How Automated Recommendations Affect the Playlist Creation Behavior of Users

Iman Kamehkhosh
TU Dortmund, Germany
iman.kamehkhosh@tu-dortmund.de

Dietmar Jannach
AAU Klagenfurt,
Klagenfurt, Austria
dietmar.jannach@aau.at

Geoffray Bonnin
LORIA
Nancy, France
bonnin@loria.fr

ABSTRACT
Modern music platforms like Spotify support users to create new playlists through interactive tools. Given an empty or initial playlist, these tools often recommend additional songs, which could be included in the playlist based, e.g., on the title of the playlist or the set of tracks that are already in the playlist. In this work, we analyze in which ways the recommendations of such playlist construction support tools influence the behavior of users and the characteristics of the resulting playlists. We report the results of a between-subjects user study involving 123 subjects. Our analysis shows that users provided with recommendation support were more engaged and explored more alternatives than the control group. Presumably influenced by the recommender, they also picked significantly less popular items, which leads to a higher potential for discovery. The effort required to browse the additional alternatives, however, increased the users' perceived difficulty of the process.

ACM Classification Keywords
Information Retrieval : Recommender Systems; Human Computer Interaction (HCI) : User studies

Author Keywords
Music Recommendation; Playlist Creation; User Study

INTRODUCTION
Creating and sharing playlists is a common feature of most of today's music platforms. The manual construction of playlists by users can however be a comparably complicated and time consuming task [6]. One way to help playlist creators is to provide them with automated suggestions for additional tracks while they are creating a playlist. Such a functionality can be found on some of today's music platforms, like Spotify and Pandora. In the literature, a number of algorithms were proposed in the past to determine a set of suitable tracks given, e.g., a partial playlist [5]. The evaluation of such algorithms is mostly based on offline experiments with a focus on prediction accuracy, i.e., on the algorithms' capability of predicting the next tracks that were picked by users.

MILC '18, March 11, 2018, Tokyo, Japan

While accurate track relevance predictions are important, such experiments cannot inform us about whether users will actually adopt the recommendation functionality and how the recommendations influence their behavior. With this work, we aim to obtain a better understanding of these aspects which, to our knowledge, have not been studied in that form in the literature before. We conducted a between-subjects user study involving 123 subjects, where the participants' task was to create playlists for a given topic using a web application that was developed for the study. One group of participants was provided with additional recommendation functionality, whereas the control group could only rely on the provided search functionality. A post-task questionnaire was used to assess additional aspects like the perceived difficulty of the task.

Our analyses revealed that the recommendation support was well accepted by the participants and that recommendations therefore represent a valuable tool for users. Almost half of the users that were provided with recommendations picked at least one of the recommended tracks, which is an unusually high proportion for the domain of recommender systems.

Furthermore, we could observe that the selection of tracks was seemingly influenced by the recommendations even if no track was actually included in the playlists, i.e., the recommendations served as inspirations for the participants. Participants with recommendation support also significantly explored more tracks within the same period of time, and they more often chose less popular tracks based on the suggestions by the recommender. As users are more engaged and explore more options from the long tail, the recommender therefore measurably increases the potential of discovering new tracks or artists. Also, this increased user engagement can help providers to collect more information about the users' preferences.

Finally, the post-task questionnaire, to some surprise, revealed that the participants found the playlist creation task to be more difficult when recommendations were provided. This was the case even when they actually did not pick any of the tracks. This observation suggests that the user interface (UI) has to be carefully designed, since parts of this might be caused by the increased complexity of the web application that included a recommendation component.

PREVIOUS USER STUDIES
Compared to the number of papers that report results of offline experiments, the number of user studies is limited. Many of these previous user studies in the music domain focus on how

users search for music and on the social or contextual aspects of listening to music [6, 7, 8, 19, 20]. In [6], for example, interviews and web forum posts related to the construction of playlists were analyzed. The authors found that playlists (mixes) are often created with a theme or topic in mind, e.g., a genre, an event or a mood. In our study, we therefore also ask participants to create a playlist for one of several given topics. According to the suggestion in [6], our online application does not *automatically* include additional tracks, but presents recommendations as a side information.

Questions of the UI design were also in the focus in [3] and [24]. For instance, [3] proposed an interactive track recommendation service called *rush*, optimized for touchscreen devices and, among other aspects, analyzed its usability for left-handed and right-handed users. In the application we used for our study, recommendations were positioned as a horizontal list on the bottom of the screen, which is common also for e-commerce sites like Amazon.com.

A few studies explore the users' interactions with music recommender systems [1, 16] and the quality perception of music recommendations [2, 17]. The recent work of [17] provided evidence that users prefer recommendations that are coherent with the recently played tracks in different dimensions. Their results also indicated that the participants tend to evaluate recommendations better when they know the track or the artist. In contrast to [17], we do not compare different recommendation algorithms but the effects of the existence of a recommender.

Finally, questions related to the factors that influence which tracks users choose for inclusion in a playlist in different situations were analyzed in [15, 23] and [25]. The study in [15], for example, showed that mood, genre, and artists are the most important factors for users when selecting the tracks, which is in line with the outcomes of the study of [6]. Similar to their work, we explicitly asked users about their decision factors after the task and report the results in Section 5.

STUDY DESIGN

A general limitation of laboratory studies is that users might behave differently in a "simulated" situation than when they normally listen to music, e.g., at home. To alleviate this problem, we provided an online application to enable users to participate in the study when and where they wanted to.

All participants were asked to create a playlist – using the developed application – for one of the following pre-defined and randomly ordered themes: *rock night*, *road trip*, *chill out*, *dance party*, *hip-hop club*.[1] The participants were randomly assigned to one of two groups. One group (called `Rec`) received additional recommendations, as shown at the bottom of Figure 1. The control group (`NoRec`) was shown the same interface but without the recommendation bar at the bottom.

All participants could use the provided search functionality. When a user of the `Rec` group added an item to the playlist, the

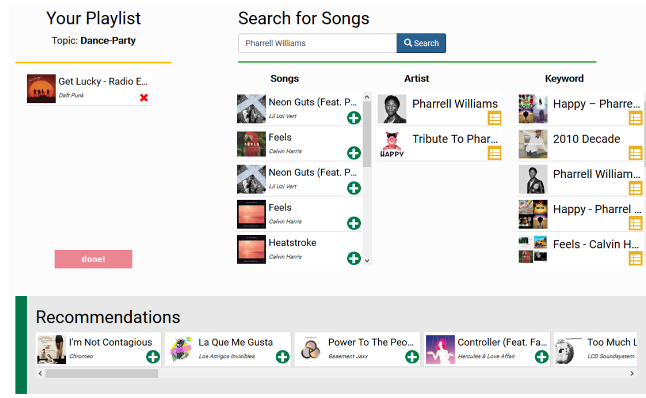

Figure 1. Web application used in the study.

20 provided recommendations were immediately updated.[2] When the playlist contained at least six tracks, the participants could proceed to the post-task questionnaire.

The first part of the questionnaire was presented to all participants and contained a list of quality factors for playlists mentioned in the literature, which were either related to individual tracks (e.g., popularity or freshness) or to the list as a whole (e.g., artist homogeneity). The participants were asked to rank these quality factors or mark them as irrelevant.

In the next step, participants of the `Rec` group were asked if they had looked at the recommendations during the task and if so, how they assessed their quality in the following dimensions: *relevance*, *novelty*, *accuracy*, *diversity* (in terms of genre and artist), *familiarity*, *popularity*, and *freshness*. Participants could express their agreement with our provided statements, e.g., "The recommendations were novel", on a 7-point Likert scale item or state that they could not tell.

In the final step, all participants were asked (a) how often they create playlists, (b) about their musical expertise, and (c) how difficult they found the playlist creation task, again using 7-point Likert scale items. Free text form fields were provided for users to specify which part of the process they considered the most difficult one and for general comments and feedback. Finally, the participants could specify their age group.

SPOTIFY'S RECOMMENDATION ALGORITHMS

Both the search and the recommendation functionality in the study were implemented using the public API of Spotify.[3] This allowed us to rely on industry-strength search and recommendation technology. As of 2017, Spotify is a leader across most streaming-service markets[4] and the recommendations produced by the service result from several years of A/B testing [21].

Companies do not usually reveal the details of the algorithms that they use. According to the documentation of the Web API of Spotify, recommendations are aimed to create a *playlist-style* listening experience based on seed artists, tracks and

[1] Note that such themes actually also convey an intended use or purpose. The best example is the "road trip" theme, whose corresponding playlists are supposed to be played when driving a car for a prolonged amount of time.

[2] Recommendations were displayed after the first track was added.

[3] https://developer.spotify.com/web-api/

[4] http://www.midiaresearch.com/downloads/midia-streaming-music-metrics-bundle/

genres. In this context, the available information for a given seed item is used to find similar artists and tracks.

Some additional information can be obtained from what the company presents in public presentations and industry reports. Public presentations around the year 2014, such as [13] and [14] indicate that Spotify relied at that time mainly on collaborative filtering techniques for generating recommendations. The latest presentation was made right after Spotify acquired The Echo Nest, a music intelligence platform that focused on the analysis of audio content, and Spotify announced they were going to also utilize content-based techniques. In more recent presentations, such as [22], the authors report that Spotify uses an ensemble of different techniques including NLP models and Recurrent Neural Networks as well as explicit feedback signals (e.g., thumbs-up / thumbs-down), and also audio features for certain recommendation tasks, but it is unclear from the presentations which techniques are used for which types of recommendations (radios, weekly recommendations, playlists, etc.).

Table 1. Description of the collected information for the tracks, as provided by Spotify: https://developer.spotify.com/web-api/get-audio-features/.

Information	Description
Acousticness	Absence of electrical modifications in a track.
Danceability	Suitability of a track for dancing, based on various information including the beat strength, tempo, and the stability of the rhythm.
Energy	Intensity released throughout a track, based on various information including the loudness and segment durations.
Instrumentalness	Absence of vocal content in a track.
Liveness	Presence of an audience in the recording.
Loudness	Overall loudness of a track in decibels (dB).
Popularity	Popularity of a track, based on the its total number of plays and the recency of those plays.
Release year	Year of release of a track.
Speechiness	Presence of spoken words in a track.
Tempo	Speed of a track estimated in beats per minute (BPM).
Valence	Musical positiveness conveyed by a track.

RESULTS AND OBSERVATIONS

General Statistics

Participants. Most of the 123 participants who completed the study were students of universities in Germany and Brazil; a smaller part was recruited via invitations on social network sites. Most (84%) of the participants were aged between 20 and 40. On a scale between 1 and 7, the median of the self-reported experience with music was 5, i.e., the majority of the participants considered themselves experienced or interested in music. Most of the participants, however, do not create playlists regularly; about 25% answered the question with a 5 or greater (median=3). The median of the perceived difficulty of the playlist creation task was 4, i.e., the majority of the subjects found the task comparably difficult.[5]

Topics and playlist length. *Rock night* and *road trip* were the most often selected themes. Each of them was selected in about 30% of all trials. *Chill out* (20%), *dance party* (15%) and *hip-hop club* (5%) were less frequently chosen. The average time for the participants to create one playlist (with at least 6 tracks) was 7.29 minutes and a created playlists contained, on average, 8.44 tracks.

Recommendation use. 57% of the participants were assigned to the `Rec` group (with recommendations). Almost half of these participants (49%) drag-and-dropped at least one of the recommended tracks to their playlists. We denote this group as `RecUsed`. The other half will be denoted as `RecNotUsed`.

Study Outcomes

Impact of Recommendations on Users and their Behavior

Adoption of recommendations. When users actively used the recommendations, they relied on them to a significant extent. At the end, about 38% of the tracks of the playlists that were created by users of the `RecUsed` group were taken from the recommendations (mean=3.2 recommendations per playlist). This is a strong indicator of the general usefulness of a recommendation component in that domain, considering that in general e-commerce settings sometimes only every 100th click of a user is on a recommendation list and recommendations are used only in about 8% of the shopping sessions [12].

Increased exploration. The participants that received recommendations played significantly[6] more tracks when creating their playlist than the participants of the `NoRec` group (mean value 14.4 and 9.8, respectively). This value is even higher for the `RecUsed` subgroup, i.e., those who actively used the recommendations, with an average of 20.3. However, the participants of the `Rec` group, on average, only needed 30 seconds more to create the playlists[7], i.e., the recommender helped them explore, and potentially discover, many more options in about the same time.

Such a high exploration is interesting as it not only increases the chances of music discovery but also allows the service provider to gather more information about users' preferences. This can in turn lead to providing better recommendations and a better listening experience for users, as well as higher customer retention and business value for the provider [10].

Difficulty of playlist creation. To some surprise, the recommendation component did not make the playlist creation task easier for users but slightly added to the perceived complexity, with a median of 4 (higher complexity) for the `Rec` group and 3 for the `NoRec` group. This slight but not statistically significant difference could be caused by the more complex recommendation UI, as well as by the fact that users explored more options in the `Rec` condition as mentioned above (see, for example, [4] for a discussion on "choice overload" in recommender systems).

[5] The collected data is ordinal, i.e., a ranking of the response levels is possible. However, we cannot assume equidistance between the response levels, and reporting mean and standard deviation values is often considered questionable in the literature.

[6] To test for statistical significance for the ordinal data, we use the Mann-Whitney U test and for the interval data we use the Student's t-test, both with $p < 0.05$.

[7] The differences were not statistically significant.

Table 2. **Average (Avg) and standard deviation (Std) of the musical features of the resulting playlists in different groups. * indicates statistical significance in comparison with the RecUsed group.**

	RecUsed (34 participants)		RecNotUsed (36 participants)		NoRec (53 participants)	
Feature	Avg	Std	Avg	Std	Avg	Std
Acousticness	0.22	0.28	0.17*	0.23	0.17*	0.26
Danceability	0.56	0.18	0.59*	0.17	0.54	0.17
Energy	0.68	0.24	0.70	0.19	0.73*	0.23
Instrumentalness	0.16	0.32	0.12	0.28	0.12	0.27
Liveness	0.20	0.17	0.21	0.18	0.21	0.17
Loudness (dB)	-7.68	4.59	-7.60	3.67	-7.52	4.72
Popularity	50.7	21.9	55.7*	17.1	54.3*	21.3
Release year	2005	12.47	2002*	15.73	2003*	13.08
Speechiness	0.07	0.07	0.08	0.08	0.08	0.08
Tempo (BPM)	123.0	28.7	122.5	27.9	122.9	28.6
Valence	0.50	0.26	0.53	0.25	0.49	0.24

Impact of Recommendations on the Resulting Playlists

To analyze the impact of the provided recommendations on the resulting playlists, we queried the musical features of the tracks contained in the created playlists through Spotify's API. Table 1 shows a list of these features. The average values and standard deviations of the features for each of the study groups are shown in Table 2. Several differences can be observed. We limit the discussion here to the most pronounced and statistically significant effects, which can be found with respect to the popularity and the freshness of the tracks that were used by the participants.

Popularity effect – promoting less popular items.[8] Users of the recommendation service added significantly less popular tracks to their playlists. This is in line with the observations from [11] where the recommendations by a commercial service were less popular (in terms of play-counts) than the tracks that users selected manually.

Recency effect – promoting newer tracks. Using recommendations also slightly but statistically significantly increased the freshness (release year) of the selected tracks. About 50% of the tracks of the playlists created by the participants who used the recommendations (RecUsed) were released in the last five years. This value is 40% for the RecNotUsed group and 34% for the NoRec group.

"Mere-Presence" Effect of Recommendations

When we compare the average musical features of the tracks recommended to the Rec group and those that were manually selected by the control group (NoRec), we can observe several significant differences regarding, e.g., danceablity, energy, popularity, freshness (release year), speechiness, or tempo i.e., the recommender often picks quite different tracks than users would do, see [11]. On the other hand, when comparing the recommendations to what the participants in the RecUsed group selected, the differences are no longer statistically significant, except for the popularity aspect. This is not surprising, as these participants often accepted the recommendations. The somewhat surprising aspect, however, is that the differences between the recommendations and the tracks of the RecNotUsed participants are also no longer significant. It is thus possible that the subjects in this group were biased (or inspired) by the presence of the recommendations. One indicator in favor of that possibility is that an overlap of 34% could be observed in terms of the artists that appeared in the recommendations and that were selected manually by the subjects. This means that the recommendations presumably influenced what users selected. This effect was previously investigated in a user study [18], where the participants exhibited a tendency to select items that were content-wise similar to a (random) recommendation.

[8]The popularity of the tracks were also determined using the Spotify's API, with values between 0 and 100 (lowest to highest popularity).

Table 3. Modified Borda Count: Ranking of Playlist Quality Criteria.

Criteria	All	RecUsed	RecNotUsed	NoRec
Homogeneity of musical features, e.g., tempo	250	68	79	103
Artist diversity	195	55	62	78
Transition	122	30	46	46
Popularity	106	39	34	33
Lyrics	95	32	34	29
Order	74	12	33	29
Freshness	32	12	11	9

A possible explanation why participants still selected tracks that are on average more popular than the recommended ones might lie in the intended use or context of the playlist to be created. If participants, for example, assumed that the playlist is designed to be played for a group of listeners (e.g., at a dance party), they might prefer to pick tracks that are presumably known by many people. In fact, the tracks selected for "dance" playlists were significantly more popular than those for "chill out" playlists.

Investigating Quality Criteria for Playlists

In order to better understand which quality characteristics one should consider when designing algorithms for playlist construction support, we analyzed the rankings that were provided by the participants in the post-task questionnaire, see the section on the study design. To determine the overall ranking, we used the Modified Borda Count method [9], which can be applied when some rankings are only partial, i.e., when not all items are ranked. The results are shown in Table 3.

The results indicate that, overall, the participants consider the *homogeneity of musical features*, e.g., tempo, energy or loudness along with the *artist diversity* of the resulting playlist as the most relevant quality criteria for playlists. On the other end of the spectrum, the *order* of the tracks in a playlist and their *freshness* were the least relevant aspects for the participants.

When looking at the different study groups, some smaller and not statistically significant differences in the rankings can be observed. Participants who used the recommendations considered track transitions less relevant than participants of the other groups. One potential explanation could be that the recommendations by the system (and, likewise, the created playlists) were perceived by users to be comparably coherent, e.g., in terms of the tempo, and the participants of the Rec group therefore paid less attention to the transitions. Another explanation is that the RecNotUsed participants are in general more demanding, and did not use some of the recommendations because they did not allow to make satisfying transitions. Other differences with respect to the quality criteria rankings were observed for the "lyrics" aspect, which was considered more important for road-trip and hip-hop playlists. Finally, popularity was considered a very important criterion almost exclusively for dance playlists for the above-discussed reasons.

General Quality Perception of the Recommendations

About 75% of the participants in the Rec group stated that they have looked at the recommendations when constructing the playlists. In the post-task questionnaire, these users were asked about their quality perception regarding the recommendations (as provided by Spotify's service). Different quality dimensions were assessed and users could express their agreement with related statements using 7-point Likert scale items. In our analysis, we considered answers that were greater than 4 as positive responses.

Specifically, we asked users to what extent the recommendations (1) matched the given topic; (2) helped them discover novel tracks or artists; (3) matched their general interests; (4) were already known to them; (5) were diverse in terms of genre and artist; (6) were generally popular or mainstream; and (7) were from trending music.

The results show that 62% of the respondents perceived the recommendation as topic-related (e.g., they match the playlist topic). More than half of the respondents also found the recommendations to match their interests and diverse in terms of genre and artist. With respect to novelty and familiarity, the results were mixed. In both cases many participants provided feedback with values above 4 (e.g., 45% in case of the novelty); a substantial amount of the participants however also found the recommendations of very limited novelty and familiarity.

We further compared the quality perception of the recommendations for the two groups of RecUsed and RecNotUsed. The only statistically significant difference we observed was in terms of novelty. In other words, those participants who used the recommendations in their playlists believed that the recommendations were significantly more novel (median=5) than the other participants who looked at the recommendations but did not accept them (median=3). One plausible explanation of this result could be that the participants who did not use the recommendations had a higher music expertise and were more demanding. However, this explanation is not in line with what the participants claimed in their answers, as the RecUsed group had a generally higher music expertise score than the RecNotUsed group (although the difference was not statistically significant). Another explanation is that participants were generally willing to discover novel music and the recommendations were mainly adopted by those participants who found the recommendations to be so.

Overall, the results indicate that Spotify's algorithms were quite successful in determining suitable tracks for recommendation. At the same time, the mixed results for some other factors are not necessarily negative either, because the importance of factors like the appropriate popularity level of the tracks – according to our analysis – can depend on the topic and intended purpose of the playlist.

Qualitative Analysis

We finally asked the participants to specify what they found to be the most difficult aspect when creating a playlist. *To remember the right tracks* was most often mentioned as a difficult task by the participants (e.g., "*Keeping up with the options. I often can't remember all the songs and performers I know*"). *To find good music for everyone* was another difficult task indicated by many (e.g., "*To meet the taste of all, as playlists are often created for occasions in which several people listen to them*"). Some participants also emphasized on *the smoothness of the track transitions* (e.g., "*The most difficult thing is to think of tracks that are connected on several dimensions at the same time such as their energy, the transitions, their level of sophistication, etc.*").

Recommendation tools have the potential to help users in many of these dimensions, e.g., remind them of tracks or artists they know and liked in the past, to find tracks that are popular in a certain community, or to find tracks that are coherent with the current playlist with respect to musical features. In fact, several participants that had no recommendation support have written about the potential advantages of having such a system: "*Sometimes I do not know what to search for and I expect the system suggest something. I prefer to choose just some songs and then listen/receive recommendations related to my initial choices*".

CONCLUSIONS

With our study we aim to shed more light on the perception and adoption of recommender systems that are designed to support users in the playlist construction process. The results of the study not only showed a high general adoption rate – almost half of the participants picked at least one track from the provided recommendations, while the proportion in the domain of recommender systems is often lower than 10% – but also that the presentation of recommendations can indirectly influence the choices of the users. Our study, therefore, provides additional evidence that recommenders can be a valuable tool to steer consumer behavior.

Generally, participants who received recommendations also explored more alternatives during the process (without needing much more time), which can be interpreted as a higher engagement in the task. As a result, a recommender system has the potential to increase the involvement of the user with the service, which from a business perspective in the best case leads to higher customer retention rates.

Finally, the study also showed that including a recommendation component can lead to a higher perceived difficulty of the task for users, which emphasizes the importance of focusing on intuitive UI designs.

ACKNOWLEDGEMENTS

We would like to thank Adrian Skirzynski who contributed to this work with the results of his master's thesis.

REFERENCES

1. Ivana Andjelkovic, Denis Parra, and John O'Donovan. 2016. Moodplay: Interactive Mood-based Music Discovery and Recommendation. In *UMAP '16*. 275–279.
2. Luke Barrington, Reid Oda, and Gert R. G. Lanckriet. 2009. Smarter than Genius? Human Evaluation of Music Recommender Systems. In *ISMIR '09*. 357–362.
3. Dominikus Baur, Sebastian Boring, and Andreas Butz. 2010. Rush: Repeated Recommendations on Mobile Devices. In *IUI '10*. 91–100.
4. Dirk Bollen, Bart P. Knijnenburg, Martijn C. Willemsen, and Mark Graus. 2010. Understanding Choice Overload in Recommender Systems. In *RecSys '10*. 63–70.
5. Geoffray Bonnin and Dietmar Jannach. 2014. Automated Generation of Music Playlists: Survey and Experiments. *ACM Computing Surveys* 47, 2 (2014), 26:1–26:35.
6. Sally Jo. Cunningham, David. Bainbridge, and Annette. Falconer. 2006. 'More of an Art than a Science': Supporting the Creation of Playlists and Mixes. In *ISMIR '06*. 240–245.
7. Sally Jo Cunningham, David Bainbridge, and Dana McKay. 2007. Finding New Music: A Diary Study of Everyday Encounters with Novel Songs. In *ISMIR '07*. 83–88.
8. J. Stephen Downie. 2003. Music Information Retrieval. *Annual Review of Information Science and Technology* 37, 1 (2003), 295–340.
9. Peter Emerson. 2013. The Original Borda Count and Partial Voting. *Social Choice and Welfare* 40, 2 (2013), 353–358.
10. Dietmar Jannach and Gedas Adomavicius. 2016. Recommendations with a Purpose. In *RecSys '16*. 7–10.
11. Dietmar Jannach, Iman Kamehkhosh, and Geoffray Bonnin. 2016. Biases in Automated Music Playlist Generation: A Comparison of Next-Track Recommending Techniques. In *UMAP '16*. 281–285.
12. Dietmar Jannach, Malte Ludewig, and Lukas Lerche. 2017. Session-based Item Recommendation in E-Commerce: On Short-Term Intents, Reminders, Trends, and Discounts. *User-Modeling and User-Adapted Interaction* (2017).
13. Chris Johnson. 2014. Algorithmic Music Discovery at Spotify. Online. (2014). `https://de.slideshare.net/MrChrisJohnson/algorithmic-music-recommendations-at-spotify`
14. Chris Johnson and Edward Newett. 2014. From Idea to Execution: Spotify's Discover Weekly. Online. (2014). `https://de.slideshare.net/MrChrisJohnson/from-idea-to-execution-spotifys-discover-weekly/12-Insight_users_spending_more_time`
15. Mohsen Kamalzadeh, Dominikus Baur, and Torsten Möller. 2012. A Survey on Music Listening and Management Behaviours. In *ISMIR '12*. 373–378.
16. Mohsen Kamalzadeh, Christoph Kralj, Torsten Möller, and Michael Sedlmair. 2016. TagFlip: Active Mobile Music Discovery with Social Tags. In *IUI '16*. 19–30.
17. Iman Kamehkhosh and Dietmar Jannach. 2017. User Perception of Next-Track Music Recommendations. In *UMAP '17*. 113–121.
18. Sören Köcher, Dietmar Jannach, Michael Jugovac, and Hartmut H. Holzmüller. 2016. Investigating Mere-Presence Effects of Recommendations on the Consumer Choice Process. In *IntRS Workshop at RecSys '16*. 2–5.
19. Alexandra Lamont and Rebecca Webb. 2011. Short- and Long-term Musical Preferences: What Makes a Favourite Piece of Music? *Psychology of Music* 38, 2 (2011), 222–241.
20. Jin Ha Lee and J Stephen Downie. 2004. Survey Of Music Information Needs, Uses, And Seeking Behaviours: Preliminary Findings. In *ISMIR '04*. 441–446.
21. Ali Sarrafi and Evan Shrubsole. 2015. A/B testing at Spotify. Online. (2015). `https://de.slideshare.net/alisarrafi3/ab-testing-at-spotify`
22. Harald Steck, Roelof van Zwol, and Chris Johnson. 2015. Interactive Recommender Systems with Netflix and Spotify. Online. (2015). `http://de.slideshare.net/MrChrisJohnson/interactive-recommender-systems-with-netflix-and-spotify`
23. S. Stumpf and S. Muscroft. 2011. When Users Generate Music Playlists: When Words Leave Off, Music Begins?. In *ICME '11*. 1–6.
24. Kirsten Swearingen and Rashmi Sinha. 2002. Interaction Design for Recommender Systems. In *Designing Interactive Systems, ACM*.
25. Nava Tintarev, Christoph Lofi, and Cynthia C.S. Liem. 2017. Sequences of Diverse Song Recommendations: An Exploratory Study in a Commercial System. In *UMAP '17*. 391–392.