A WEB-BASED GAME FOR COLLECTING MUSIC METADATA

Michael I Mandel
Columbia University
LabROSA, Dept. Electrical Engineering
mim@ee.columbia.edu

Daniel P W Ellis
Columbia University
LabROSA, Dept. Electrical Engineering
dpwe@ee.columbia.edu

ABSTRACT

We have designed a web-based game to make collecting descriptions of musical excerpts fun, easy, useful, and objective. Participants describe 10 second clips of songs and score points when their descriptions match those of other participants. The rules were designed to encourage users to be thorough and the clip length was chosen to make judgments more objective and specific. Analysis of preliminary data shows that we are able to collect objective and specific descriptions of clips and that players tend to agree with one another.

1 MOTIVATION AND GAME PLAY

The easiest way for people to find music is by describing it with words. Whether hearing about a new band from a friend, browsing a large catalog, or locating a specific song, verbal descriptions, although imperfect, generally suffice. While there are notable community efforts to verbally describe large corpora of music, only an automatic music description system can adequately label brand new, obscure, or unknown music. To train such a system, however, requires human generated descriptions.

Thus in this project, we endeavor to collect ground truth about specific, objective aspects of music by asking humans to describe short musical excerpts, which we call clips, in the context of a web-based game [1]. Such a game entertains people while simultaneously collecting useful data. Not only is the data collected interesting, but the game itself makes novel contributions to the field of "human computation."

Here is an example of a how a player experiences the game. First she requests a new clip to be tagged. This clip could be one that other players have seen before or one that is brand new, she does not know which she will receive. She listens to the clip and describes it with a few words: *harp*, *female*, and *sad*. The word *harp* already has been used by exactly one other player, so it scores her one point. In addition, the player who first used it scores two points. The word *female* has already been used by at least two players, so it scores our player zero points. The word *sad* has not been used by anyone before, so it scores no points immediately, but has the potential to score two points should another player subsequently use it.

The player then goes to her game summary. The summary shows both clips that she has recently seen and those that she has recently scored on, e.g. if another user has agreed with one of her tags. It also reveals the artist, album, and track names of each clip and allows the user to see another user's tags for each clip. The next time she logs in, the system informs her that three of her descriptions have been used by other players in the interim, scoring her six points while she was gone.

A number of authors have explored the link between music and text, especially Whitman [4]. More recently, [2] has applied ideas from the image retrieval literature to associate text with music. In the ESP Game [3] pairs of players describe the same image and score points when they agree. This game popularized the idea of allowing free form responses that only score points when verified.

2 DESIGN CONSIDERATIONS

We designed the game with many goals in mind. Our main goal, which shaped the design of the scoring rules, was to encourage users to describe the music thoroughly, to be original, yet relevant. Our second goal, which informed the method for picking clips to show, was for the game to be fun for both new and veteran users. We also wanted to avoid cheating, collusion, or other manipulations of the scoring system or, worse, the data collected.

While games like the ESP game pair a player with a single partner, ours in a sense teams a player with all of the other players who have ever seen a particular clip. It is possible that a pair of players could vary widely in skill level or familiarity with the clip under consideration, frustrating both players. The non-paired format allows the most creative or expert players to cooperate with each other asynchronously. It also allows the systematic introduction of new clips, avoiding a "cold start." These benefits come at the price of vulnerability to asynchronous versions of the attacks that afflict paired games.

The design of the game's scoring rules reflects our first goal, to encourage users to thoroughly describe clips. To foster relevance, users only score points when other users agree with them. To encourage originality, users are given more points for being the first to use a particular description on a given clip and are given no points for a tag that

[1] The game is available to play at: http://game.majorminer.com

Label	Verified	Label	Verified
drums	793	vocals	120
guitar	720	jazz	120
male	615	voice	119
rock	571	vocal	118
synth	429	hip hop	118
electronic	414	slow	112
pop	375	80s	94
bass	363	beat	89
female	311	fast	84
dance	297	drum machine	83
techno	224	british	68
electronica	155	country	65
piano	153	soft	58
rap	140	instrumental	55
synthesizer	136	house	53

Table 1. The 30 most popular tags and the number of clips on which each was verified by two players.

two players have already agreed upon. Currently, the first player to use a particular tag on a clip scores two points when it is verified by a second player, who scores one point. By carefully choosing when clips are shown to players, we can adjust the difficulty of scoring and use the tension created by the scoring rules to inspire originality without inducing frustration.

When a player requests a new clip, we have the freedom to return whatever clip we like and we adjust the choice based on the players experience level. We either draw a clip from the pool of clips that have been seen by other players, which are the only clips on which immediate scoring is possible, or a brand new clip, introducing it into that pool. For players who have not seen many clips yet, we draw clips from this pool to facilitate immediate scoring. For players who have seen a fraction, γ, of this pool, we usually draw clips that have already been seen, but with probability γ draw a brand new clip. While new clips do not allow immediate scoring, they do offer the opportunity to be the first to use many tags, thus scoring more points when others agree later.

Once a player has labeled a clip, he has the opportunity to see the name of the song and the performer along with the labels that another player has used to describe that same clip. We choose to reveal the labels of the first player to describe a given clip, these labels will remain the same no matter how many subsequent players see the clip. Because of the clip choice described above, the labels revealed are likely to be those of an experienced user and can then serve as exemplars for new players.

3 DATA COLLECTED

The type of music present in the database affects the labels that are collected and our music comes from four sources. The first, and biggest source, contained electronic music, drum and bass, post-punk, brit pop, and indie rock. The second contained indie rock and hip hop. The third contained pop, country, and contemporary rock. And the last contained jazz.

At the time of this paper's writing, the site had been live for 3 months, in which 361 users had registered. A total of 2183 clips had been labeled selected at random from 2547 tracks, each with an average of 25.0 clips. Each clip had on average been seen by 6.03 users, and described with 27.56 tags, 4.48 of which had been verified. See Table 1 for some of the most frequently used descriptions.

Certain patterns are observable in the collected descriptions. As can be seen in Table 1, the most popular tags describe genre, instrumentation, and the gender of the singer, if there are vocals. People do use descriptive words, like soft, loud, quiet, fast, slow, and repetitive, but less frequently. Emotional words are less popular, perhaps because they are difficult to verbalize in a way that others will likely agree with. And there are hardly any words describing rhythm, except for an occasional *beat*. Lyrics also prove to be useful tags, and a corpus of music labeled with lyrics might facilitate lyric transcription.

Players also use the names of artists they recognize. For example, *cure* has been verified 12 times, *bowie* 8 times, and *radiohead* 6 times. Of the clips verified as *bowie*, however, three were performed by Gavin Friday, Suede, and Pulp. This label most likely indicates songs that sound like David Bowie's music regardless of the actual performer. Artists used in this way could be the anchors in an "anchor space" where music is described by its similarity to that of well known artists [1].

4 FUTURE WORK

There is much that we would like to do with this data in the future: train models to automatically describe music, analyze the similarities between clips, between users, and between words, investigate ways to combine audio-based and word-based music similarity to help improve both, use automatic descriptions as features for further manipulation, investigate an anchor space built from the data collected here, use descriptions of clips to help determine the structure of songs, and so forth.

4.1 Acknowledgments

Thanks to Marios Athineos, Graham Poliner, Neeraj Kumar, and Johanna Devaney. This work was supported by the Fu Foundation School of Engineering and Applied Science via a Presidential Fellowship, and by the Columbia Academic Quality Fund, and by the National Science Foundation (NSF) under Grant No. IIS-0238301. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

5 REFERENCES

[1] Adam Berenzweig, Daniel PW Ellis, and Steve Lawrence. Anchor space for classification and similarity measurement of music. In *Proc Intl Conf on Multimedia and Expo (ICME)*, 2003.

[2] Douglas Turnbull, Luke Barrington, and Gert Lanckriet. Modeling music and words using a multi-class naive bayes approach. In *Proc Intl Symp Music Information Retrieval*, October 2006.

[3] Luis von Ahn and Laura Dabbish. Labeling images with a computer game. In *Proc SIGCHI conference on Human factors in computing systems*, pages 319 – 326, 2004.

[4] Brian Whitman and Daniel PW Ellis. Automatic record reviews. In *Proc Intl Symp Music Information Retrieval*, 2004.