Development of a Music Recommendation System for Motivating Exercise

Jiakun Fang*, David Grunberg*†, Simon Lui†, and Ye Wang*
*School of Computing
National University of Singapore, Singapore
Email: {fangjiak, grunberg, wangye}@comp.nus.edu.sg
†Department of Information Systems Technology and Design
Singapore University of Technology and Design, Singapore
Email: {david_grunberg, simon_lui}@sutd.edu.sg

***Abstract*—While the health benefits of regular physical activity are well-established, many people exercise much less than is recommended by established guidelines. Music has been shown to have a motivational effect that can encourage people to exercise more strenuously or for longer periods of time, but the determination of which songs should be provided to which exercisers is an unsolved problem. We propose a system that incorporates user profiling to provide a strong set of initial recommendations to the user. Reinforcement learning is then used as each recommendation is accepted or rejected in order to ensure that subsequent recommendations are also likely to be approved. Test subjects who used the proposed system rated the playlists it provided more highly than those provided by a prior state-of-the-art reinforcement learning-based music recommendation system and also did not need to reject as many songs before being satisfied with their recommendations, both when receiving recommendations based on individual profiles, and when receiving recommendations based on aggregate profiles formed by grouping the users.**

***Keywords*-Music recommendation, exercise, reinforcement learning.**

I. Introduction

The health benefits of exercise have been widely established [1]. People who exercise regularly have been shown to have lowered risks of developing a variety of medical conditions, including obesity, certain cancers, diabetes mellitus [2], stroke [3], and coronary heart disease [1]. And though many people do not like to exercise, music has been shown to have a motivational effect which can encourage people to exercise more strenuously or for longer, thus helping them stay healthier [4], [5]. However, it has also been shown that not all music has an equally strong motivational effect; while some music is very motivational for certain people, other music has no apparent effect on exercise habits [4]. Furthermore, there may not be universally-motivational music due to individual preferences and cultural backgrounds; for instance, different groups of subjects have alternately indicated that sets of music are more likely to be motivational if they are fast [6], slow [7], or of mixed tempi [8].

While determining the motivational content of a piece of music in general may not be tractable, studies have shown that it is possible to reliably select motivational music for *specific users* by incorporating knowledge of their musical preferences. Algorithms such as reinforcement learning can be used to help determine such preferences; however, without additional knowledge about the user, such algorithms can require users to evaluate literally hundreds of tracks to obtain enough information to provide quality results [9]. This can fatigue users and can take a very long time, making such systems inconvenient to use. We therefore propose a system which incorporates a user profiling step that represents a user's musical and exercise preferences, allowing the algorithm to provide satisfactory music in a practical manner. We also extend the user profile to represent groups as well as individuals so that the proposed system can recommend music for group activities, such as physical therapy, that is appropriate for all participants.

II. Literature Review

In addition to general-purpose music recommendation systems [10], several task-specific music-recommendation systems have been developed by researchers [11]. However, the specific activity of exercise has aspects which require special care that is not required in the general case or for other tasks such as driving a car. For example, a user might move at a very specific pace or period during exercise, while they are unlikely to do so when trying to fall asleep, driving a car, reading, or simply listening to music in general. As such, systems designed to recommend music for exercise must take aspects such as strict tempo restrictions into account.

Some systems have been developed to recommend music specifically for exercise [12], [13], [14]. TripleBeats provides a user with faster or slower music depending on whether the system determines that the user should speed up or slow down [14]. PersonalSoundtrack measures a user's running pace and recommends music with an equivalent tempo [13], and IM4Sports uses a training phase which predicts future recommendations based on music a user has previously selected for exercise [12]. These systems, however, mostly offer recommendations based just on the tempo of the music (IM4Sports also incorporates artist and genre features). They neglect other features such as physiological arousal which are likely to influence how motivational a piece of music is [5].

III. Algorithm

Initially, the proposed system has approximately 384,500 songs that can be recommended. These songs are

978-1-5386-3276-5$31.00 ©2017 IEEE

obtained from tracks in the Million Song Dataset (MSD) and which also have previews available from 7Digital.com [15]. However, much of this audio is likely to be a poor match to any specific user. The proposed system therefore uses a User Profiling (UP) step. Music therapy programs often use questionnaires to help therapists find music that is appropriate for specific users; this system therefore uses a questionnaire to create a user profile [16]. The information obtained from the questionnaire is compared with musical features to prune the initial 384,500 songs to 1000 songs the user is more likely to enjoy in order to make the recommendation problem more tractable. The questionnaire is also used to cluster users into groups with similar musical tastes; aggregate user profiles are then created in order to find recommendations that are likely to be acceptable to the group as a whole. Finally, Reinforcement Learning (RL) is used to refine initial results provided by the UP recommendation system.

A. User Profiling System

Upon subscribing to the system, a user is first directed to a page containing the questionnaire which will be used to determine their profile. The questions were designed to correlate with features which were mentioned frequently in the literature as having a potentially strong influence on the motivational quality of a piece of music such as tempo [4], [5].

Users enter their favorite songs, artists, and genres into text boxes. Research indicates that a user is more likely to be motivated by a piece of music if they personally like the song [5], and this portion of the questionnaire allows for a user to list that information to help the system find similar music that they are also likely to enjoy. Users then input their tempo preferences for exercise music on a five-point scale, and also have the option of tapping their tempo. Finally, users determine on a five-point scale how loud, energetic, positive, and familiar they like their exercise music to be, and how prominent they like the rhythm in their exercise music to be.

B. Clustering Technique

After the questionnaires are filled in, the system can use them to cluster users into groups that are likely to have similar musical tastes. Clustering users allows for recommendations that suit multiple people, which is useful in a variety of situations. Group exercise or group therapy sessions featuring music, for example, will ideally use music that all members of the group find acceptable.

Before similarity is calculated, each user is first assigned a vector based on their questionnaire responses. The elements of each user's vector are the six five-point ratings they listed for the loudness, energy, positiveness, tempi, familiarity, and rhythmic prominence of their preferred exercise music, plus the tapped tempo value if provided (otherwise that element is null), and a series of binary values for each of the artists, songs, and genre entries in the database. The ternary value for each artist, song, or genre is '0' for entries that the user did not list as a favorite, a positive weight for entries that the user listed as a favorite, and half that weight for artists that the user did not list as a favorite but which the MSD index says are 'similar to' an artist the user did list. The weights are set so that the artists, songs, and genres are all equal in weight to the other features.

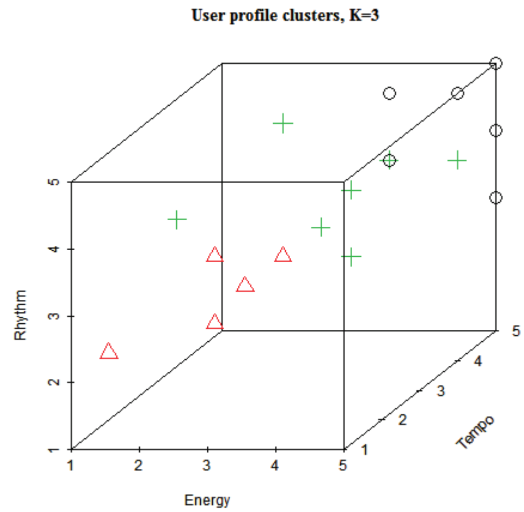

Figure 1. Three clusters of users, plotted on the 'Rhythm', 'Energy', and 'Tempo' axes.

The similarity between each pair of vectors is then calculated via their Euclidian distance, and the vectors are then clustered into groups via the K-Means algorithm (Figure 1). Finally, each group's profile is made by averaging and rounding the user's ratings for loudness, energy, happiness, tempo, popularity, and rhythmic prominence listed by that group's users, then aggregating the song, artist, and genres entries made by that group's users.

C. Initial Recommendations via UP

The user profiles, whether for an individual or a group, are then used to narrow the initial set of 384,000 tracks down to approximately 1,000 tracks according to features in the publicly-available MSD [15] (which is used in order that the proposed system be adoptable by others). This is done as follows:

- If any songs in the database have titles or artists that the user entered as favorites, those songs are likely to be included. Each artist in the database is also mapped to a genre according to the AllMusic.com genre structure, and songs whose artists are linked to genres that the user entered as a favorite are also more like to be included.
- If a user profile contains an exact tempo entered via the tapping procedure, then music with tempi near the tapped tempo are more likely to be retained. Otherwise, the five-point scale is first converted to five equally-spaced ranges between 50-100 beats per minute (BPM) and 150-200 BPM. Music within the appropriate range is more likely to be retained.

- The familiarity response is compared with the MSD's 'hottness' feature, on the basis that 'hotter' music is more likely to be familiar to a user than unpopular, or 'less hott' music.
- The rhythmic prominence response is checked against the 'beat confidence' feature from the MSD, on the basis that stronger beat confidence is often indicative of a strong, clear beat.
- The loudness, arousal, and valance values are not used for this filtering process. While MSD features exist for some of these elements, we did not find them to be reliable enough for our purposes.

D. Reinforcement Learning System

After the 1000 potential recommendations are obtained, the first set of 10 recommendations is selected at random for a user to evaluate. The user is provided with the song title, artist name, year of release, and a 30-second preview of the music obtained from 7Digital.com. The user then listens to the music and decides whether he thinks that music would be suitable for exercise. Then he can accept it or reject it. If he accepts it, that track is saved to his playlist. If he rejects it, however, more recommendations must be generated to replace the rejected track.

For our RL algorithm, we follow the non-greedy approach of Xing et al [9]. This approach models the problem as an *n-armed bandit* problem, which optimizes results over multiple iterations, instead of just a single iteration. The user's preferences for latent features found via collaborative filtering are determined based on the users' prior acceptances and rejections, and then the system makes selections from within the 1000-song subset found via UP. These selections are chosen to balance exploiting the user's known preferences with exploring the search space to allow for better recommendations in subsequent iterations. The new recommendations are then passed to the user, who can accept or reject them as before, and can also remove songs from their playlist if they determine that the newer songs are in fact more acceptable than the music they already have. This process continues until a user has accepted exactly 10 songs.

IV. Experiment and Results

Sixty college-age volunteers were divided into 3 groups of 20 students each for testing. One group utilized the proposed system, and its members received their initial set of music recommendations based on their individual questionnaires. The second group also tested the proposed system, but the members of this group were sorted into clusters and received their initial recommendations based on the aggregate questionnaire for their cluster. We empirically set the number of clusters to 3. Finally, the remaining group utilized a reference system with no UP system [9]. Students were not informed of their group.

All students filled out a user profile, then later used the music recommendation system while being monitored by the researchers in a lab environment. They received their initial recommendations, then rejected songs they felt were unsuitable and received new recommendations as per Section III-D. Students used the system until they had completed playlists of ten songs that they judged to be acceptable for use during exercise. At that point, they were asked to evaluate each song individually as well as the playlist as a whole on a five-point scale in terms of how suitable they were for exercise. They were also asked to indicate whether or not they would use the program in a real-world environment to select music for exercise. Overall results for the experiment are shown in Table I.

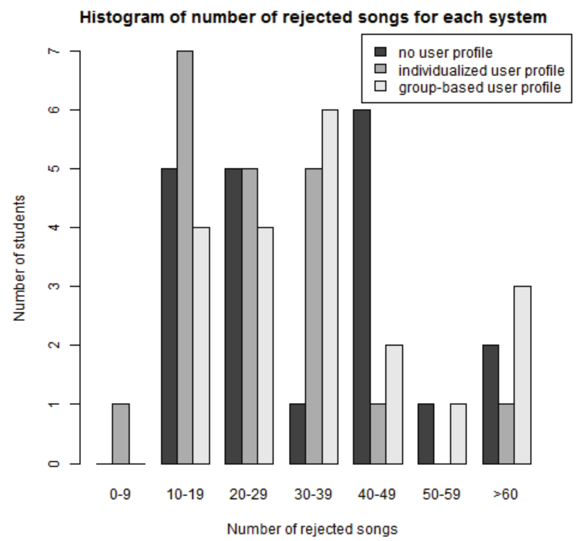

Figure 2. Histogram of the number of songs rejected by students in each group.

The percentage of users that said they would be willing to use each system in the future to recommend music for exercise is shown in Row 2 of Table I. The results show that both user-profile systems outperformed the reference system which did not use user profiles. Fully half of the students using the reference system did not want to use the system again, while three-quarters of the students using the other systems did. A t-test confirms that these results are statistically significant ($p<0.05$).

The average user of the individualized user profile-based system rejected about 27 unacceptable songs during the test, while the users of the reference system had to reject about 34 unacceptable tracks before obtaining 10 acceptable songs–25% more than for the proposed system. In fact, the reference system had far more students rejecting 40 or more songs than either proposed system (Figure 2). The group-based user profile system reported an average rejection of about 44 songs, but this was due to an outlier who rejected 294 songs, more than double the next highest value. Without this outlier, the average is reduced to about 31 songs, also less than for the reference system. Students using the user profile systems also removed fewer songs from their playlists after accepting them than users of the reference system. The user profile-based systems thus

	No user profile	Individualized user profile	Group-based user profile
Would use system again	50%	**75%**	**75%**
Average number of rejected songs	33.50	**26.50**	44.30/31.16
Average number of removed songs	1.35	**0.75**	**0.85**
Average rating (playlist)	3.45	**3.60**	3.50
Average rating (individual tracks)	3.80	3.83	3.81

Table I
RESULTS OF 60 STUDENTS TESTING THE RECOMMENDATION SYSTEMS. REJECTED SONGS WERE SONGS A USER SAW BUT DID NOT ADD TO THEIR PLAYLIST; REMOVED SONGS WERE ADDED TO THE PLAYLIST AT ONE POINT BUT LATER REMOVED. AVERAGE RATINGS ARE OUT OF 5 POINTS.

provided more acceptable music more quickly to users than the reference system. Though some improvement may be expected since the proposed system does incorporate the user's history of music preferences via the profile, this improvement nonetheless shows the benefits of our proposed approach.

The students' average ratings of their playlists and individual songs were also recorded. Even though students were told to continue until they found music that they found acceptable, the ratings still differed between the systems. While there was minimal improvement in the individual track ratings, the playlists as a whole were rated more highly for the UP systems. This indicates that the reference system caused the users to 'settle' for less acceptable music, likely due to lower quality recommendations.

V. CONCLUSION AND FUTURE WORK

The systems which used the UP algorithm outperformed the reference system. Students reported that they were more likely to use the new system, and they obtained acceptable recommendations without having to evaluate as many unacceptable ones. We also found that the system which used group-based user profiles produced comparable results to the system which used individualized profiles, validating the UP system for use in group environments like classes.

In the future, we will have our subjects exercise to the recommended music and monitor their physiological and psychological responses. This will allow us to better understand the effects that the recommended music has on people during exercise. We will also expand our results to other age groups. We are making arrangements with a local primary school for their students to use our system during physical education class.

REFERENCES

[1] F. Penedo and D. Hargreaves, "Exercise and well-being: A review of mental and physical health benefits associated with physical activity," *Current Opinion in Psychiatry*, vol. 18, no. 2, pp. 189–193, 2005.

[2] S. General, "Physical activity and public health: a report of the Surgeon General: executive summary," U. S. Public Health Service Commissioned Corps, Tech. Rep., 1996. [Online]. Available: http://www.cdc.gov/nccdphp/sgr/index.htm

[3] I.-M. Lee and R. Paffenbarger, "Physical activity and stroke incidence: the Harvard alumni health study," *Stroke*, vol. 29, pp. 2049–2054, 1008.

[4] M. Flores, "Uncovering the essence of music in exercise experiences," Master's thesis, California State University, Fresno, 2011.

[5] C. Karageorghis and D.-L. Priest, "Music in the exercise domain: a review and synthesis (part II)," *International Review of Sport and Exercise Psychology*, vol. 5, no. 1, pp. 67–84, 2012.

[6] J. Edworthy and H. Waring, "The effects of music tempo and loudness level on treadmill exercise," *Ergonomics*, vol. 49, no. 15, pp. 1597–1610, 2006.

[7] B. Copeland and B. Franks, "Effects of types and intesities of background music on treadmill endurance," *The Journal of Sports Medicine and Physical Fitness*, vol. 31, no. 1, pp. 100–103, 1991.

[8] A. Szabo *et al.*, "The effects of slow- and fast-rhythm classical music on progressive cycling to voluntary physical exhaustion," *Journal of Sports Medicine and Physical Fitness*, vol. 39, no. 3, pp. 220–225, 1999.

[9] Z. Xing *et al.*, "Enhancing collaborative filtering music recommendation by balancing exploration and exploitation," in *Proceedings of the International Symposium on Music Information Retrieval*, 2014, pp. 445–450.

[10] Y. Song *et al.*, "A survey of music recommendation systems and future perspectives," in *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval*, 2012, pp. 395–410.

[11] X. Wang *et al.*, "Context-aware mobile music recommendation for daily activities," in *Proceedings of the 20th ACM International Conference on Multimedia*, 2012, pp. 99–108.

[12] G. Wijnalda *et al.*, "A personalized music system for motivation in sport performance," *IEEE Pervasive Computing*, vol. 4, no. 3, pp. 26–23, 2005.

[13] G. Elliott and B. Tomlinson, "PersonalSoundtrack: context-aware playlists that adapt to user pace," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2006.

[14] R. de Oliveira and N. Oliver, "TripleBeat: enhancing exercise performance with persuasion," in *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services*, 2008, pp. 255–264.

[15] T. Bertin-Mahieux *et al.*, "The million song dataset," in *Proceedings of the 12th International Society for Music Information Retrieval*, 2011.

[16] M. Salokivi, "The individualized music therapy assessment profile as an initial assessment tool of social emotional functioning," Master's thesis, University of Jyvaskyla, Finland, 2012.