As a result, crowdsourcing systems are becoming increasingly gamified, that is, organizations are making the act of contributing more like a gaming activity rather than working [6]. There are many types of applications that can make use of crowdsourcing, and one such type is the music recommender system.

Following the rapid expansion of digital music formats, the need of a good music suggesting system is becoming very important. A multitude of music recommender systems and algorithms have been developed. These systems either utilizes data taken from their own user base to suggest music to other users (Collaborative Filtering method) or analyzes the songs to find similar songs in terms of instrument/tempo or mood of the song (Content-based Modeling method) [5]. However, none has employed crowdsourcing as a source of data.

Along those lines, we implemented a music recommender system that uses crowdsourcing. By gathering users' contributions, the app builds up a database of inter-related song, which can then be queried to find a list of related song from a given song. The app incorporates many gamification elements to provide motivations and incentives for the users (or players) to contribute. The purpose of this study is to determine the effectiveness of gamification when used with crowdsourcing, by analyzing the user's impression of the app. We have conducted an experiment with volunteering students from Gustavus Adolphus College.

The organization of this paper is as follows: We begin with exploring several related projects on crowdsourcing and gamification. We then go into details of the app's implementation, with some discussion of the various gamified elements utilized in the app. In the following section, we describe the design of the experiment we conducted, and then followed by the result and analysis of the experiment itself. The paper ends with a conclusion of the topic and a discussion of the results.

## II. RELATED WORKS

There have already been a huge amount of other related works using similar ideas of gamified crowdsourcing, but they serve different purposes. These works all aim to increase the crowdsourcee's motivation and participation through gamification. Generally, it has been shown that gamification has had a positive effect. [3].

One example of a similar work is a mobile phone app by Altmeyer et al. [1]. The app has the capability of using OCR to capture and read receipts, turning physical receipts into digital ones. However, because the OCR technology is not always reliable in reading correct text, a crowdsourcing system was added, so other people can look at the results of the OCR, and then confirm whether if it was correct or not. To make this crowdsourcing process more effective, gamification elements were employed. Namely, a system of points, achievements and leaderboards.

Another example is an image labeler that asks the player to label the images with tags. It was designed in the form of a cooperative game: Two strangers are randomly matched with each other, then have to try to come up with different words to describe a given image, and for every same words, both players are awarded with score. Those same words are then considered the label of that image. The author claimed that "If the game is played as much as popular online games, we estimate that most images on the Web can be labeled in a few months." [2]

In the same spirit, we wanted to make a similar system: one that is capable of producing valuable output, while being fun and engaging at the same time: A game in which players will contribute towards building a music suggestion database.

## III. IMPLEMENTATION

The application is a webapp, with a dedicated server as the backend. The technologies used include jQuery for the front end, Node.js as the backend, and MongoDB as the database. The app also made use of third-party API from Spotify and Youtube for building its database and for previewing songs.

### *A. Game Design*

The game is a single player game. For each playthrough, the player will have to go through two stages, called "Survey Mode" and "Quiz Mode", respectively.

In Survey Mode (Fig 1), the player will be given a song, and then challenged to list out similar songs to the one given. "Similar songs" might mean similar in genre, musical style, or the feel of the song or "vibe", depending on the player. There is a time limit of 5 minutes in this mode. The song that appears as the challenge is called the "seed song", as opposed to other normal songs. There is a search box for the player to search for songs. Players have the ability to preview the songs in the search results before they click on "Add" to add the song to the list. When players are happy with the current list of selection, they can click on "next" to move forward to next stage.

In Quiz Mode (Fig 2), the player will be presented with 10 different pairs of song, and are asked to rate if that the songs in one pair are similar or related to each other. It is simply a list of Yes / No questions. This quiz mode was inspired by the short comings of the program created by L. von Ahn and L. Dabbish [2].

### *B. Database Visualization*

The database is organized like a graph, with two types of objects: Song object and Relation object. The song objects are equivalent to nodes in the graph, and the Relation objects are equivalent to the edges, as shown in Fig 3.

### *C. Rating the relations*

We use a simple way to rate the strength of the relations between songs. Each time the song is listed in Survey Mode, it counts as an upvote. Each time the relation is voted "Yes" in Quiz Mode, it also counts as an upvote. Consequently, the strength of the relation is simply the amount of upvote over the total amount of votes.

## IV. EXPERIMENT DESIGN

In order to test weather our application effectively creates an enjoyable environment for the users, we must put the application through a user study that consists of two parts. The first part is simply letting the participants play through the game itself, and the second part is a post-play survey.

Part one: In this part of the test, we explained the study and the application briefly to the participant to ensure there is as little of a learning curve as possible. The participants then go through the application one time to get a grasp on the fundamentals of it. To ensure there is no bias of a learning curve, participants are asked to play through the application two times before going on to part two.

Part two: This part of the test is the most important piece of the experiment as it will show us weather or not the application is enjoyable or not. Participants are presented with a short survey and are asked to fill it out. For the survey, we utilized the Pleasure Arousal Dominance Emotional State Model as developed by Mehrabian and Russell (1974)[8] as a way to measure the impression of the participants towards the app.

## V. EXPERIMENT

As stated above, our application is a new approach from current music recommendation services such as Spotify, YouTube or Pandora. With that said we were able to test the application independent of comparing to another service. We now present an experiment that tests the enjoyment levels of Rhythmic Unicorn to see if the application is worth pursuing further. We recruited 10 participants (all 18-25 age bracket, 1 female) from the Gustavus Adolphus College student body for the 20-minute experiment. To begin the experiment, we asked users to play the game, we also briefly explained the study we were conducting and the fundamentals of the game. The users then played through the game two times, asking questions of us when needed. The users then finished the game and were prompted to take a six-question survey which rated how the application made users feel. The survey was based off of the PAD Emotional State Model [8]. The PAD scale consisted of six questions that had polar opposite adjective pairs rated on a nine-point scale. There were two questions for each emotional state (pleasure, arousal, and dominance) in which we chose words that we believed to be appropriate and simple to understand for the experiment.

## VI. RESULTS

The main goal of this experiment is to test whether our application is enjoyable to the user. Due to the limited amount of time and resources available, we were unable to address all of the bias that came up while gathering data. All participants were in the same age bracket (18-25) and most were male (90 percent). We also noticed that our standard deviation was considerably high due to the low number of participants. We are interested in taking this experiment further in the future when we are not constrained on time. To achieve the goal of this experiment, we preformed the experiment in the way listed above.

When utilizing our modified PAD test (see Figure 3) it is important to note that a score of 4.5 or less translate to a non-enjoyable feeling by the user. It shows that the smallest average that we observed though the experiment was a 5.4, proving that, on average, users experienced enjoyment from the application. Also, we determined that the application was found to be enjoyable to at least 90 percent of users when using the PAD test. When we observe the PAD test in the same way as Baker et al (1992) [7], we can determine that

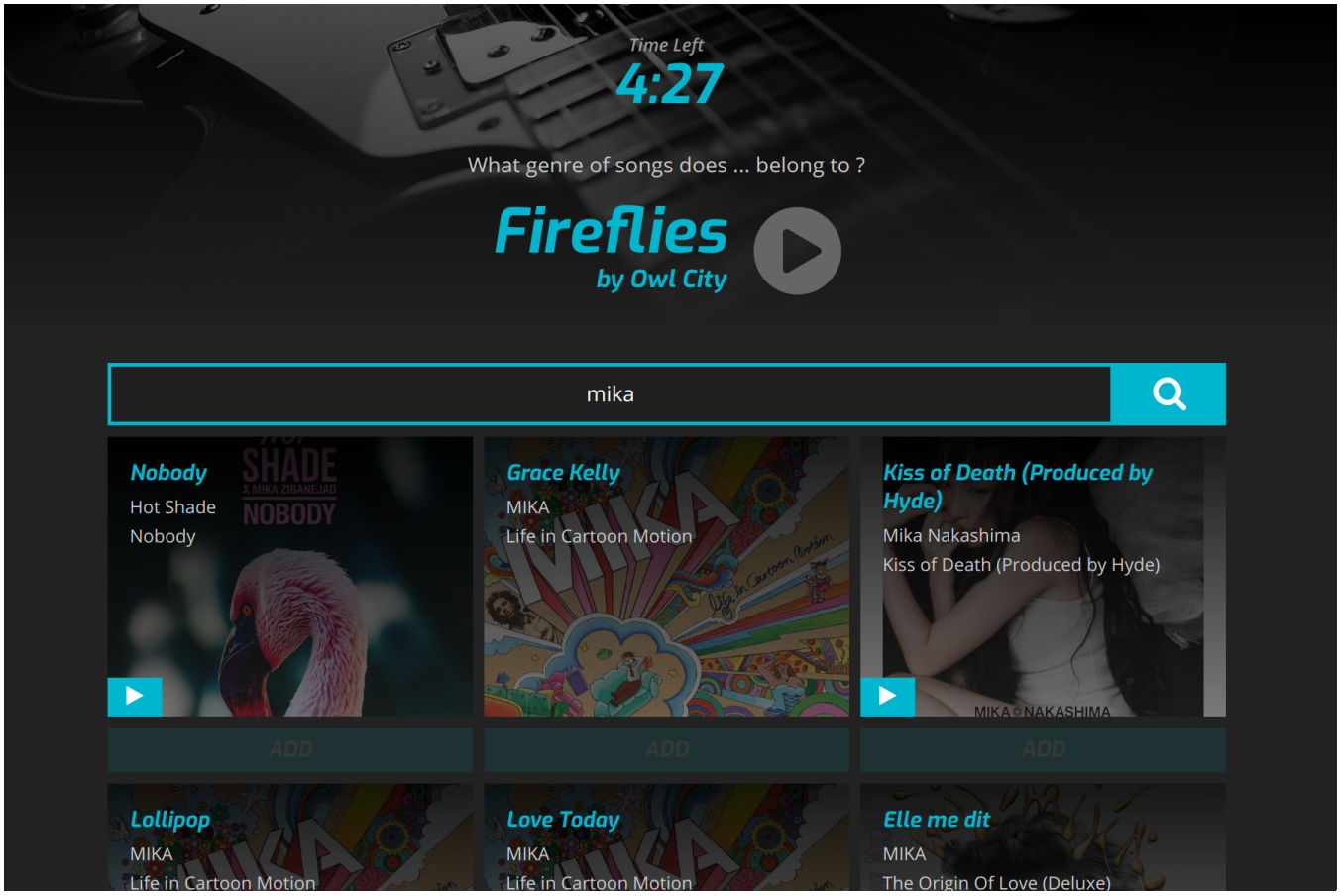


Fig. 1: Interface of Survey Mode.

our application is enjoyable to just under 95 percent of users. The data shows that users get pleasure the most out of our application (mean of 6.5), followed by arousal (5.9) and then dominance (5.75). Through these numbers we conclude that there is sufficient evidence that we should move forward with this application and test at a higher rate and without bias.

## VII. CONCLUSION

To conclude, based on our experiment, this application is enjoyable to users and should be pursued further in the future. With more time and resources we would be able to conduct a better experiment and get better results from it. We believe that this system is a stepping stone in the direction of giving users more input to what type of songs are related to one another.

## REFERENCES

[1] M. Altmeyer, P. Lessel and A. Krger, "Expense Control: A Gamified, Semi-Automated, Crowd-Based Approach For Receipt Capturing", Proceedings of the 21st International Conference on Intelligent User Interfaces - IUI '16, 2016.

[2] L. von Ahn and L. Dabbish, "Labeling images with a computer game", Proceedings of the 2004 conference on Human factors in computing systems - CHI '04, 2004.

[3] Morschheuser, B., Hamari, J., Koivisto, J. and Maedche, A. (2017). Gamified crowdsourcing: Conceptualization, literature review, and future agenda. International Journal of Human-Computer Studies, 106, pp.26-43.

[4] Howe, J., 2006. The rise of crowdsourcing. Wired 14, 14.

[5] Y. Song, S. Dixon, M. Pearce, 2012. A Survey of Music Recommendation Systems and Future Perspectives. 9th International Symposium on Computer Music Modelling and Retrieval - CMMR 2012.

[6] J. Hamari, J. Koivisto and H. Sarsa, "Does Gamification Work? – A Literature Review of Empirical Studies on Gamification", 2014 47th Hawaii International Conference on System Sciences, 2014.

[7] Baker, J., Levy, M., and Grewal, D. (1992). An experimental approach to making retail store environmental decisions. Journal of Retailing, 68(4), 445460.

[8] Mehrabian, A., and Russell, J. A. (1974). An approach to environmental psychology (pp. 216217). USA: The Massachusetts Institute of Technology.

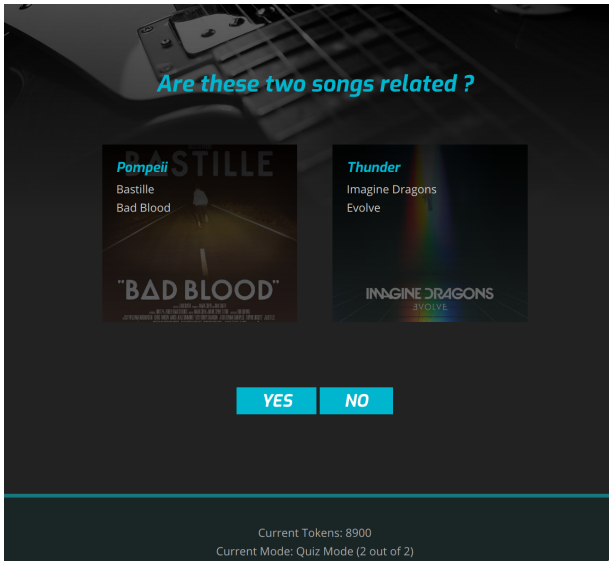


Fig. 2: Interface of Quiz Mode.

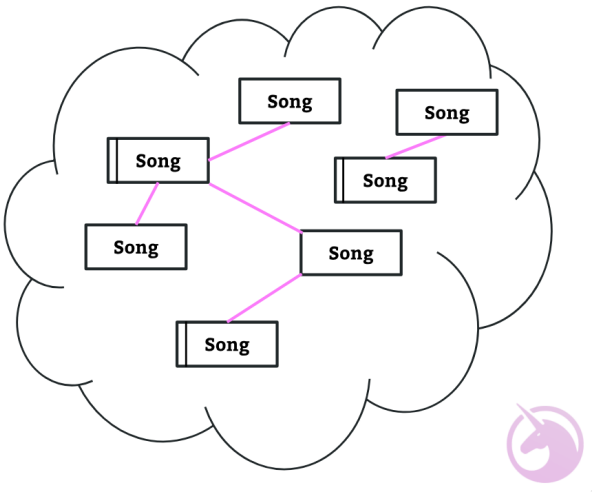


Fig. 3: Visualization of the database.

Post Survey for Rhythmic Unicorn

Please rate your experience with the application based on how it made you feel.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Annoyed | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Pleased |
| Unsatisfied | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Satisfied |
| Relaxed | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Stimulated |
| Calm | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Excited |
| Controlled | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Controlling |
| Influenced | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Influential |

Fig. 4: Example of the post-play survey.

| Question Asked | Mean | Standard Deviation | 95% CI for Mean | Percent Chance of it Being Negative |
|---|---|---|---|---|
| Annoyed-Pleased | 6.5 | 1.72 | 5.4-7.5 | 0.01% |
| Unsatisfied-Satisfied | 6.5 | 2.01 | 5.2-7.6 | 0.23% |
| Relaxed-Stimulated | 6.4 | 2.22 | 5.1-7.7 | 0.23% |
| Calm-Excited | 5.4 | 2.29 | 4.0-6.8 | 12.6% |
| Controlled-Controlling | 5.7 | 2.54 | 4.1-7.1 | 7.5% |
| Influenced-Influential | 5.8 | 3.46 | 3.7-7.7 | 12% |

Fig. 5: Shows the mean, standard deviation, 95 percent Confidence Interval, and the percent change of us getting a less than 4.5 on future tests. The categories are ordered by pairs, the first two being pleasure, second two being arousal and the third pair being dominance.

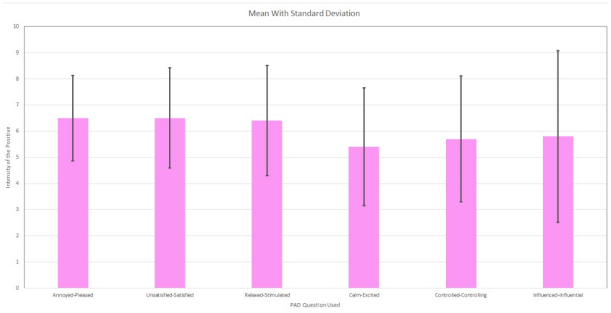


Fig. 6: Shows the mean with the standard deviation laid on top.