On Skipping Behaviour Types in Music Streaming Sessions

Francesco Meggetto, Crawford Revie, John Levine, Yashar Moshfeghi
{francesco.meggetto,crawford.revie,john.levine,yashar.moshfeghi}@strath.ac.uk
University of Strathclyde, Glasgow, UK

ABSTRACT

The ability to skip songs is a core feature in modern online streaming services. Its introduction has led to a new music listening paradigm and has changed the way users interact with the underlying services. Thus, understanding their skipping activity during listening sessions has acquired considerable importance. This is because such implicit feedback signal can be considered a measure of users' satisfaction (dissatisfaction or lack of interest), affecting their engagement with the platforms. Prior work has mainly focused on analysing the skipping activity at an individual song level. In this work, we investigate different behaviours during entire listening sessions with regards to the users' session-based skipping activity. To this end, we propose a data transformation and clustering-based approach to identify and categorise skipping types. Experimental results on the real-world music streaming dataset (Spotify) indicate four main types of session skipping behaviour. A subsequent analysis of short, medium, and long listening sessions demonstrate that these session skipping types are consistent across sessions of varying length. Furthermore, we discuss their distributional differences under various listening context information, i.e. day types (i.e. weekday and weekend), times of the day, and playlist types.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

Spotify; Session; Skipping; User Behaviour; Listening; Music

1 INTRODUCTION

Online music streaming services such as Spotify and Amazon Music have surged in popularity in recent years due to content digitisation. If before the listening experience was solely based on cassettes, CDs and vinyl, now entire album collections are easily accessible on demand and within a few clicks. This new listening paradigm as well as the ability for users and systems to create personalised playlists and change songs, has led to a significant change in users' behaviour and interaction with these systems [14]. Listening behaviours such as skipping and scrubbing (i.e. seeking forward and backward by moving the cursor) are becoming prominent and of acquired importance [20, 21]. Such behaviours could be considered as an implicit feedback indicating the lack of interest of the user in the song being played, and in turn affecting the engagement of users with such platforms. Thus, understanding the session skipping patterns plays a role of greater importance in today's systems than it did years ago.

Existing work has focused on analysing the skipping patterns as a function of the time at which this takes place within a song (skip profile) [12, 23]. In this paper, however, we seek to analyse and identify different behaviours during entire listening sessions with regards to users' session-based skipping activity. In particular, we aim at understanding the influence of various listening contexts in skipping activity. Such analysis is performed on different day types, times of the day, and playlist types.

To better understand the variations in the distribution of types and the interplay with listening context information, we investigate the following research questions: **RQ1:** What are the main types of skipping behaviour that we can identify at a session level? **RQ2:** For those types, how does weekday/weekend day type affect their overall distribution? **RQ3:** Furthermore, do different times of the day, i.e. morning, afternoon, evening, and night affect users' skipping interaction with the streaming service? **RQ4:** Finally, in what ways do playlist types (e.g., personalised playlist, radio, etc.) affect users' skipping behaviour?

We investigate our research questions by analysing a large real-world music streaming dataset and on sessions of varying length, thereby providing generalisation of our findings. The contributions of this paper are three-fold. Firstly, we perform an extensive investigation on users' skipping behaviour during entire listening sessions. Secondly, we identify four main types of session skipping behaviour. Thirdly, we investigate the influence that various listening contexts have on the distribution of these patterns.

2 BACKGROUND

Research revolving around skipping behaviour in online platforms spans across ads on social media platforms [5–7] and music [12, 20, 23, 24, 26]. The latter, however, has been largely under-researched. Modelling and understanding skipping behaviour in music listening sessions arguably plays a crucial role in understanding user behaviour in modern streaming services. For instance, the skipping signal has already been used as a measure in heuristic-based playlist generation systems [9, 25], user satisfaction [16, 28], relevance [17], and as counterfactual estimators in Recommender Systems [22].

In a preliminary analysis of skip profiles [20], it was noted how a quarter of all streamed songs are skipped within the first couple

of seconds, and only half are listened to in their entirety. Also, Montecchio et al. [23] identified a connection between skip behaviour and musical structure. They show that users are more likely to skip a song right after a change of musical sections. The skip identity of a song is both very specific to the song as well as stable across time and geographical region. Such skip profile also follows a universal U-shaped pattern, with spikes in skipping rate at the beginning and end of the playback. In the work by Donier [12], the idea of skips being for the most part reactions to salient musical events is further reinforced and confirmed. Ng and Mehrotra [24] demonstrate that fluctuations in audio features are common in music streaming sessions and relate to the skipping behaviour of users. Taylor and Dean [26] find that people who usually listen to songs in their entirety (users were asked in advance for this information), show higher listening duration than those who do not.

To encourage research in this under-developed field, Spotify released the Music Streaming Sessions Dataset (MSSD) [10] and the *Sequential Skip Prediction Challenge* [1]. The challenge focused on predicting whether individual tracks encountered in a listening session will be skipped or not. Numerous deep-neural network [2, 8, 11, 15, 18, 27, 29] and supervised learning [13] based approaches have been proposed to predict such sequential music skip behaviour. Afchar and Hennequin [3] proposed using interpretable deep neural networks for skip interpretation via feature attribution.

In contrast to prior works, we propose an in-depth analysis on users' skipping behaviour during entire listening sessions, with the aim of finding and characterising different types of session skipping behaviour. This analysis is performed on the entire training set of the real-world MSSD dataset.

3 APPROACH

Session Skipping Pattern Extraction. In order to analyse and identify skipping types, we first need to extract the skipping activity across entire listening sessions. We refer to such pattern as the *session skipping pattern*. The skip features available in the selected dataset are *skip_1*, *skip_2*, *skip_3*, and *not_skipped*. These represent defined thresholds that respectively indicate whether the track has been played very briefly, briefly, mostly, or in full. In Table 1, a summary of the interplay among these features at record-level is presented, together with an integer transformation (ID) for every observed pattern. This is motivated by the easing of result reporting in later sections. It is important to note that a rarely occurring pattern, namely "*False, True, False, False*", as well as sessions with missing values are ignored in our analysis.

With a scalar representation of the skipping activity for every record of a session, we can now construct a session-level vector representation. It is the ordered sequence by session position of all individual skipping activities that form a session. Thus, for a session of length 20, a vector of 20 elements captures the *session skipping pattern*. For example, "1, 1, 2, 1, 5, 3, 1, 1, 1, 2, 1, 1, 2, 1, 1, 1, 2, 5, 3, 5", indicates that in this session a user played very very briefly the first two songs, followed by very briefly and very very briefly for the 3rd and 4th songs respectively, before the 5th song was played in full, etc. An interpretation of the session skipping patterns can also be used to indicate how long each track was played for.

Table 1: Summary of skip patterns and their corresponding translation in terms of for how long the current track was played. ID is an integer value associated with each pattern, used in the construction of session skipping patterns.

skip_1	skip_2	skip_3	not_skipped	Listening Length	ID
True	True	True	False	Very Very Briefly	1
False	True	True	False	Very Briefly	2
False	False	True	False	Briefly	3
False	False	False	False	Most	4
False	False	False	True	All	5

Session Skipping Type Identification. With a set of session skipping patterns, we apply Principal Component Analysis (PCA) [1]. This is motivated by the fact that the input space contains noisy patterns as well as that reducing the dimensionality of such large sets decreases the computational requirements. Later, the transformed data, i.e. the PCA components, are used as the input space for k-means clustering, which uses Euclidean distance to choose the centroids. We also adopt the kmeans++ initialisation scheme, known to augment speed and quality of the clusters [4].

A common, but subjective, cut-off point of the total explained variability by PCA is 70% [19]. To find the optimal variance to retain, we performed local evaluation on the number of PCA components. Increasing the retained variance and/or applying clustering on the raw data yields similar results. We believe this to be due to the noise and amount of outliers. By decreasing the variance, only the outliers are less represented in the PC space. Since our goal is to find the main behavioural types, the difference is minimal given that the highly relevant patterns are well described in the PC space. We also perform extensive parameter experimentation on the number of clusters, in the range of 2 to 40. Finally, each analysis is performed 5 times and with 5 arbitrarily selected random seed numbers.

4 EXPERIMENTAL SETTINGS

Dataset. We conduct our experiments on Spotify's Music Streaming Sessions Dataset (MSSD) [10]. The publicly available training set consists of approximately 150 million logged streaming sessions, collected over 66 days from July 15th and September 18th 2018, with each day comprising 10 logs. Each log includes streaming sessions uniformly sampled at random throughout the entire day. Sessions are defined as sequences of songs/tracks that a user has listened to, containing from 10 up to at most 20 records (one record per song). They include contextual information and user interaction history (e.g., if the user is premium, context type of the playback) and audio features and metadata describing the tracks (e.g., duration, popularity). Implicit user feedback is provided in the form of actions such as song skipping behaviour, pauses during playbacks, and user actions that led to the current track being played or ended. Important to note is the fact that there is no user identification, nor access to demographic or geographical information. Hence, we do not know whether two sessions have been played by the same user or by two different users. Finally, we perform our analysis on the complete training set. The testing set is not used since most of the metadata as well as the skipping attributes are missing.

Experimental Conditions. After removal of sessions with unrecognised skipping activity, as described in Section 3.1, of the resulting 125 million listening sessions, the majority (47.7%) corresponds to sessions of length 20. The remainder are distributed in decreasing order from sessions of length 10 (8.8%) to those of length

19 (2.8%). Given this, we perform our analysis on long sessions (length 20), but also, to show the adaptability and extendability of our approach to various lengths, generalise our results by including data from medium (length 15) and short (length 10) sessions.

Then to answer our RQs, the following contextual scenarios are formulated: *Weekdays and Weekends:* The 66 days of training data can be grouped into 9 full weeks, hence 45 weekdays and 18 weekend days. The 10th Week is only partially available, with only 3 available days, i.e. 16th, 17th, and 18th of September. For a fair comparison they are therefore ignored when answering RQ2. *Time of The Day:* We define 5 time windows, where numbers in brackets correspond to the hour range: night (0-5), morning (6-11), afternoon (12-17), evening (18-23), all (0-23). If a session spans across two hours, we round up and consider the whole session as either part of start or end hour. *Playlist Type:* The playlist type is defined as the type of playlist that the playback occurred within. It includes editorial playlist, user collection, catalog, radio, charts, and personalized playlist. As it can be subject to change throughout a session (playlist switch), for the scope of this condition we restrict our analysis only to sessions with a single playlist type. *All:* The analysis is performed on all days and all their listening sessions. **Experimental Procedure.** By retaining 70% of the explained variance, the results in this paper are based on using 7 PCA components for long sessions, 6 for medium, and 5 for short. Moreover, we empirically identify four to be the optimal number of clusters.

To allow comparison of the four identified types for different experimental conditions, we perform cluster matching so as to pair highly similar clusters from one experiment (e.g. morning) to another (e.g. afternoon). For every cluster, we group all of its session skipping patterns and produce a single averaged vector representation. That is, every position in the session is the average for that position of all session skipping patterns. We call this single averaged vector representation the *average skipping session* for that cluster. Clusters matching is then performed via pairwise Euclidean distance. Each average skipping session from one experiment is matched to its closest cluster in another experiment. We find Euclidean distance to be the most suitable metric for this task. Other metrics, such as cosine similarity, generated inconsistencies and mismatches when applied on the average skipping sessions, even when the averages were computed on the PCA components rather than session positions. Local evaluation on the different runs suggests high similarity and stability of the results, and hence the reporting of results is for seed 0. The highest observed Euclidean distances among seeds for all reported results are of 0.2259, 0.2363, and 1.1756 for long, medium, and short sessions respectively.[2]

5 EXPERIMENTAL RESULTS

In Figure 1, the average skipping sessions and their corresponding type name (i.e., *listener*, *listen-then-skip*, *skip-then-listen*, *skipper*) for a "All" analysis is reported. As can be observed, the four identified types appear to be consistent across sessions of different length. This suggests that such dominant behaviours have no strong relation with the absolute length of a session and are thus generalisable. Additionally, a closer examination reveals that in fact we observe two main distinctive behaviours: *listener* and *listen-then-skip*. The

remaining two types, i.e. *skipper* and *skip-then-listen*, can be seen as their respective complimentary behaviour (addressing **RQ1**).

(a) Long Sessions (Length 20)

(b) Medium Sessions (Length 15)

(c) Short Sessions (Length 10)
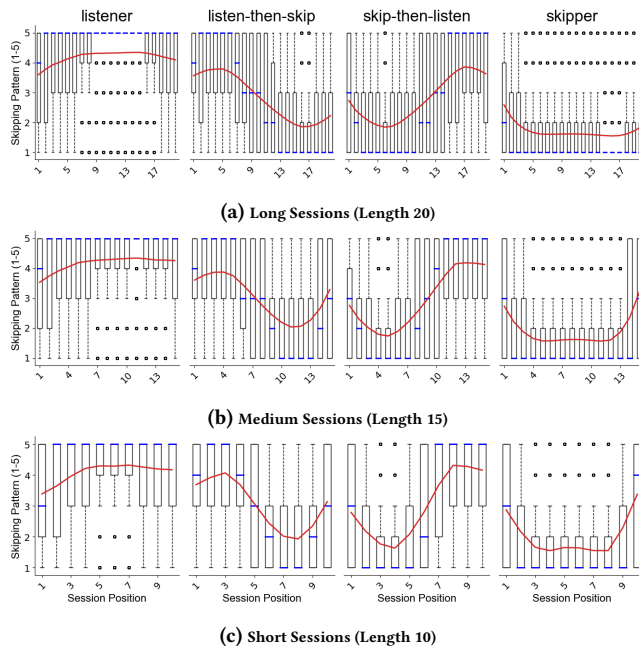
Figure 1: Box plots of the identified skipping types for different lengths and for a "All" analysis. The x-axis is the range of session positions $[1..n]$, where n is 20, 15, or 10 depending on selected length. The y-axis represents the skipping patterns (ID 1-5 in Table 1). The red line indicates the average skipping session.

Intuitively, categorising all listening sessions in four types yields a high variance, as shown in Figure 1. This is because each type agglomerates sessions that are considered similar but yet contain a noisy sequence of session skipping patterns. For a *skipper* type, an example is a pattern that contains all ID1s except for a middle ID5. These sessions can arguably be a different, and more refined, behaviour and not necessarily a member of *skipper*. Those branches of sub-behaviours with lower variance emerge with a higher number of clusters. Successive analysis also highlights that some of these refined types are the transposed representation of other types. That is, for a long session the ID5 shift of the previous example can happen in any of the 20 positions. Therefore, this can result in a likewise number of different behaviours, where the resulting bell shape is shifted along the session's position. Ultimately, however, they can arguably represent a single type of skipping behaviour which is position-independent. Finally, by refining types there also emerges the behaviour type that consists of the continuous alternation of skipping and no skipping. This yields an average skipping session similar to a cosine function.

However, as we perform refining, we note divergence in common behaviour across experiments. Whereas with a low number of types the clusters matching has high accuracy, with a higher number the comparisons might be performed on different behaviours. This is due to the identification of ungeneralisable behaviours that tend to differ across experiments.[3] Further, since the goal of this paper is

---

[3]Due to space limitations differences among cluster validity indices (e.g. Calinski-Harabasz, Silhouette, Dunn, etc.), are omitted from this paper.

to identify the main generalisable behaviours and their variations under different listening context information rather than dissimilarities, in the next subsection we report results for the analysis on the four types previously outlined.
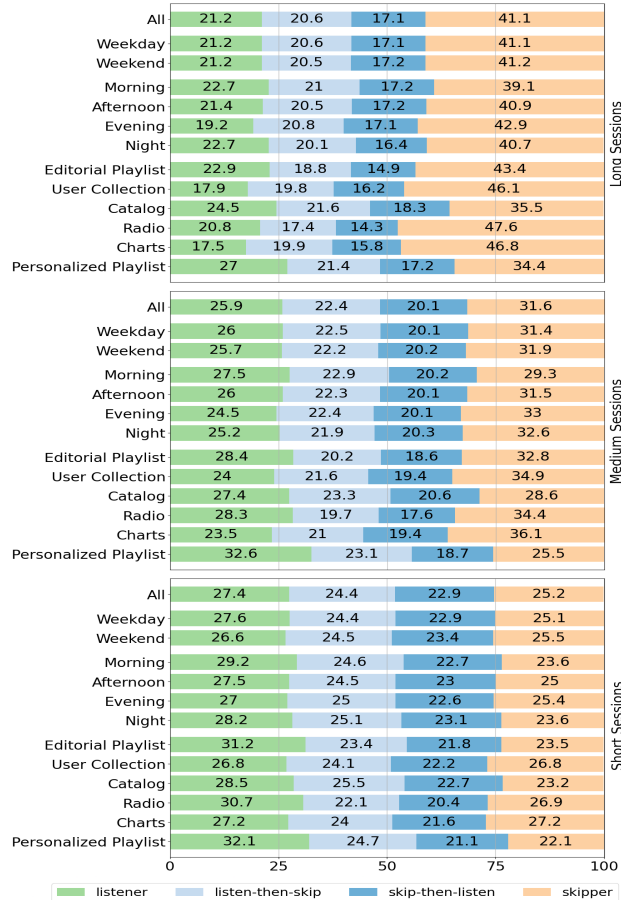
Figure 2: Distribution of types under different experimental settings for long (*top*), medium (*centre*), and short (*bottom*) sessions. The x-axis represents the membership distribution in percentage value for each type.

Distribution Differences. We observe from Figure 2 that there are no significant differences in distribution across weekdays and weekends (addressing **RQ2**). However, given Spotify's global user distribution and cultural differences around the definition of the "weekend", it may be unsurprising that no differences were observed. Further inclusion of geographical and time-of-day variables may be required to properly evaluate any 'weekend' effect. Differences emerge when we narrow down our analysis to the different times of the day. Morning users have the lowest *skipper* activity. In contrast, such behaviour is always at its highest in the evening. This suggests that the amount of *skipper* activity increases as we progress through the day. We hypothesise this to be due to morning users paying higher attention to their daily tasks rather to the songs currently playing. In contrast, evening users may have more spare time, thus allowing them to pay closer attention to the streamed content (addressing **RQ3**). To address **RQ4**, we can immediately

note how the *skipper* activity is consistently lower for catalog and personalised playlists. This steep decrease in comparison with other types is highly prevalent in long and medium sessions but less pronounced for short sessions. It would also appear to be the case that the user collection, radio, and charts categories are associated with the highest degree of skipping activity regardless of session length. These findings partially meet our expectations. Catalog users have higher control in music selection than charts. Surprisingly, charts has consistently one of highest *skipper* activity, despite being a collection of the most popular titles of the moment. This suggests that music popularity has an impact in skipping behaviour. Lastly, for the two under-represented types, namely *listen-then-skip* and *skip-then-listen*, we note how the former has always a higher activity, independently on the absolute session length. This suggests that it is more likely for users to listen to the first half and skip the second rather than the opposite. In the case of the radio, such under-represented types are the lowest of all playlist types. This suggests that radio users vary less their skipping patterns, thus consistently be a *listener* or *skipper* type.

Major differences can also be seen when looking at the distribution proportions for different session lengths. In the example of the "Weekday" analysis, we note how that the proportion of *skippers* drops from a 41.1% in long sessions, to 31.4% in medium, and 25.1% in short. Conversely, the proportion associated with *listener* type increases from 21.2% to 26%, eventually reaching 27.6% in short sessions. Overall, as sessions become shorter, we see a decrease of *skipper* type behaviour and an increase of the other three types. If the *skipper* type prevails in long and medium sessions, we note how for short sessions the gaps are reduced, with *listener* becoming the highest representative behaviour. In other words, under our experimental settings and with the provided data, we observe that in comparison with long listening sessions, in short sessions users tend to increase their listening activity and thus skip less frequently. Understanding the factors that lead to such an increase in listening activity as sessions become shorter is left as future work.

6 CONCLUSION

In this work, we tackled the task of identifying different behaviours during entire listening sessions with regards to the users' session-based skipping activity. This analysis is performed on a large real-world dataset of music streaming listening sessions. By adopting an effective data transformation and clustering-based approach, we identified four dominant skipping types, namely *listener*, *listen-then-skip*, *skip-then-listen*, and *skipper*. We also observed that: (i) types are consistent across sessions of varying length, thus representing a generalisable, common behaviour; (ii) two pairs of complimentary behaviour can be observed, i.e. *listener* with *skipper* and *listen-then-skip* with *skip-then-listen*; (iii) the distribution of types varies under different listening context information as well as when varying the session length. The main limitation of our work is that it observes distribution variations under a limited set of features. We hope our findings to inspire future research in better understanding and modelling music skipping behaviour.

REFERENCES

[1] Hervé Abdi and Lynne J Williams. 2010. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* 2, 4 (2010), 433–459.

[2] Sainath Adapa. 2019. Sequential modeling of Sessions using Recurrent Neural Networks for Skip Prediction. *arXiv preprint arXiv:1904.10273* (2019).

[3] Darius Afchar and Romain Hennequin. 2020. Making neural networks interpretable with attribution: application to implicit signals prediction. In *Fourteenth ACM Conference on Recommender Systems*. 220–229.

[4] David Arthur and Sergei Vassilvitskii. 2006. *k-means++: The advantages of careful seeding*. Technical Report. Stanford.

[5] Snehasish Banerjee and Anjan Pal. 2021. Skipping Skippable Ads on YouTube: How, When, Why and Why Not?. In *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)*. IEEE, 1–5.

[6] Daniel Belanche, Carlos Flavián, and Alfredo Pérez-Rueda. 2017. User adaptation to interactive advertising formats: The effect of previous exposure, habit and time urgency on ad skipping behaviors. *Telematics and Informatics* 34, 7 (2017), 961–972.

[7] Daniel Belanche, Carlos Flavián, and Alfredo Pérez-Rueda. 2020. Brand recall of skippable vs non-skippable ads in YouTube. *Online Information Review* (2020).

[8] Ferenc Béres, Domokos Miklós Kelen, András Benczúr, et al. 2019. Sequential skip prediction using deep learning and ensembles. (2019).

[9] Klaas Bosteels, Elias Pampalk, and Etienne E Kerre. 2009. Evaluating and Analysing Dynamic Playlist Generation Heuristics Using Radio Logs and Fuzzy Set Theory.. In *ISMIR*, Vol. 9. 351–356.

[10] Brian Brost, Rishabh Mehrotra, and Tristan Jehan. 2019. The music streaming sessions dataset. In *The World Wide Web Conference*. 2594–2600.

[11] Sungkyun Chang, Seungjin Lee, and Kyogu Lee. 2019. Sequential Skip Prediction with Few-shot in Streamed Music Contents. *arXiv preprint arXiv:1901.08203* (2019).

[12] Jonathan Donier. 2020. The universality of skipping behaviours on music streaming platforms. *arXiv preprint arXiv:2005.06987* (2020).

[13] Andres Ferraro, Dmitry Bogdanov, and Xavier Serra. 2019. Skip prediction using boosting trees based on acoustic features of tracks in sessions. *arXiv preprint arXiv:1903.11833* (2019).

[14] Benjamin Fields et al. 2011. *Contextualize your listening: The playlist as recommendation engine*. Ph.D. Dissertation. Goldsmiths College (University of London).

[15] Christian Hansen, Casper Hansen, Stephen Alstrup, Jakob Grue Simonsen, and Christina Lioma. 2019. Modelling sequential music track skips using a multi-rnn approach. *arXiv preprint arXiv:1903.08408* (2019).

[16] Casper Hansen, Christian Hansen, Lucas Maystre, Rishabh Mehrotra, Brian Brost, Federico Tomasi, and Mounia Lalmas. 2020. Contextual and sequential user embeddings for large-scale music recommendation. In *Fourteenth ACM Conference on Recommender Systems*. 53–62.

[17] Christian Hansen, Rishabh Mehrotra, Casper Hansen, Brian Brost, Lucas Maystre, and Mounia Lalmas. 2021. Shifting Consumption towards Diverse Content on Music Streaming Platforms. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 238–246.

[18] Olivier Jeunen and Bart Goethals. 2019. Predicting Sequential User Behaviour with Session-Based Recurrent Neural Networks. (2019).

[19] Ian T Jolliffe and Jorge Cadima. 2016. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374, 2065 (2016), 20150202.

[20] Paul Lamere. 2014. *The Skip*. Retrieved Jun 9, 2021 from https://musicmachinery.com/2014/05/02/the-skip/

[21] Paul Lamere. 2015. *The Drop Machine*. Retrieved Jun 9, 2021 from https://musicmachinery.com/2015/06/16/the-drop-machine/

[22] James McInerney, Brian Brost, Praveen Chandar, Rishabh Mehrotra, and Benjamin Carterette. 2020. Counterfactual evaluation of slate recommendations with sequential reward interactions. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1779–1788.

[23] Nicola Montecchio, Pierre Roy, and François Pachet. 2020. The skipping behavior of users of music streaming services and its relation to musical structure. *Plos one* 15, 9 (2020), e0239418.

[24] Aaron Ng and Rishabh Mehrotra. 2020. Investigating the Impact of Audio States & Transitions for Track Sequencing in Music Streaming Sessions. In *Fourteenth ACM Conference on Recommender Systems*. 697–702.

[25] Elias Pampalk, Tim Pohle, and Gerhard Widmer. 2005. Dynamic Playlist Generation Based on Skipping Behavior.. In *ISMIR*, Vol. 5. 634–637.

[26] John R Taylor and Roger T Dean. 2021. Influence of a continuous affect ratings task on listening time for unfamiliar art music. *Journal of New Music Research* (2021), 1–17.

[27] Charles Tremlett. 2019. Preliminary Investigation of Spotify Sequential Skip Prediction Challenge. (2019).

[28] Hongyi Wen, Longqi Yang, and Deborah Estrin. 2019. Leveraging post-click feedback for content recommendations. In *Proceedings of the 13th ACM Conference on Recommender Systems*. 278–286.

[29] Lin Zhu and Yihong Chen. 2019. Session-based Sequential Skip Prediction via Recurrent Neural Networks. *arXiv preprint arXiv:1902.04743* (2019).