

Building a game-with-a-purpose to annotate mood on musical media data

Alissa Powers

June 6, 2017

Abstract

Current music algorithms do not recognize or generate the mood and musicality found in man-made music. To achieve this, automated systems must be trained on a reliable dataset that relates the qualities of music to human moods. Existing mood-media databases are limited for this purpose, so I propose a way to develop a large-scale database of this kind. The method I describe uses crowd-sourcing to annotate music elements with mood using a game-with-a-purpose (GWAP), a method that has been proven to be fast, inexpensive, and accurate. I describe how to combine programming and human computer interaction techniques to build a GWAP that collects data from human users and populates a database of different types of musical media annotated with mood.

1 Introduction

Music recommendation algorithms and music generation algorithms have become increasingly fundamental as technology continues to push the boundaries of music. Current music players, like Spotify and Google Play Music, use algorithms to recommend music to listeners. Meanwhile, music generators claim to be the future of music because they can create an infinite number of unique songs for listeners. Unfortunately, current music recommendation and generation algorithms fail at recognizing or producing the moods we hear in music that is composed or played by a human.

In order for computers to learn this skill, a reliable database of music annotated by mood is required as a training data set. Although some mood-media databases exist, the moods they index are vague (Kim, 2009; Rowe, 1992). For example, moodfuse plays songs classified by broad moods, such as angry, instead of more nuanced moods, such as enraged, annoyed, frustrated, etc. Additionally, these mood-media databases only contain music and lyrics, instead of including videos, animations, and 3D models (Kim, 2009; Rowe, 1992). It is important to have multiple types of media in the database because humans convey mood not just through the words they use, but also through body language, motion and tone of voice.

My goal is to create a large database of media clips from multiple musical mediums classified by mood. Hiring a specialist to label moods for the media clips would be expensive and would result in biased data. However, similar classification tasks have been successfully completed using aggregated human computation. It has been proven that in cases where individual humans may have some error, aggregating responses from many humans can quickly converge on the truth (Surowiecki, 2005). One example of the human computation approach is a game-with-a-purpose (GWAP), which is a game that uses crowd-sourcing to solve large-scale computational problems (Ahn & Dabbish, 2008). GWAPs have already successfully allowed humans to classify large amounts of data in the past; in 2006, Google used a GWAP that labeled images to improve image search results. For my purpose of classifying media clips by mood, I can assume that if a threshold number and percentage of users agree on the mood of a media clip then the mood is accurate for that clip.

Drawing on programming and human-computer interaction techniques from computer science and informatics, I plan to build a GWAP to collect data from human users to create a database of different types of media, such as lyrics, music, videos, and animations accurately labeled by mood.

2 Research

2.1 Why do we need a media-mood database?

Classifying the mood of music and lyrics has already become very important with the development of advanced music prediction algorithms, like those used to suggest songs on systems such as Google Play Music and Spotify. However, there is a lot more that can be done with media classified by mood.

First, we could improve the current systems for predicting music by having the music classified by more specific moods. They are currently classified by very broad moods or situations, like "breakup songs", "wake up call", and "good vibes". Everyone has different ideas or moods that fit into these categories, and they don't often predict exactly what people want to hear. For example, one person might want to hear really sad, painful, depressing songs after a breakup, while another person might want angry, vengeful songs, while still another might want songs of bitter insults and reassuring haughtiness. It seems obvious that a vague category of "breakup songs" would not satisfy most people.

Second, computers have become increasingly successful at "implement(ing) algorithmic methods for generating musical material" (Rowe, 1992). There are already experimental interactive music systems, machines that respond to performers' actions to produce music in a collaborative way. For example, the VIVO software allows a computer to play ambient music that reacts to motion. A dancer's movements, recorded by a camera and fed to the computer, can create changes in the ambient music being played by the VIVO system in real time (Paolizzo, 2016). Such interactive systems have a wide variety of uses, including services for audio on demand, improving access to musical heritage, education and training activities, music therapy, and music making. However, current interactive systems do not reflect the mood and musicality that can be produced by a human player. Currently, only placement within the frame and change in velocity can really affect the output of the algorithm, but it would be much more useful to have the music also match the mood of the dance. However, it would not be possible to do this without mood classifications for videos or three dimensional vector models.

Finally, there are many more research possibilities on the horizon but just out of reach due to lack of sufficient media-mood databases. Current databases classify entire songs with one or two moods, rather than classifying the mood of specific lines or stanzas. It would be really interesting to study how the varying moods of individual lines build on each other to make up the overall mood(s) of the piece. It would also be interesting to study how music affects human cognition and mood. Having the mood classifications for small clips of media is necessary for both of these studies as well as for future research possibilities that we have not even thought to explore yet.

2.2 How can large amounts of specialized data be collected?

It is very difficult to label media by mood. Hiring a specialist to take on such a large-scale task would be expensive (both in time and money) and would be biased: one person's opinions on the mood of a piece may reflect only their specific education and background. Luckily there is a better approach. Similar tasks have been successfully completed using human computation on a large scale.

As an example of the human computation approach, a game-with-a-purpose (GWAP) is a game that uses crowd-sourcing (a wisdom-of-the-crowd approach) to solve large-scale computational problems (Ahn & Dabbish, 2008). This method is good for populating a large database because it is fast, inexpensive, and accurate. Google used this method to label all of the images that display under Google Images, so that users can search and find relevant images. It is assumed (and has been proven through implementation in the Google Image Labeler and other GWAPs) that the percentage of humans that agree on a label for something often correlates with the validity of that label. In the case of populating a database with media classified by mood, getting a threshold number and percentage of users to agree on the mood of a media clip will be a good indicator that the media clip conveys that mood, and therefore can be labeled with that mood in the database. Although each individual person may be prone to error and bias, combining the efforts of a large group of people (on the scale of thousands) can result in stabilized data, with much smaller margins of error.

2.3 GWAPs

A game-with-a-purpose (GWAP) is a game that a large number of people can play and results in accomplishing a task that is impossible for computers. Because games can be played on-line, a GWAP can be played easily and inexpensively by many people all over the world. This makes it extremely easy to crowd-source and get a large enough group of people to respond, especially if the game is fun and encourages people to play not just for the contribution to research but actually because they enjoy playing.

There are different types of GWAPs and strategies for making a successful GWAP. Different types of successful GWAPs include output agreement games, input agreement games, and inversion-problem games.

2.3.1 Output Agreement Games

This is a generalization of the ESP Game, shown in Figure 1, which is the game used by Google to get labels for images so that when someone searches in Google Images, relevant images come up.

Initial setup: Two strangers are randomly chosen by the game itself from among all potential players.

Rules: In each round, both are given the same input and must produce outputs based on the input. Game instructions indicate that players should try to produce the same output as their partners. Players cannot see one another's outputs or communicate with one another.

Winning condition: Both players must produce the same output; they do not have to produce it at the same time but must produce it at some point while the input is displayed onscreen. (Ahn & Dabbish, 2008)



Figure 1: In this output-agreement game, the partners are agreeing on a label. (Ahn & Dabbish, 2008)

2.3.2 Input Agreement Games

This is a generalization of TagATune, shown in Figure 2, which is a game that labeled audio clips with descriptions of the sound.

Initial setup: Two strangers are randomly chosen by the game itself from among all potential players.

Rules: In each round, both players are given inputs that are known by the game (but not by the players) to be the same or different. The players are instructed to produce outputs describing their input, so their partners are able to assess whether their inputs are the same or different. Players see only each other's outputs.

Winning condition: Both players correctly determine whether they have been given the same or different inputs. (Ahn & Dabbish, 2008)

2.3.3 Inversion-Problem Games

This is a generalization of Peekaboom and Verbosity. Peekaboom, shown in Figure 3, labeled specific objects within images and Verbosity labeled words with common knowledge facts about them.

Initial setup: Two strangers are randomly chosen by the game itself from among all potential players.

Rules: In each round, one player is assigned to be the “describer,” and the other player is assigned to be the “guesser.” The describer is given an input. Based on this input, the describer produces outputs that are sent to the guesser. The outputs from the describer should help the guesser produce the original input.



Figure 2: In this input-agreement game, players received a sound, described it to each other, then guessed whether they both were given the same sound or different sounds. (von Ahn et al., n.d.)

Winning condition: The guesser produces the input that was originally given to the describer. (Ahn & Dabbish, 2008)

3 Approach

GWAPs have certain attributes, some that are general and apply to all video games and games in general, and others that are specific to GWAPs. Some of these general attributes include number of players, challenges, and rewards. GWAP-specific attributes include type of GWAP and human player attention (Ahn & Dabbish, 2008; Dulacka, Simko, & Bielikova, 2012). These attributes highly contribute to the overall success of the game. I researched the implementation of these attributes in GWAPs that have already been successful, some of which are listed in Table 1. I considered each of these when I was designing the basic structure and game-play and implemented the current design of the game with respect to these considerations.

3.1 Number of Players

One reason that many GWAPs are two-player games, using random partners, is because humans enjoy social connection, competition, and teamwork. It engages more players and motivates repeated play (Ahn & Dabbish, 2008). Unfortunately, this option is not always feasible. Google is a world-renowned company

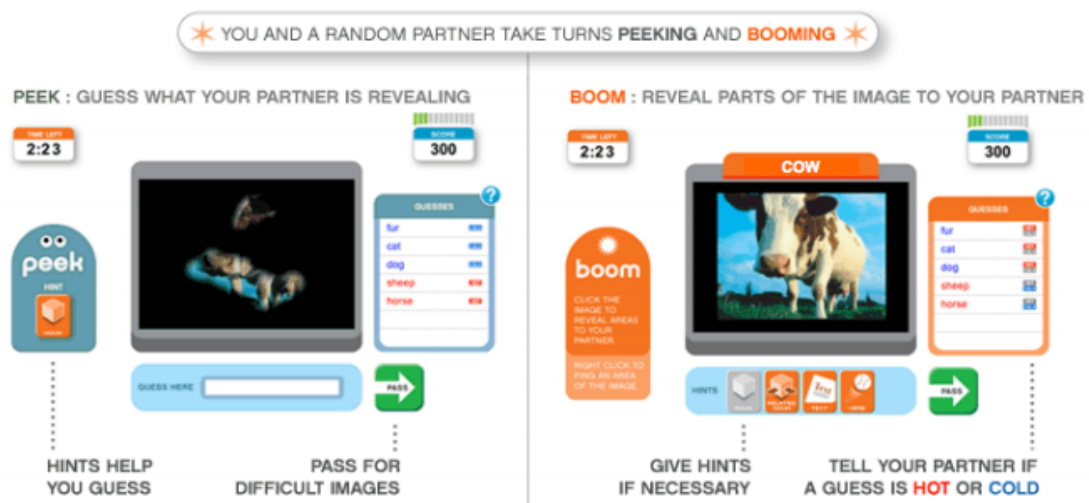


Figure 3: In this inversion-problem game, Boom gets an image along with a word related to it, and must reveal parts of the image for Peek to guess the correct word. Peek can enter multiple guesses that Boom can see. (von Ahn et al., n.d.)

Table 1: Comparison Table of GWAPs

Type	GWAP	Rewards	Players	Timed
Output Agreement	ESP Game	Score	Random partners	Yes
Input Agreement	TagATune	Score	Random partner	Yes
Inversion-Problem	Peekaboom	Top Scores List and Ranks	Random partners	Yes
	Verbosity	Score	Random partner	No
	Word Sleuth	Leaderboards, Rank, Badges	Single players	No

Overview of GWAP types, including example GWAP, reward mechanism, player mechanics, and game timing (Ahn & Dabbish, 2008; Dulacka et al., 2012; Law, von Ahn, Dannenberg, & Crawford, n.d.; Pearl & Steyvers, 2013; Speer, Havasi, & Surana, 2010; von Ahn et al., n.d.).

that can get millions of people all over the world to play their GWAP. They are always guaranteed to have enough players and can have pre-recorded games to handle cases where there is an odd number of players playing. They pair one person with the pre-recorded game and pair everyone else together, and this pre-recorded game does not affect their data much. However, being located at UC Irvine and not being quite as large as Google, we would mainly have people in California playing and would not be guaranteed to have a lot of players at any one time. It would be difficult to guarantee that there would be enough players to pair together randomly and we would often have to pair someone with a pre-recorded game that would impact the data. Instead, I decided to have individuals playing on their own and to incorporate the social connection when scores are calculated using all past players' responses. This approach was already shown successful in WordSleuth, where players never directly connected with other players in the moment, but still were affected by the other players as they were labeling a player's sentence or creating a sentence for others to label.

3.2 Type of GWAP

I decided to use a take on an output agreement GWAP, similar to the ESP Game, where players receive the same media clip and provide a label for it.

An input agreement game would not work for this because it requires that two players are playing at the same time so that they can both provide each other with labels to guess whether or not they both received the same clip of media. However, as discussed in the previous section, this could not be guaranteed. I needed to make a game that could be played by anyone at any time, regardless of how many others were playing.

An inversion-problem game is irrelevant to my goal, because it requires pre-labeled data that players are either verifying or revealing in a way that provides new data. For example, in Peekaboom there was an image pre-labeled 'cow.' The users revealed parts of the image that helped their opponent guess that it was a cow. So rather than labeling the image, they were helping show which parts of the image were relevant to the label. Another example is in WordSleuth. The players could choose to either provide sentences that expressed a given meaning or guess what other players' sentences expressed. Because players were supplying the data, an inversion-problem game worked. However, this kind of game does not work if you already have the data but do not have labels for the data.

This left me with an output agreement GWAP, which works perfectly for my goal. I have media clips, and want players to provide mood-labels for them. I know that a mood-label is accurate for that media clip when many different players all provide the same mood-label for the media clip.

3.3 Challenges and Rewards

Introducing challenge is crucial in order to get players to enjoy a game. As Ahn and Dabbish point out, "much of the previous work describing game-design principles cites challenge as a key aspect of any successful game." (Ahn & Dabbish, 2008) They go on to discuss how challenge manifests itself in a game, with timed responses, score-keeping, high-score lists, and randomness.

I decided to incorporate score-keeping, high-score lists, badges, and randomness into the GWAP in order to challenge the players. Scores are earned for each mood-label given as well as at the end of a game. Players can check the leaderboards to see the scores of the top players and are provided with their own rank information after each game. Randomness is incorporated by randomizing the media clips a player receives. Some are harder than others, and rather than starting with all the easy ones and getting harder, the player receives a variety of difficulty levels throughout the game.

I decided not to impose a time limit. As I learned during development, thinking of a mood-label for a given media clip is inherently challenging, and I did not want challenge to turn into stress by giving time limits for the responses. Also, I want specific, detailed, accurate mood-labels for the media clips, rather than the generic, obvious, repetitive mood-labels given under a time constraint.

3.4 Considerations: Human Player Attention

Because I am relying on humans to voluntarily play the game in order to collect data, the game must be enjoyable. People do not often spend time playing games that they do not enjoy, and with so many video games and apps competing for people's time and attention, the game has to be fun so that people will be drawn to it. The focuses listed below, recommended by leaders in GWAP design, are aimed at increasing the number of users as well as the average length of time that each user plays.

Design focuses include:

- Goals that are well-specified and challenging
- Attractive design and simple game-play
- Proper and motivating scoring
- Building up a player base

([Ahn & Dabbish, 2008](#); [Dulacka et al., 2012](#))

4 Design

4.1 The Current Design

The current design involves the user getting a snippet (that is, a media clip) and responding through text with a mood that they think the snippet creator was trying to convey. Points are awarded in relation to the popularity of their response compared to other players' responses to the same snippet. This encourages users to give a mood word or description that they think others will agree with. When a mood word reaches a threshold of popularity, it is added to the database as an appropriate annotation for that snippet. Through this method, the database is being populated.

The design and layout are simple, so that users will not get confused and discouraged. Also, the goal is very clear (to label the given snippet by the mood it conveys). In addition, there is an in-game tutorial that all new users play through as well as a standard tutorial that players can revisit as often as they like.

Players select to play a game and receive a randomized media clip. They can view the media clip and type in a response, as shown in Figure 4. The design is simple so that the style of the page does not bias their response. Players receive scores after each snippet they label, as shown in Figure 5. If they provide a new mood-label, they receive 100 points. Once there are more responses in the database, they receive scores that are multiplied by the percentage of players that also provided the same mood-label for the snippet and scaled to the number of total responses for the snippet.

The encouraging message and note provided to the player change as the player's score increases. If they give a really good mood-label (over 90% of players gave the same mood-label) then the score is much higher with added multipliers and the wording reflects the increased score. This provides feedback to the player about their score as well as encourages them to continue trying to think like other players, as has

Play Game

Snippet: Be rather the consoling angel of my family. Violetta. Think You still have time. Young lady, it is God who inspires these words on a father's lips.

Response:

*responses must only contain letters

SUBMIT

Skip this Snippet

Quit Game

Figure 4: Simple game-play with randomized snippet.

been proven successful in past GWAPs.

The player can continue the game and respond to more snippets or exit the game to receive their total game score.

At the end of a game, there are many motivators for the player to continue playing, as shown in Figure 6. The player receives their total score from the game along with their rank compared to other players. Rank is calculated from the total score of all games combined, so users are encouraged to play more games and earn more points during games to increase their rank and jump ahead of other players.

Players also receive badges during the game for accomplishments (i.e. reaching a high score, responding

Results

GOOD JOB!

You earned 100 points for your response. Keep trying to think of the best moods that other players would agree with to earn more points!

Note: You are unique! You are the first person to answer with this mood descriptor.
You have played 1 snippet so far!

Quit Game

Continue



Figure 5: Results after responding to one snippet with a new mood-label.

to a certain number of snippets during a game, etc.). The badges that they win during a game are shown at the end of the game along with the badges that they still need to earn. Players can see how close they are to earning another badge and be motivated to play another game.









Game Results

Congratulations, aap! You earned 100 points this round!

View the competition below and try to get ahead!

	rank	user	total score
 aap score: 600 rank: 15	1	 K.J.	228325
	14	 enrico tomaselli	650
	16	 123	300

Badges to Earn:

-  100 Meter Sprint - Respond to 10 snippets in a game.
-  Adventurer - Play 10 games
-  Around the World - Respond to 1000 snippets.
-  Explorer - Play 50 games
-  Marathon Runner - Respond to 50 snippets in a game.
-  Precious Gem - Respond with 50 High Scoring words.
-  Special Snowflake - Respond with 50 unique words.
-  The Whole Enchilada - Earn every badge.

Play Again

(a) Total game score and rank compared to other players.

(b) Badges the player has not earned yet.

Figure 6: Results of a game.

4.2 More Design Focuses

Other design focuses involve limiting opportunities for cheating, limiting the effectiveness of belligerence, encouraging accuracy, and rewarding the user. I address each of these in turn.

To decrease cheating, I began by using a large dataset currently made up of opera lyrics.¹ After acquiring this large dataset, I started pulling randomized snippets for the user to respond to. This decreases cheating because two users sitting next to each other do not get the same snippets and therefore are less likely to share answers. Also, duplicate snippets are never given to a user unless they already responded to all of the snippets in the database, so that they cannot make their mood annotation the most popular by using it multiple times for the same snippet. Once a user responds to every snippet in the database, they can continue to play for points, but their responses do not continue to increase the popularity of mood annotations in the database.

One common type of belligerence is attempting to compromise the data by entering junk as responses, irrespective of the impact on the score. However, the effect of these belligerent players is minimized in the current design. Because scores are based on popularity and because moods need to pass a threshold of popularity before they are added as a real annotation for the snippet, junk entries remain low in popularity and do not affect the data.

To encourage accuracy, players are awarded more points for more popular mood annotations. Players who care about their score attempt to respond with an accurate mood annotation, hoping to match the most common annotation. This leads to the most common annotations being accurate ones that can be added to the database.

To reward players for accurate responses and to encourage them to aim for high scores, the following reward systems are in place:

- Ranks that the user reaches by gaining some total number of points
- Leaderboards
- Badges that users can earn

One of the difficulties I faced during the design of the game was how to cause the least amount of bias in user responses. Because the layout, background music and sound effects, and graphics can all influence the mood of the players, which can then influence their perception of the mood conveyed by the snippet, it is important that these background features be as neutral as possible. This posed a problem, as these are also the things that help a player enjoy a game and make a game more fun. It was necessary to find a balance of design that interests the user while leaving their responses as unbiased as possible. I opted to stick with a very simple design, without flashy graphics, for the time being. There are a few colors, icons for badges, and profile pictures or avatars so that there is some visual appeal for players, but the simple aesthetics allow the players to absorb the media and feel the mood of the snippet rather than the mood of the game.

¹This is because there are many operas with expressive lyrics, they are all text and thus are relatively easy to retrieve, and because there are many available sources.

The game was implemented using WordPress. I decided to start with WordPress because it allowed for a relatively quick development of the prototype. It also worked well with the simple design and layout of the game. Although it does not seem intuitive for a web application, it works well for websites and provides themes for layout, account creation and management, and plug-ins for many more necessities that I would have otherwise had to build from scratch. One of the main challenges of getting a web application into WordPress was transferring form data and tracking data between pages. Using database queries and browser sessions, I was able to overcome this challenge to develop the game.

5 Evaluation of the GWAP

There are two main ways to evaluate the success of the GWAP. Because this GWAP is a video game, evaluation methods generally used to assess video games can be used. One of the best ways to evaluate video games is the playtest method (Fullerton, 2008; Schell, 2008), in which users start the game from the beginning in a controlled setting. After a certain amount of time, they answer questions about the game. Some of the questions are general, such as “How much fun did you have playing the game?” and “How clear were the controls and instructions?”. Others are specific to the game. The general questions are used to compare the current game to pre-existing games, and the specific questions allow the user to give feedback that directly applies to the game, which may provide the designer with specific information about what should be improved. Rating questions with a number scale are typically avoided for these game-improvement-oriented questions, as receiving only a number is not helpful when deciding specifically what to change. The playtest method provides the user with the opportunity to give direct feedback and specific examples of bugs, usability and enjoyment of a game.

Another method that can be used to evaluate the success of the game is the expected contribution (Ahn & Dabbish, 2008), which is an evaluation technique specific to GWAPs. The expected contribution is the throughput (average number of problems solved per human hour) times the average lifetime play (the average time a player plays across all players). This is calculated after the release of the game to measure its success and is a standard way to evaluate GWAPs.

Unfortunately, both of these methods require more time (especially more time with the game released to the public) than I currently have. They will be done in the future of the game after it is released to the public.

In order to evaluate the success of the current game before the public release, I went through a series of alpha and beta testing within the members of the Computation of Language Laboratory at the University of California, Irvine.

Research lab members were asked to play the alpha version of the game and fill out a short survey after. They were asked general questions such as "How enjoyable was the game?" and "How easy was it to navigate through the game and website?" These were rated on a scale of 1-10. The enjoyability of the game was rated at an average of 4.5 out of 10 and the ease of navigation was rated at an average of 8 out of 10. More specific questions such as "How difficult was it to think of a mood for each snippet" were rated on a scale of 1-5 with an average of 3 out of 5. The survey also included some short response questions so that players could provide explanations, complaints, ideas, and report bugs. Most players commented that the scoring and feedback after a game were confusing and inadequate, and a few players started playing without going to the tutorial first and were confused on how to play. The other main comment was that it did not keep player interest and that there was not enough motivation to play.

After the alpha testing, I improved the scoring system, making it more fair and intuitive. It took some adjustment of the ratios and bonus multipliers to get a scoring system that made more sense. Now, players get 100 base points if they are giving a new mood label for a snippet. If other players have given moods for the snippet, then the player also gets 10 extra points for every person that gave the same label as them. Once 100 players give moods for a snippet, a multiplier is applied to the score. This multiplier is calculated using a percentage (the number of players that gave the same label as the current player for this snippet out of the number of players that have given any label for this snippet) multiplied by 1000. Although I made the scoring more obvious, the game still needed to provide better feedback to the player about the quality of the score they earned. The quality of the score is not obvious to players at first because it takes a while for game players to grow accustomed to scoring systems for new games. For example, in some games 100

is a really good score and in others, like this game, it is the lowest number of points a player can earn. This is when I added in the congratulatory words and notes in the results page that change at certain thresholds of scores to provide feedback, explain the score, and show how many people gave the same mood-label. Now, if a player earns a score of 100 points, they also are told that they are giving a new mood label for this snippet. If the player earns a score of 150 points, then they are told that 5 people responded with the same mood label as them. I implemented an in-game tutorial that all new players go through the first time they play. This ensures that nobody dives into the game unaware of how to play. I also added the badges with icons, pictures and avatars for users, and displayed the player's rank compared to others after each game. After implementing these along with other improvements and bug fixes, I sent the game to the research lab members for beta testing.

Members were asked to play the beta version of the game and fill out another short survey after. They were asked the same general questions but this time the enjoyability of the game was rated at an average of 6.5 out of 10 and the ease of navigation was rated at an average of 9 out of 10. The survey again included some short response questions so that players could provide explanations, complaints, ideas, and report bugs. The scoring and lack of tutorial was no longer an issue. Although a few players said that the game was boring, most complained about the quality of the snippets and the lack of context in the snippets. Unfortunately, the lack of context in the snippets is required for the game; I want the mood conveyed by small clips of media out of context so that the labels can be useful for studying how the small parts make up the whole. However, the quality of snippets will improve. I was using a small test batch of snippets of varying lengths and quality for the test and will have a much larger dataset of well-chosen snippets and multiple media types for the public release of the game.

Based on the increased ratings from the alpha to the beta testing and the shift in the comments from focusing on the quality of the game to focusing on the quality of the snippets, the GWAP is successful. When a game is running smoothly, the players no longer notice the mechanics of the game and are more attentive to other pieces, like the snippets. Therefore I can conclude that the game has succeeded and will be able to fill a database of media labeled by mood.

The game is going into its final stages of development before being released to the public.

6 Acknowledgements

I would like to thank my mentors, Dr. Lisa Pearl and Dr. Fabio Paolizzo, members of the Computation of Language Laboratory at UC Irvine, all of the play-test participants, and audiences at the 2017 Undergraduate Research Symposium at UC Irvine.

References

- Ahn, L. V., & Dabbish, L. (2008, aug). Designing games with a purpose. *Communications of the ACM*, 51(8), 58-67.
- Dulacka, P., Simko, J., & Bielikova, M. (2012, sep). Validation of music metadata via game with a purpose. *Proceedings of the 8th International Conference on Semantic Systems*, 177-180.
- Fullerton, T. (2008). *Game design workshop: A playcentric approach to creating innovative games*. Elsevier Inc.
- Kim, C. (2009). *Running head: Interactive music systems for the web* (Unpublished master's thesis). The Steinhardt School: New York University.
- Law, E. L. M., von Ahn, L., Dannenberg, R. B., & Crawford, M. (n.d.). *Tagatune: A game for music and sound annotation*.
- Paolizzo, F. (2016, apr). *Vivo (video interactive vst orchestra): An interactive and adaptive musical instrument for self-reflection in music*.
- Pearl, L., & Steyvers, M. (2013). C'mon you should read this: Automatic identification of tone from language text. *International Journal of Computational Linguistics (IJCL)*, 3.
- Rowe, R. (1992). *Interactive music systems: Machine listening and composing*. MIT Pres.
- Schell, J. (2008). *The art of game design: A book of lenses*. Elsevier Inc.

- Speer, R., Havasi, C., & Surana, H. (2010). Using verbosity: Common sense data from games with a purpose. *Proceedings of the Twenty-Third International Florida Artificial Intelligence Research Society Conference*.
- Surowiecki, J. (2005). *The wisdom of crowds: A book of lenses*. Anchor Books.
- von Ahn, L., Liu, R., & Blum, M. (n.d.). *Peekaboom: A Game for Locating Objects in Images*.