User-Centered Design of a Social Game to Tag Music

Luke Barrington
Electrical & Computer Engineering
U.C. San Diego
lukeinusa@gmail.com

Douglas Turnbull
Computer Science
Swarthmore College
turnbull@cs.swarthmore.edu

Damien O'Malley
Music Search, Inc.
San Diego
555damien@gmail.com

Gert Lanckriet
Electrical & Computer Engineering
U.C. San Diego
gert@ece.ucsd.edu

ABSTRACT

We present "Herd It", a competitive, online, multi-player game that has the implicit benefit of collecting tags for music. We describe Herd It's user-centered design process and demonstrate that the game can collect both musical and social data. This data can be used to build machine learning models that automatically associate music with tags. Herd It differs from previous "games with a purpose" in that it is designed to be social: the game runs on the Facebook online social network and scoring is based on consensus between a large group of listeners - "the Herd". By presenting music in a social context, Herd It adds demographic context to the semantic music descriptions that it collects.

Keywords

Auditory I/O and sound in the UI, Distributed knowledge acquisition, Computer audition, Web-based games with a purpose

Categories and Subject Descriptors

I.2.6 [**Learning**]: Knowledge acquisition; H.5.3 [**Group and Organization Interfaces**]: Web-based interaction; H.5.5 [**Sound and Music Computing**]: Systems; I.2.1 [**Applications and Expert Systems**]: Games

1. INTRODUCTION

With hundreds of millions of songs available on the Internet, there is a need for new music search and discovery tools that can automatically analyze this data[11]. We are interested in developing a *query-by-description* system in which an individual can use words to describe the music that they want to retrieve (e.g., "Google for music"). For example, a user may request "energetic classic rock songs that feature distorted electric guitar riffs".

Before such a system can retrieve music, we must first annotate (i.e., index) each song with *tags*: short text-based tokens, such as "energetic", "classic rock" and "distorted electric guitar" that describe some meaningful aspect of the song. However, both the huge (and growing) number of songs and the potentially large vocabulary of useful tags make annotating music a difficult and labor-intensive task. This task is made more difficult by the fact that music is inherently subjective: different listeners will often annotate the same song with different tags.

In this paper, we present *Herd It*[1]: a social game that collects tags about music. Our "game with a purpose" is deployed on Facebook (`www.facebook.com`), a popular online social network, and allows simultaneous data collection from multiple players for a diverse set of songs. In addition, we collect personal information from each game player which allows us to combine data from demographically or psychographically similar players. This, in turn, can help provide personalized search results for a query-by-description music retrieval system. While the annotations collected with our game can be used directly to index the songs that are presented in the game, they are also useful as training data for our computer audition system that can automatically tag music without requiring direct human involvement.

In the following sections, we describe how the game has been designed to collect both musical and social data. Herd It differs from previous games in that it was built to be highly social in nature. Since music is often used in a social setting (e.g., concerts, parties, icebreakers, etc.[9]), the game is found to be more interesting to users when it encourages social interaction.

2. TAGGING MUSIC WITH GAMES

Tags are text-based tokens that can be used to annotate songs. They represent a rich source of semantic information that is useful for text-based music retrieval (e.g., [12]), as well as recommendation, discovery, and visualization [6]. Tags can be collected from humans using surveys (e.g., `www.pandora.com`) or social tagging websites (e.g., `www.last.fm`). They can also be generated by text mining web-documents [5, 18] or by autotagging audio content [12, 2, 14]. All these methods suffer from drawbacks - surveys are expensive, social tags and web documents introduce a lot of noise, non-automatic methods suffer from the "cold-start problem" where it takes a long time for new or unpopular songs to be richly tagged and automatic tagging requires large amounts of well-tagged data to achieve human levels of accuracy - currently only available for a small number of musically-relevant tags [12].

[1]Play Herd It at `http:www.herdit.org`

So-called “games with a purpose" [15] have been designed to tag images [16], collect common-sense facts [17] and even label galaxies (www.galaxyzoo.org). ListenGame [13], Tag-a-Tune [7], MajorMiner [8] and MoodSwings [4] are examples of games with a purpose that attempt to collect music tags accurately, rapidly and cheaply. However, these games have largely been small-scale research efforts and have attracted limited user traffic. As yet, no compelling music-tagging game has been able to harvest data in sufficient quantities to tag the huge and rapidly growing collection of digital music. For example, during a two-week pilot study, ListenGame collected 26,000 annotations (best and worst votes) for 250 songs using a 120-tag vocabulary. Each of the 27,250 possible song-tag pairs were presented to users an average of 1.8 times [13]. Contrasting this to the scale of real-world problems, the online social music network Last.fm catalogs over 100 million songs that users have associated with more than 1.2 million unique tags[11].

A truly compelling, well designed, viral music tagging game might succeed in collecting data on a sufficiently large scale to tag the world’s music. Alternatively, the solution to this problem may lie with “auto-tagging" - machine learning methods that can automatically generalize information collected from a small set of well-tagged songs to all songs. When trained with precise data, these auto-tagging systems have shown the ability to predict tags with accuracy equivalent to humans [12]. Designed with precisely this goal in mind, ListenGame collected tags that were used to train an auto-tagger that achieved accuracy on a par with a system trained using expensive, laborious tags collected from a survey [13].

In the remainder of this paper, we describe the design and development of “Herd It" - a new game for collecting music tags that has just been launched. Herd It improves on the games described above in two main areas. First, it has been designed to be a fun, social, multi-player, real-time game, emulating the scenario in which music is often enjoyed and maximizing its potential for collecting large amounts of data by engaging high volumes of users. Second, Herd It has been designed to work within the Facebook social network, facilitating distribution and user interaction and also adding social context to the data collected. Given the extra information about users and their social networks that the Facebook platform offers, Herd It can add demographic and sociographic context the musical descriptions that it collects.

3. HERD IT GAMEPLAY

Herd It is a Facebook application. A player arrives at the game, often as the result of an invitation from a friend who has played before. The first screen presents the player with some game-related information (high scores, instructions and tutorial, a list of friends who are playing) and invites them to choose the genre of music they would like to listen to (currently “pop", “rock", “blues", “hip hop", “electronic" or “everything", although more are being added). Once they choose a genre, they are connected in realtime with all other players who are currently listening to the same genre - “The Herd" - and the game begins. If there are not at least 10 human players currently active, artificial “bots" make sure that there are plenty of faces in the Herd.

The same clip of music is played to all players and a minigame quizzes players on their opinions about the music. The minigame might ask the player to select an appropriate musical sub-genre, the most prominent instrument, or a time when it would be good to listen to the piece of music. The minigame lasts 10-20 seconds and requires just a single click for the player to enter their choice. For example (see Figure 1), one minigame shows the player six bubbles floating around the play area, each containing adjectives that might describe the emotion evoked by the music. The player clicks the most appropriate bubble and all other bubbles disappear with a pop. After the minigame, the player sees feedback on their choice (e.g., the bubble that they clicked remains on screen) as well as on the Herd’s choices (e.g., lots of new bubbles, representing the choices of the other players in the Herd, float on screen). During this feedback display, the player is awarded points equal to the percentage of the Herd that agreed with them: consensus with the Herd is rewarded. Following each minigame, the player can earn 20 bonus points by correctly naming the song or the artist they have been listening to in a multiple-choice trivia round.

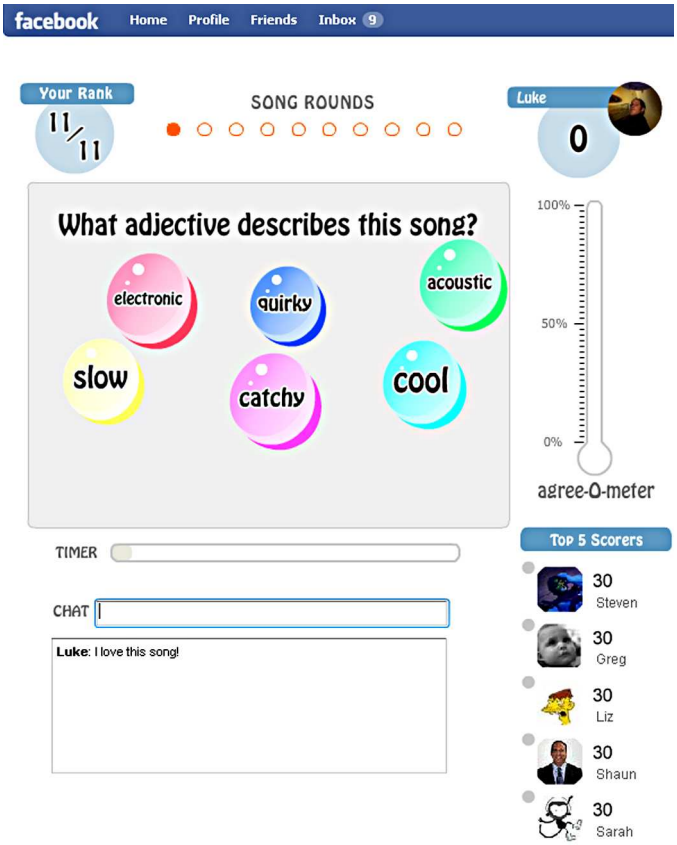

Figure 1: Herd It game screen.

The sequence of one minigame and one trivia round is repeated for 5 different songs for a total of 10 rounds. A variety of fun, simple minigames are currently deployed including clicking floating bubbles, measuring yes/no responses on a weighing scale, selecting the most appropriate sub-genre from a grid and an XY valence plot measuring emotion and activation [10, 4]. At the end of the game (lasting 2-3 minutes), a summary screen presents the final scores, lists the songs that were played during the game and identifies which of the other players in the Herd most agreed with the current player - their “musical twin".

4. DESIGN GOALS OF HERD IT

We had two requirements: design a game that is entertaining (and therefore popular) and collects accurate music tag data. Research involved demographic and psychographic analysis, both quantitative and qualitative. Extensive competitive analysis was made of existing casual games on the internet, consoles and stand-alone PC games, both traditional games and games with a purpose. Based on this

research and data collated from user interviews, it emerged that our target audience were “casual gamers" - people who play simple games as a brief distraction. Unlike traditional video gamers, the main demographic for casual games is women aged 35 to 50, who play for 5 to 20 minutes [1].

4.1 Goal 1: Engage players

The game should attract and retain a large volume of players with fun games and cool music using:

- visually appealing and intuitive user interface
- genre-specific games matching each player’s musical tastes
- competitive scoring & realtime feedback
- foster a sense of community by displaying other players’ selections, enabling real-time chat and suggesting friendships based on musical compatibility

4.2 Goal 2: Collect Data

The game should collect data that represents:

- accurate associations between songs and tags to directly index music
- accurate associations between songs and tags to train an auto-tagging system - allows us to annotate additional songs without direct human involvement
- personal information about players, for personalized music search and recommendation
- a musically diverse set of songs

5. USER-CENTERED DESIGN

The game underwent a user-centered design process. Rapid prototyping produced a playable game every 4-6 weeks, or at every significant milestone in its development. User tests were conducted on each prototype with 5-10 people to determine many factors including playability and appeal, UI intuitiveness, viral potential and stability. Users were observed both in person (to determine efficacy of the interface) and remotely (the game is designed to be played alone at a computer connected to the internet). User test results very quickly highlighted the flaws and weaknesses of the game and the challenge of creating an appealing game based around the strict data/tagging requirement proved extremely difficult.

Invaluable to the design process were suggestions by the users on how to make the game better. Many features currently in the game were developed in response to user requests and, on testing, these features often proved to be great improvements. The key issues that were highlighted by this design process and the resulting implementations are discussed below. Quotes are taken from a series of user interviews (roughly one every 2 months over the past year) and statistics are taken from our final beta test, after which 16 users completed an online questionnaire.

5.1 A Social Game

Music is an inherently social experience: music is the most common topic of conversation between strangers getting acquainted and individuals use their music preferences to communicate information about their personalities to observers [9]. Many existing music games are played either alone [8] or in pairs [7, 4]. Beyond simple high score lists, none foster social interaction between players.

Herd It began as a five-player game on a stand-alone web site. It was discovered that the game had a far greater likelihood of becoming popular and getting repeat visitors if placed in a social networking environment where users benefit from the social aspects of both music and casual games (e.g., where users can be notified of their friends performance and vice versa). Facebook proved to be the ideal platform as users primarily use Facebook to keep in touch with friends and make contact with potential new acquaintances [3]. Furthermore, Facebook provides an API for third-party applications which can then take advantage of the users’ social networks. Playing games and using applications is another major use of Facebook [3] and Facebook applications have increased potential to reach a mass audience, compared to a stand-alone site with low promotional resources. Tables 1 and 2 demonstrate user awareness and effect of social interactions.

Very aware	Somewhat aware	Not at all aware
64.3%	28.6%	7.1%

Table 1: Are you aware of other people playing the game with you?

Definitely	Likely	Maybe	Unlikely	No Way
42.9%	50%	7.1%	0%	0%

Table 2: Would you try to beat a friend’s high-score if you saw it on your Facebook homepage?

As the game evolved though many iterations of testing and design, basic playability and UI problems were solved. The focus then shifted towards adding features that enhanced the appeal of the game, such as realtime chat, automatically notifying a player’s friends when new high-scores are set, discovering each player’s musical profile and, determining the player’s “musical twin" at the end of a game. In subsequent testing, these new features - common in social games but not available in other music games - proved to broaden and deepen its appeal. *“This would allow me to flirt by contacting girls with similar tastes!"* J, male, 35.

5.2 Focus on Music

Incorporating the auto-tagging system into the game provides intelligent tag suggestions for each song. This makes the game more enjoyable by ensuring that minigames present players with relevant tags for describing each song and also creates a feedback loop that maximizes data collection efficiency. We make music the focus of the gameplay by ensuring that attention to the music is required to complete all minigames. Keeping the minigame interactions simple (one click) allows players to listen to and analyze the music without being distracted by the response interface.

Players begin the game by selecting the musical genre of the songs that they will hear in the game, motivating them to listen carefully, encouraging them to play for longer, and making it more likely that they will have experience with the music and understanding of the words used to describe it.

“One of the coolest games I’ve played - and with the added bonus of getting to listen to good music!" D, female, 33.

“I couldn’t bear to play this game if I thought I’d hear any terrible J-pop tunes." B, female, 32.

The possibility of serendipitous discovery of new, enjoyable music (the song and artist names are revealed at the end of the game) encourages repeat play.

“I liked hearing songs I would not otherwise hear." R, female, 30.

5.3 Engaging Gameplay

Initial iterations of Herd It used fast-paced, gestural minigames that required traditional gaming traits such as quick reactions, hand-eye coordination and puzzle solving. It was quickly determined that these challenges did not appeal to the casual gaming audience and that they distracted from the main goal of attending to the music. In the final design, all minigames require just a single click

Very Clear	Quite Clear	Yes	Sort of	Confusing
0%	37.5%	62.5%	0%	0%

Table 3: Did the questions and words clearly correspond to the music?

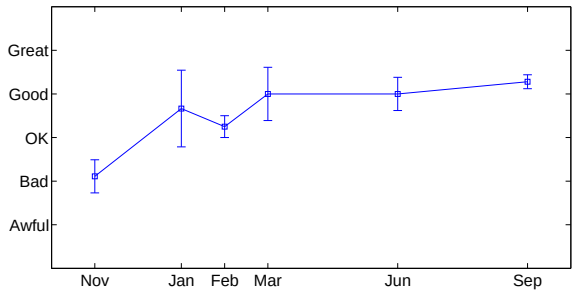

Figure 2: Evolution of overall game enjoyment over 10 months of user-centered design and development

to enter a response, focusing the gameplay on listening to, enjoying and analyzing the music, comprehending the available tags and predicting the responses of the Herd to maximize scores.

Likewise, the scoring system began by having complicated methods of evaluating the level of agreement between players, including negative scoring when a player's chosen tag disagreed with the Herd. The scoring was displayed in a player-by-player matrix that showed all pairs of player agreement. This display was found to be overwhelming and unnecessary in our user tests. We also found that negative scoring (e.g., subtracting points) turned players off the game because they were being explicitly penalized for their unique opinion. The final design uses a more positive scoring mechanism: in each minigame, players score points equal to the percentage of the Herd that agrees with them. This percentage agreement is displayed on the "Agree-O-Meter" (see Figure 1), a thermometer that ranges from 0 to 100. In addition to the Agree-O-Meter, each minigame has a custom animated recap screen. The recap screen gives an intuitive visual representation, specific to the minigame used to input a choice, of the popularity of each choice among the Herd.

Tables 4 and 5 demonstrate overall player enjoyment and likelihood of telling a friend about Herd It. Figure 2 demonstrates the effectiveness of the user-centered design process by charting the evolution of overall player enjoyment throughout the one-year development of the game.

Great	Good	OK	Bad	Awful
35.7%	57.1%	7.1%	0%	0%

Table 4: Overall, how did you like the game?

Definitely	Likely	Maybe	Unlikely	No Way
44.4%	44.4%	11.1%	0%	0%

Table 5: Would you recommend the game to your friends?

5.4 Reliable Data

Scoring that rewards consensus encourages players to select the tag they think is most relevant to the music since this will likely agree with the Herd. This ensures that tags collected by Herd It are likely to be reliable. With a new minigame every 20 seconds, Herd It collects tags at a rate of approximately 3 per player per minute.

6. HERD IT GOES LIVE

On April 15, 2009 Herd It was released into public beta-testing and is now available at `www.herdit.org`. A limited alpha-test has already verified that Herd It has the potential to attract lots of users and collect large volumes of data. 1049 players have played 13,831 rounds of the game. This has produced 9,941 song / tag pairs that have been verified by at least 2 users, including 643 unique tags and 3442 unique songs. For comparison, a much wider distribution and significant promotion for ListenGame resulted in 26,000 verified song / tag pairs but for only 120 tags and 250 songs [13].

7. REFERENCES

[1] International Game Developers Association. Casual games white paper. 2006.

[2] D. Eck, P. Lamere, T. Bertin-Mahieux, and S. Green. Automatic generation of social tags for music recommendation. In *NIPS*, 2007.

[3] A.N. Joinson. 'Looking at', 'looking up' or 'keeping up with' people? Motives and uses of Facebook. In *ACM CHI*, 2008.

[4] Y. Kim, E. Schmidt, and L. Emelle. Moodswings: A collaborative game for music mood label collection. In *ISMIR*, 2008.

[5] P. Knees, T. Pohle, M. Schedl, and G. Widmer. A music search engine built upon audio-based and web-based similarity measures. In *ACM SIGIR*, 2007.

[6] P. Lamere and O. Celma. Music recommendation tutorial notes. ISMIR Tutorial, September 2007.

[7] E. Law and L vonAhn. Input-agreement: A new mechanism for collecting data using human computation games. In *ACM CHI*, 2009.

[8] M. Mandel and D. Ellis. A web-based game for collecting music metadata. In *ISMIR*, 2007.

[9] P.J. Rentfrow and S.D. Gosling. Message in a ballad: The role of music preferences in interpersonal perception. *Psychological Science*, 17(3), 2006.

[10] J.A. Russell. Core affect and the psychological construction of emotion. *Psychological Review*, 110(1), 2003.

[11] D. Turnbull. *Design and Development of a Semantic Music Discovery Engine*. PhD thesis, University of California, San Diego, 2008.

[12] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. Semantic annotation and retrieval of music and sound effects. *IEEE TASLP*, 16(2):467–476, February 2008.

[13] D. Turnbull, R. Liu, L. Barrington, D. Torres, and G Lanckriet. Using games to collect semantic information about music. In *ISMIR*, 2007.

[14] G. Tzanetakis and P. R. Cook. Musical genre classification of audio signals. *IEEE Transaction on Speech and Audio Processing*, 10(5):293–302, 7 2002.

[15] L. von Ahn. Games with a purpose. *IEEE Computer Magazine*, 39(6):92–94, 2006.

[16] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *ACM CHI*, 2004.

[17] L. von Ahn, M. Kedia, and M. Blum. Verbosity: A game for collecting common-sense facts. In *ACM CHI*, 2006.

[18] B. Whitman and D. Ellis. Automatic record reviews. In *ISMIR*, 2004.