PRACTICAL EVALUATION OF REPEATED RECOMMENDATIONS IN PERSONALIZED MUSIC DISCOVERY

Brian Manolovitz
University of Miami
bmm157@miami.edu

Mitsunori Ogihara
University of Miami
m.ogihara@miami.edu

ABSTRACT

Studies have shown that repeated exposures to novel songs cause an increase in a person's memory and liking. These studies are commonly verified through self-reporting emotion-based surveys. This paper proposes the "retention rate" as an additional parameter for evaluation, which examines the rate at which the listener revisits the novel items. The authors hypothesize that when a person listens to novel (i.e., both unfamiliar and interesting) pieces of music, the retention rate will be proportional to the number of times the discovery engine suggests the pieces to her, as long as they remain novel. The authors have tested the hypothesis through a six-week human-subject experiment which simulates a real-world listening environment and a follow-up survey. During the experiment period, each subject received, through Discover Weekly in Spotify, suggestions for novel songs up to three times and provided evaluation. One month after the evaluation experiment, the human-subjects answered whether they had revisited the novel songs. Through the analysis of the response and survey data, the researchers conclude that the more times a listener is exposed to a song during the discovery process, the more likely she is to return to the song.

1. INTRODUCTION

The arrival of online music streaming services, such as Spotify[1] and Apple Music[2], has greatly changed the way people listen to music. They allow their users to make dynamic selections of music from vast libraries and thus provide an improved exposure outlet for musicians. By adopting a music streaming service, listeners are far more likely to broaden their tastes and explore songs and artists appearing in the long tail of the popularity distribution [3]. The instant accessibility to a myriad of songs through streaming platforms addresses the long-tail problem [13], where most listening data corresponds to few songs and the vast majority of songs have very little listening data, especially through use of collaborative filtering (CF). CF is a technique that finds a group of users whose tastes and activities show substantial similarity to each other and makes recommendations based upon what the other members of the group liked. The more people in the group who enjoy a piece, the more likely that the system recommends it, which gives rise to a recommendation bias towards popular songs, though other techniques can help correct this issue.

Personalized music discovery tools, such as Spotify's Discover Weekly playlist, aim at recommending music independent of user groups. They utilize the personal listening history of a user and try to suggest new and interesting songs specific to her interests. These "serendipitous recommendations" [9] are useful for extracting music from the long tail, which would otherwise be difficult for the user to find. They make use of content-based filtering techniques, which determine song similarity through the audio features and are a potential solution to the popularity contest that tends to be created by collaborative filtering methods.

However, even when presented with novel songs, the onus is on the user to remember to revisit them, usually by saving them or adding them to a playlist. The goal of the user during the experience of a new piece may not be to record what she liked for future relistening, but instead, she may want to listen to something in the background [4]. As a result, some songs may disappear not only from the memory of the user but also from her song collection, despite that the listener enjoyed them on first listen. These songs in oblivion create missed opportunities to expand both the listening repertoire of the user and the audience of the artist. The goal of this study is to show that when repeating recommendations during the music discovery process, the listeners are significantly more likely to revisit the discovered songs. This addresses the long-tail problem by focusing on data saturation, rather than item selection, in order to help items break out of the long tail.

Missed opportunities also arise when a user has a neutral or uncertain initial response to a song and discards it, since subsequent listens may have yielded a more favorable response. The music domain has the somewhat unique characteristic that users are expected to revisit songs many times. Studies have shown that repeated listens to a piece of music cause an initial increase in liking, which subsides with satiation. Simultaneously,

[1] www.spotify.com
[2] www.apple.com/apple-music

memory steadily increases throughout [7,11,14,15]. This is not an issue in the typical radio format for music discovery, where songs are presented many times upon their release, ensuring listeners will become familiar. Though streaming services have adopted the radio format to an extent [5], this study should provide justification for expanding its use in the music discovery tools.

In the present paper, the authors studied whether repeated recommendations of novel songs cause users to return to the songs later. A six-week human-subject experiment with a follow-up survey was conducted, which simulated a real-world listening environment. During the experiment period, the subjects were provided, through Discover Weekly in Spotify, suggestions for novel songs. Each novel song appeared no more than three times during the experiment. Each subject provided her response to each song on the list. One month after the conclusion of the six-week experiment, the human-subjects answered whether they had revisited the novel songs. Although it is typical to evaluate the effects of repeated recommendations with respect to a person's familiarity and preference to the songs, this study will introduce two additional factors, retention rate and forgetting rate, which allow the standard listening behaviors of the subjects to dictate the results. This paper will describe the details of the experiment and present the analysis of the data that was collected.

2. RELATED WORKS

When examining research on novelty and the effects of repeated recommendations, it is essential to trace everything back to [1], where Berlyne coined the inverted-U theory for collative variables. This theory postulates that as a collative variable (i.e., familiarity, complexity) increases, a person's liking increases to a point, then decreases, creating an inverted-U shape. Chmiel and Schubert [2] examined the validity of this theory throughout the past several decades of research. They found that, in general, the theory holds in the results of the studies they surveyed. It is, therefore, safe to assume that the first time a person listens to a song will not be their most enjoyed listening experience. Vargas and Castells [16] point out that flaws exist in evaluating novel recommendations solely based on the accuracy of the selected songs. They offer an alternative strategy that accounts for the ranking of the chosen items and their relevance to the user.

The problem of recommending long-tail items is a popular topic in the current recommendation systems research. Park [12] proposed an adaptive clustering method that clusters items based on their popularity but chose only the data objects in the long tail for clustering. This method performed better than prior approaches, both in terms of performance and the system's ability to recommend long-tail items. Wang et. al [17] utilized the users' experience level to control the extent to which long-tail items are recommended, finding that more experienced users were more open to the items in the long tail. [6] adapts the diversification of recommendation lists based on the perceived preferences of the user towards diversity and penalizes the inclusion of popular items while increasing accuracy. Using a multi-objective simulated annealing process, the resulting recommendation lists performed very well, compared to existing methods. Finally, [10] extended an existing tripartite graph approach for long-tail recommendations by expanding full genres, allowing more connections between items and genres. The results showed an increase in diversity and recall scores over existing methods.

Several studies exist to test the effects of repeated listens on liking and familiarity [7,11,14,15]. All have a result stating that both factors typically increase after the first listen. Hargreaves [7] found that when novel songs were repeated in weekly intervals, rather than in one continuous setting, familiarity ratings plateaued as in the inverted-U shape, but the "like" ratings remained constant in both cases. In [14], Szpunar et al. tested the effects of repeated listens during focused versus unfocused listening on memory and liking. They found that inattentive (or passive) listeners exhibited a slow and steady increase in liking and memory with repeated listens, whereas attentive listeners showed an inverted-U shape for both.

Similarly, in [11], Madison and Schiolde utilized a series of user listening experiments to conclude that familiarity increases liking regardless of the complexity of the music, and that familiarity is the most critical indicator of enjoyment. Van den Bosch et. al [15] conducted experiments involving psychophysiological scans that measured electrodermal activity. They connected self-reports on liking of music with these emotional responses. They found that as the unfamiliar songs were repeated, the emotional measurements became more closely related to the self-reported liking ratings.

Ward et al. [18] tested the effects of familiarity on music choice by asking subjects to rate songs in terms of familiarity and preference. After the initial rating, the songs were paired the songs up, one familiar and one unfamiliar, and asked the subjects which they would rather hear. They found that the subjects tended to select the songs with which they were more familiar, regardless of their liking ratings, leading them to conclude that people prefer to hear familiar music despite claiming they would like to hear more novel songs. Alternatively, their results indicate a greater need to boost familiarity during the discovery process, so that novel songs are more likely to be revisited.

3. PROBLEM OVERVIEW

Automatic personalized music discovery systems can make serendipitous recommendations to their users based on their listening histories. One example of these systems is Spotify's Discover Weekly, which provides users with a 30-song playlist of new and interesting songs based on their listening histories. Since the playlist is refreshed every Monday, the users will need to remember to revisit the songs they enjoy, usually by saving them or adding them to a playlist. If the users forget to do this, or if they prefer to listen to new music in the background [4], there will be missed opportunities to broaden the listener's repertoire, since songs the users enjoyed or may have

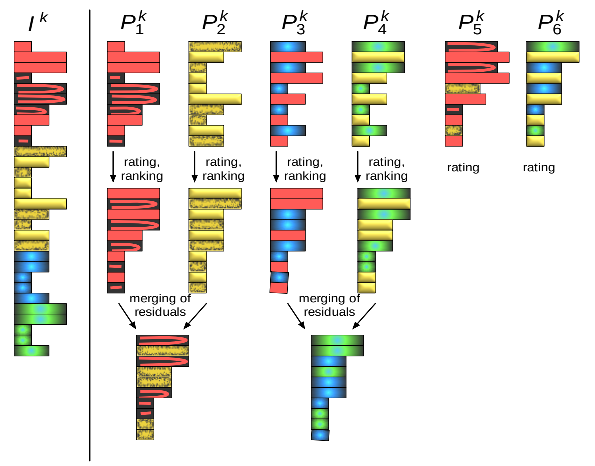

Figure 1. The playlist construction process. The boxes in red are the first ten songs of Discover Weekly, those in yellow are the next ten, those in blue are the next five, and those in green are the last five.

grown to enjoy never entered their listening rotation. We are proposing an improvement to the personalized music discovery approach, where the novel recommendations are repeated at least once to boost familiarity and increase the likelihood of the users revisiting the songs in the future. The first and main question of this research is as follows:

RQ1. Is the likelihood for a user to revisit a liked song proportional to the number of times that the song is presented to the user?

To answer this question, we introduce *retention rate*, R, as an evaluation metric. Let N be the set of songs a person listens to for the first time, and let r be the number of songs in N that person listens to more than once, $R = r / |N|$.

As shown in previous research [7,11,14,15], repeated listens of a novel song will typically lead to an initial increase in preference, followed by a decrease once the listener is satiated. The phenomenon is in line with the inverted-U theory for familiarity and preference [1,2]. Although the number of listens will need to be sufficiently large to see this pattern, we should still expect an increase in preference over the initial listening experiences. Therefore, the second research question is as follows:

RQ2. Does the preference toward the unfamiliar songs increase with repeated presentations?

Properly answering these questions will require a real-world listening simulation, where subjects are as free as possible to listen to the music and report on their behaviors. Since this approach utilizes actual listening activities for evaluation, it will yield more practical results than the usual feelings-based self-reports.

4. EXPERIMENT SETUP

To answer the two research questions, 19 Spotify users (15 female) were recruited to participate in a 6-week music listening study. This study was designed to simulate real-world listening behavior, allowing the subjects to listen freely and report on their activity. To begin, the subjects were each asked to provide their most recent 30-song Discover Weekly playlist (570 total songs) without first listening to it. Let I^k represent this playlist for user k, such that I_n^k represents the n^{th} song from the initial discovery playlist for user k. Every week, on Monday, subject k is provided with a link to a 10-song playlist, P_w^k, where w is the current week number. They were asked to listen to the playlist once, then fill out a survey before any further listens. No restrictions or requirements were placed on the setting or device of the listening. The only suggestion given to the subjects was to treat the listening as they would their normal music discovery. The survey asked the subjects to evaluate the entire playlist in terms of both their enjoyment and its effectiveness for discovery, both on a 5-point Likert scale. Additionally, they were asked to place each of the individual songs into one of three categories: "Like it," "Not sure yet/Neutral," or "Don't like it."

4.1 Playlist Construction

Let us describe in detail the playlist construction process (see Figure 1). Since the duration of the experiment is six weeks and we make weekly playlists, there are six playlists for subject k: P_1^k, …, P_6^k. The six playlists contain ten songs each. Recall that the subject k disclosed the 30 songs in her Discover Weekly list, I^k, without listening to any of the songs in it. appearing in the Discovery Weekly. Since the selection in the Discover Weekly playlists ensures that the user has never listened to the song before using Spotify, we generated the playlists under the assumption that each subject had not heard any of the 30 songs in her list before. We constructed the six ten-song lists dynamically as follows:

- P_1^k and P_2^k contained songs $[I_1^k : I_{10}^k]$ and $[I_{11}^k : I_{20}^k]$, respectively, in the same order they appeared in I.
- P_3^k and P_4^k contained songs $[I_{21}^k : I_{25}^k]$ and $[I_{26}^k : I_{30}^k]$, respectively, as well as five songs repeated from P_1^k and P_2^k, respectively. It is necessary to control for the subjects' song ratings when deciding which songs to repeat. Practical implementations would simply repeat the songs the users seemed to enjoy most (saved, liked, did not skip, etc.), but the retention rate must be independent of preference in this experiment; otherwise it would be impossible to claim that it is correlated with repetition. We chose the five repeats in P_3^k and those in P_4^k to preserve the proportion of the initial ratings (Like = Not sure/Neutral = Dislike). These ratings are unknown ahead of time due to the nature of the experiment, and so we used the subject's ratings from weeks 1 and 2 for balancing the ratings. We first sorted the ten songs in P_1^k and the ten in P_2^k in the decreasing order of rating (in the case of ties, the order from I^k was preserved). We then assigned the songs at odd numbered positions in the ranking (i.e., 1, 3, 5, 7, and 9) from the first list to P_3^k. Similarly, we assigned the five songs from the second list at odd numbered positions to P_4^k. After determining the ten-song sets from which to build P_3^k and P_4^k, we fixed the order in

which the ten songs appeared. We increased the perception of diversity by placing the repeated songs and new songs alternatingly in these playlists, following the idea from [8].

- In P_5^k and P_6^k, we kept the repeats from P_1^k and P_2^k, that we used in P_3^k and P_4^k, respectively. We selected the remaining songs as follows:
  - For P_5^k, the five remaining songs are those from P_1^k and P_2^k that we did not use in P_3^k and P_4^k (i.e., those with even numbered ranks after sorting in the decreasing order of rating). We merged the two ranked lists of five remaining songs and selected five at those at odd numbered positions in the merged list.
  - P_6^k is constructed in the same manner, but the source of the five repeats are the five non-repeats each from P_3^k and P_4^k. Again, we sorted the ten songs in the decreasing order of rating and then selected those at odd numbered positions.
- In the end, 10 songs were presented 3 times, 10 were presented twice, and the remaining 10 were only presented once and were used as a baseline.

4.2 Surveys and Evaluation Process

After the 6-week music listening portion of the experiment, the subjects were asked to complete a survey containing general questions about their music listening habits, as well as a final evaluation of the songs in *I*, which was simply a 5-point Likert scale rating of their likelihood to revisit the songs in the future. They were asked to fill out this portion without listening to the songs again, relying solely on their memory of the song based on name and artist. If the subject could not remember the song, they were asked to respond with an asterisk instead of a rating. Since the weekly surveys asked the subjects to place the song name and artist into one of the three fields, they were required to think critically about the song, and therefore, should be expected to remember this information. The analysis will refer to the *forgetting rate*, which refers to the percentage of forgotten songs with respect to some characteristic, such as initial rating or the number of presentations of the song. Finally, if the subjects had heard the song prior to the experiment, they were asked to leave the field blank, and that song would be omitted from the results.

One month after the end of the experiment, the subjects received a follow-up survey asking them to indicate whether they had chosen to listen to each song in *I* during the month since the experiment. We evaluate the *retention rate*, *R*, with respect to the number of times the songs were presented, R^x as well as with respect to the subjects' ratings of the songs.

5. RESULTS AND DISCUSSION

The study utilized Spotify's Discovery Weekly playlist to provide serendipitous recommendations to the subjects based on their listening histories on the system, which meant some of the songs were not entirely new to the subjects. This is because Spotify is not the only music listening platform the subjects use, and we rectify it by omitting previously heard songs from the results analysis (26 out of 570 – less than 0.5%). It is also worth noting that the limited number of subjects restricts the generalizability of these results, but this being a multi-week study made finding subjects difficult.

The weekly preference and discovery ratings were grouped by their composition as mentioned in Section 4, and their results are shown in Figure 2. The preference ratings were basically constant across each group, with means as follows: weeks 1-2 = 3.57, weeks 3-4 = 3.75, and weeks 5-6 = 3.67. The discovery ratings unsurprisingly declined in weeks where fewer new songs were presented, with means as follows: weeks 1-2 = 3.88, weeks 3-4 = 3.12, and weeks 5-6 = 2.41. It is worth noting, however, that an even split between new and repeated songs is likely not ideal in a real-world setting, but determining the optimal split was outside of the scope of this research.

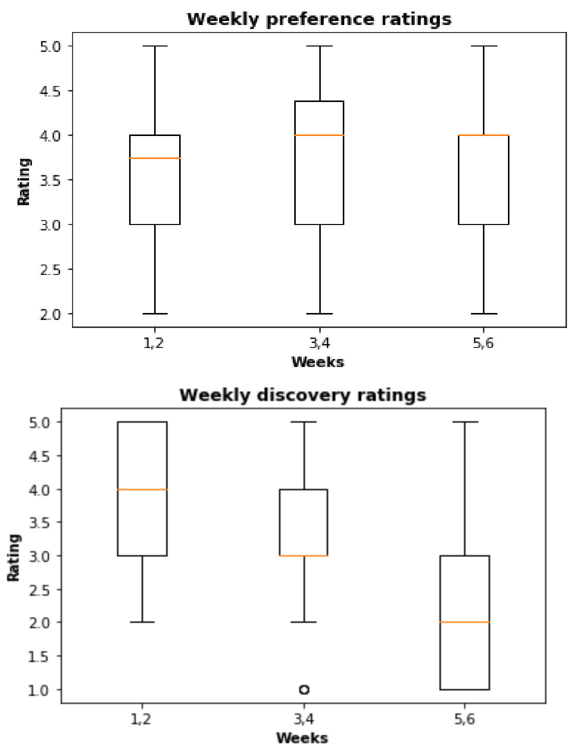

Figure 2. Boxplots of the weekly preference (top) and discovery (bottom) ratings for all subjects, grouped based on the composition of the playlists (unfamiliar vs. familiar).

Figure 3 shows the number of songs whose rating changed, either positively or negatively, after the first listen, as well as the counts of the final ratings for "neutral" songs with more than one presentation. Though it was most common for the rating to remain the same, when it changed, it increased almost twice as often as it decreased. Of all songs with more than one play, 43 of them showed a decrease in rating and 73 showed an increase, though 263 retained their initial rating. Isolating the 116 songs with an initial rating of "neutral," we see that 43 of them (37%) had a final rating of "like," versus 20 (17%) whose rating decreased to "dislike" and 53 (46%) that remained at "neutral." Therefore, there is a potential for missed opportunities with a “neutral” song, since the initial listen may not entice the listener to revisit the song, whereas subsequent listens are likely to yield a more favorable response. We can answer RQ2 by saying that it is more likely for a person’s preference toward an unfamiliar song will increase with repeated listens, but most of the time their feelings will remain constant. A rating scale with finer granularity would have answered this question more effectively.

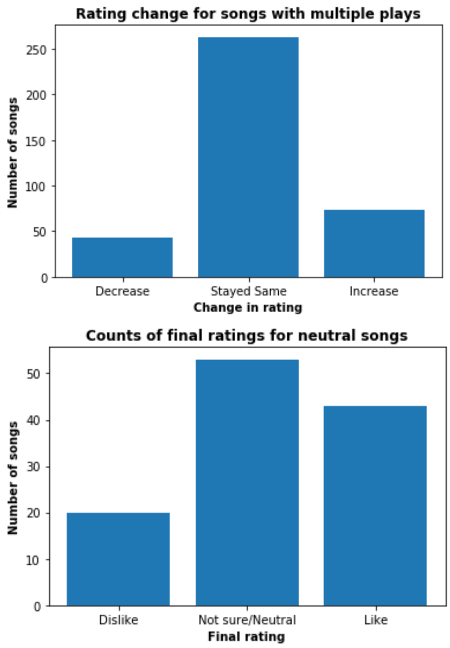

Figure 3. The rating changes of all songs (top) and neutral songs (bottom) with more than one presentation during the study.

Figure 4 illustrates the forgetting rate with respect to play count and initial rating. It also takes an isolated look at the songs with an initial “like” rating. In total, 105 of the 544 songs (19.3%) were forgotten. Two one-tailed paired samples z-tests were performed, comparing the forgetting rate between songs with 1 and 2 plays, as well as 2 and 3 plays, $z(343) = 5.88$, $p < 0.00001$ and $z(342) = 2.38$, $p < 0.01$, respectively. The forgetting rate decreases significantly as the number of plays increases, which should be reciprocated with a higher retention rate, though any subjects who saved the songs for later would be less reliant on their memory. In terms of liking, two additional one-tailed paired samples z-tests were conducted to compare the difference between the forgetting rates for songs with an initial rating of “dislike” versus “neutral,” $z(250) = 1.38$, $p < 0.1$ and “neutral” versus “like,” $z(423) = 0.44$, $p < 0.33$. In both cases, we found no significant differences on the forgetting rates, which indicates a lack of connection between preference and memory. Looking at the forgetting rate of the liked songs, we can see another potential for missed opportunities, as the songs with only one presentation are significantly more likely to be forgotten, $z(182) = 4.21$, $p < 0.00002$ for 1 versus 2 plays and $z(177) = 5.94$, $p < 0.00001$ for 2 versus 3 plays. These missed opportunities occur when a person enjoys a song but does not save it and forgets its name or the artist name and is unable to search for it later.

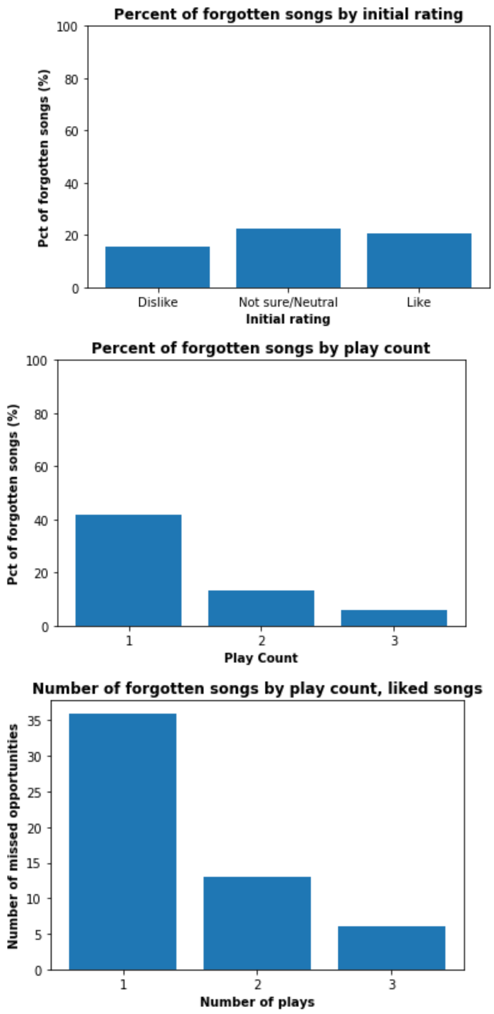

Figure 4. The percentage of forgotten songs by initial rating (top) and play count (middle), as well as the number of forgotten “liked” songs by play count (bottom)

One month after the conclusion of the listening experiment, the subjects received a survey asking them to state whether they listened to each of the songs from the experiment on their own accord (i.e. in a playlist they created or by searching for the song). Of the 19 subjects, 18 of them provided responses, and only one of those did not revisit any songs. Figure 5 shows the percentage of songs which were revisited with respect to both their play counts as well as the initial ratings. Figure 6 groups the songs by play count and shows the percentage of retained songs with respect to both initial and final rating.

A series of paired samples one-tailed z-tests were performed to answer RQ1. First, the songs were grouped by the number of times they were presented, requiring a test comparing R^1 and R^2 and another comparing R^2 and R^3, $z(343) = 2.41$, $p < 0.01$ and $z(342) = 1.22$, $p < 0.2$, respectively. Clearly, there is no significant difference between the retention rate of songs with 2 and 3 plays.

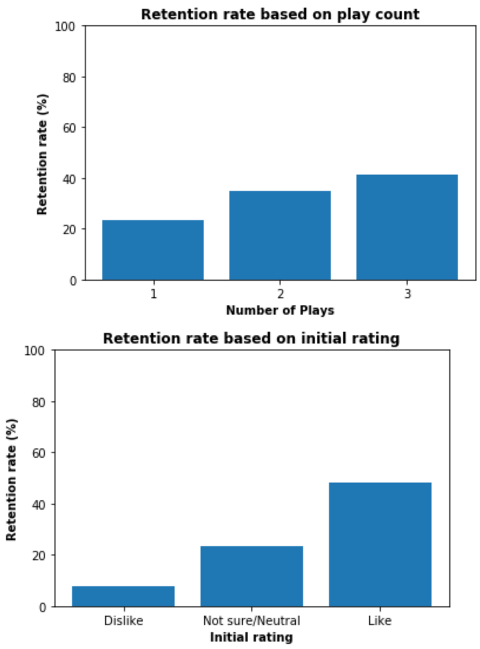

Figure 5. The retention rate with respect to play count (top), initial rating (bottom).

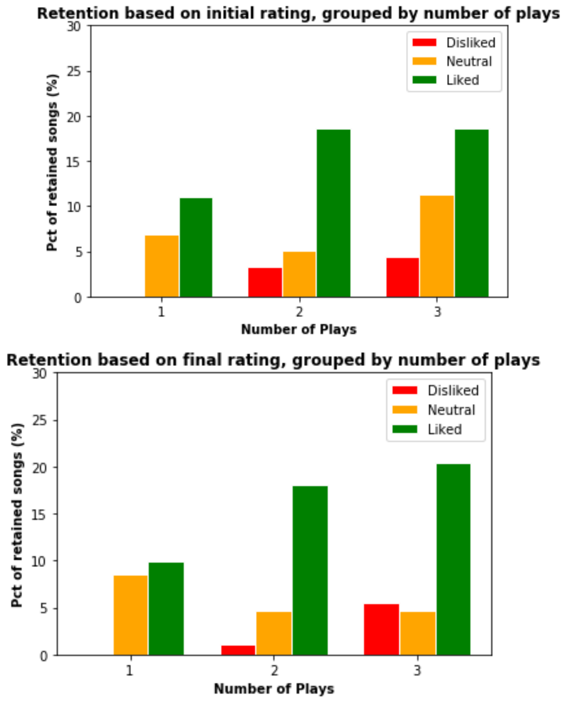

Figure 6. The number of retained songs, grouped by number of plays, with respect to initial (top) and final (bottom) ratings.

However, if we assume that all additional plays have the same effect on retention, we can combine the songs with 2 and 3 plays. Comparing the retention rate of these songs with R^1, we get $z(514) = 3.42$, $p < 0.0005$. Therefore, we can answer RQ1 affirmatively on the premise that additional presentations of novel songs beyond the first are essentially the same. It is worth noting that this experiment did not test more than 3 presentations of a novel song, and it is likely there are diminishing returns beyond the third presentation. We conducted additional z-tests by grouping the songs by their initial rating, then comparing the songs with 1 play and the songs with more than 1 play, as follows: for songs rated "dislike," $z(91) = 1.98$, $p < 0.025$; rating = "neutral," $z(159) = 0.62$, $p < 0.3$; rating = "like," $z(264) = 3.37$, $p < 0.0004$. The performance of the "neutral" songs is interesting and could be related to the change in rating for these songs with more than one play. In general, only the "liked" songs had a reasonably high retention rate, so it is possible that a neutral or worse feeling towards a song is insufficient to persuade a person to revisit it. As previously seen, the "neutral" songs increased in rating at a 37% rate, and when looking at the graph in Figure 6 which shows final ratings, the "liked" songs still show an increase with play count, but the remaining neutral songs do not.

We performed similar z-tests to evaluate the relationship between the initial rating of the songs and their retention, one for "dislike" vs "neutral/not sure" and another for "neutral/not sure" vs "like," $z(250) = 3.11$, $p < 0.001$ and $z(423) = 5.08$, $p < 0.00001$, respectively. Clearly, initial rating is a strong predictor of the retention of a song, though this should be obvious and does not deter from the results with respect to play count. In practice, a personalized discovery system can infer liking via user interaction (i.e. button clicks, skips, page visits, etc.), then use that to select which songs to repeat.

6. CONCLUSION AND FUTURE WORK

Previous research has concluded that repeated listens of novel music will increase both memory and liking, but the evaluation has typically involved the subjects self-reporting on their feelings. This study implemented a real-world listening simulation and evaluated the effects of repeated listens of novel songs with respect to the rate at which the songs were revisited by the subjects. We found that when songs were played more than once, in general, their retention rate significantly increased, and the rate at which the songs could be recalled from name and artist alone also increased. Additionally, if the ratings of the songs changed after the first listen, it was significantly more likely to be an increase.

We explored the concept of missed opportunities when assuming a music discovery process which recommends songs to users once and expects them to remember to revisit the songs. By only presenting songs once, liked songs are less likely to be remembered or revisited, and songs users feel neutral or unsure about will not have a chance to improve their favorability. Repetition of novel recommendations clearly decreases the potential for missed opportunities in both cases, giving users a greater chance to broaden their musical tastes.

One aspect we did not evaluate was the proper split between new and repeated songs, which may be a user-specific parameter and likely varies from week to week. In our future work, we intend to explore whether there is a predictable pattern to the amount of new music a person consumes on a weekly basis. In addition, we are planning several studies involving electroencephalograpy, where we will test memory and attention when listening to new music over a sustained period.

7. REFERENCES

[1] D. E. Berlyne: *Aesthetics and psychobiology*, Appleton-Century-Crofts, New York, NY, 1971.

[2] A. Chmiel and E. Schubert: “Back to the inverted-U for music preference: A review of the literature,” *Psychology of Music*, Vol. 45, No. 6, pp. 886-909, 2017.

[3] H. Datta, G. Knox, and B. J. Bronnenberg: “Changing their tune: How consumers’ adoption of online streaming affects music consumption and discovery,” *Marketing Science*, Vol. 37, No. 1, pp. 5-21, 2017.

[4] J. Garcia-Gathright, B. St Thomas, C. Hosey, Z. Nazari, and F. Diaz: “Understanding and evaluating user satisfaction with music discovery,” *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 55-64, 2018.

[5] M. Glantz: “Internet radio adopts a human touch: a study of 12 streaming music services,” *Journal of Radio & Audio Media*, Vol. 23, No. 1, pp. 36-49, 2016.

[6] E. M. Hamedani and M. Kaedi: “Recommending the long tail items through personalized diversification,” *Knowledge-Based Systems*, Vol. 164, pp. 348-357, 2019.

[7] D. J. Hargreaves: “The Effects of Repetition on Liking for Music,” *Journal of Research in Music Education*, Vol. 32, No. 1, pp. 35-47, 1984.

[8] M. Ö. Karakaya and T. Aytekin: “Effective methods for increasing aggregate diversity in recommender systems,” *Knowledge and Information Systems*, Vol. 56, No. 2, pp. 355-372, 2018.

[9] D. Kotkov, S. Wang, and J. Veijalainen: “A survey of serendipity in recommender systems,” *Knowledge-Based Systems*, Vol. 111, pp. 180-192, 2016.

[10] A. Luke, J. Johnson, and Y. K. Ng: “Recommending Long-Tail Items Using Extended Tripartite Graphs,” *IEEE International Conference on Big Knowledge (ICBK)*, pp. 123-130, 2018.

[11] G. Madison, and G. Schiolde: “Repeated Listening Increases the Liking for Music Regardless of Its Complexity: Implications for the Appreciation and Aesthetics of Music,” *Frontiers in Neuroscience*, Vol. 11, Article 147, 2017.

[12] Y. J. Park: “The adaptive clustering method for the long tail problem of recommender systems,” *IEEE Transactions on Knowledge and Data Engineering*, Vol. 25, No. 8, pp. 1904-1915, 2012.

[13] Y. J. Park and A. Tuzhilin: “The long tail of recommender systems and how to leverage it,” *Proceedings of the 2008 ACM Conference on Recommender Systems*, pp. 11-18, 2008.

[14] K. K. Szpunar, G. E. Schellenberg, and P. Pliner: “Liking and Memory for Musical Stimuli as a Function of Exposure,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 30, No. 2, pp. 370-381, 2004.

[15] I. N. Van den Bosch, V. J. Salimpoor, and R. Zatorre: “Familiarity mediates the relationship between emotional arousal and pleasure during music listening,” *Frontiers in Human Neuroscience*, Vol. 7, Article 534, 2013.

[16] S. Vargas and P. Castells: “Rank and relevance in novelty and diversity metrics for recommender systems” *Proceedings of the fifth ACM Conference on Recommender Systems*, pp. 109-116, 2011.

[17] Y. Wang, J. Wang, and L. Li: “Enhancing Long Tail Recommendation Based on User's Experience Evolution,” *IEEE 22nd International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pp. 25-30, 2018.

[18] M. Ward, K. Goodman, and J. Irwin: “The same old song: The power of familiarity in music choice,” *Marketing Letters*, Vol. 25, No. 1, pp. 1-11, 2014.