Spotivibes: Tagging Playlist Vibes With Colors

Hiba Abderrazik[*]
h.abderrazik@student.tudelft.nl
Delft University of Technology
Delft, The Netherlands

Giovan Angela[*]
g.j.a.angela@student.tudelft.nl
Delft University of Technology
Delft, The Netherlands

Hans Brouwer[*]
j.c.brouwer@student.tudelft.nl
Delft University of Technology
Delft, The Netherlands

Henky Janse[*]
h.a.b.janse@student.tudelft.nl
Delft University of Technology
Delft, The Netherlands

Sterre Lutz[*]
s.lutz@student.tudelft.nl
Delft University of Technology
Delft, The Netherlands

Gwennan Smitskamp[*]
g.m.smitskamp@student.tudelft.nl
Delft University of Technology
Delft, The Netherlands

Sandy Manolios
s.manolios@tudelft.nl
Delft University of Technology
Delft, The Netherlands

Cynthia C. S. Liem
c.c.s.liem@tudelft.nl
Delft University of Technology
Delft, The Netherlands

ABSTRACT

Music is often both personally and affectively meaningful to human listeners. However, little work has been done to create music recommender systems that take this into account. In this demo proposal, we present Spotivibes: a first prototype for a new color-based tagging and music recommender system. This innovative tagging system is designed to take the users' personal experience of music into account and allows them to tag their favorite songs in a non-intrusive way, which can be generalized to their entire library. The goal of Spotivibes is twofold: to help users better tag their playlists to get better playlists and to provide research data on implicit grouping mechanisms in personal music collections. The system was tested with a user study on 34 Spotify users.

KEYWORDS

Recommender systems; Personal experience of music; Emotion-based recommendations; Color-based tags

ACM Reference Format:
Hiba Abderrazik, Giovan Angela, Hans Brouwer, Henky Janse, Sterre Lutz, Gwennan Smitskamp, Sandy Manolios, and Cynthia C. S. Liem. 2019. Spotivibes: Tagging Playlist Vibes With Colors. In *Proceedings of Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS '19) (IntRS '19)*. CEUR-WS.org, 5 pages.

1 INTRODUCTION

Many people love to listen to music and share their music tastes with others. With music consumption largely having moved to the digital realm, music organization and discovery have moved to the digital space accordingly, opening up great opportunities for digital music services to support these experiences.

However, many popular present-day music services are very much framed as catalogues, in which users have to perform directed, linguistic searches on existing song metadata to find what they are looking for. In the Music Information Retrieval research domain, considerable work has been performed to automatically describe music objects beyond catalogued metadata. However, much of the research in this area still has focused on fairly "objective" descriptors of aspects of the music object (e.g. chords, tempo), but did not explicitly consider corresponding end user experiences [3, 6, 11].

Frequently, music is seen as a moderator of mood and emotion. A considerable body of work on automatic music emotion recognition from audio content exists [2, 9, 17]. However, generally, it is hard to get good labeled data (for which humans need to give the initial input) at scale. In order to make labeling engaging, several proposals have been made for crowdsourced tagging games [1, 8, 10]. While these are more engaging to users than traditional tagging interfaces, they explicitly ask for users to concentrate on the annotation within imposed linguistic frames (e.g. describing songs with a tag, or mapping songs in valence-arousal space), which may take away the "natural" affective experience of music consumption. Furthermore, these tagging interfaces generally reward consensus across human annotators. While this allows for labels that are more stable and generalizable across a music population, this takes away any notion of very personal and subjective perception.

Also with regard to automatic music recommendation, in which user consumption patterns are taken into account to foster automatic music discovery, it was pointed out that true user feedback is not yet optimally integrated [14]. While many algorithms are evaluated with user studies or trained on hand-labeled genre tags, not many approaches holistically incorporate user responses.

While algorithms have focused on describing musical objects, when humans listen to music in everyday life, they actually may not have their full focus on the musical object for active listening. Instead, multiple studies have shown that music is often consumed passively, e.g. in the background while performing another activity [5, 7, 12, 16]. This again gives useful dimensions of (personal) music categorization, that presently still are understudied.

To study these open challenges in the literature and music consumption in general, we propose the Spotivibes system. This system is designed to capture user reactions and associations to music in both a personal and an abstract way, in an integrated way with the

[*] All marked authors contributed equally to this research.

IntRS '19, September 2019, Copenhagen, DK
2019.

user's existing listening preferences in the Spotify music service. Taking a user's existing playlists as the basis, users are asked to tag the "vibe" of songs (with "vibe" intentionally chosen to be more abstract than "mood" or "purpose') with one or more colors. This both restricts the tag vocabulary in the backend, while at the same time, it allows for more abstract associations at the user side than would be the case when imposing a certain vocabulary.

In the backend, the system will learn associations from colors to content features in the user's consumption history. Consequently, the system can generate tailored playlists for the users based on colors. In this way, Spotivibes serves a twofold goal: on one hand, it can serve as an annotation system that is both more abstracted than existing tagging tools, while at the same time being more integrated with actual everyday listening behavior of a user. On the other hand, it also directly can serve users in gaining insight into their music preferences and associations, and setting more personal recommendations. This makes Spotivibes an interesting research tool to study the impact and usefulness of abstract color tagging of personal perception of music in recommender systems. In the current paper, we present a first functional prototype of Spotivibes, that is intended to provide a framework for conducting deeper research on tagging mechanisms in the future.

2 OVERVIEW OF THE APPLICATION

Spotivibes is a web application and as such only requires a device with Internet access and a browser, as well as a Spotify account. Upon their first login, users have to tag a certain number of songs (10 or 30) from their Spotify saved tracks, using as many colors as they want. The available colors are depicted in the upper part of Figure 1. Then, they can get personalized playlists based on a single color, a progression between two colors or a mosaic of colors. Those playlists can then be exported to their Spotify account.

Spotivibes relies on user feedback to further improve its recommendations : users can always modify existing tags or tag more songs. Users have also access to various statistics regarding their tags to give them more information about their tagging behavior and motivate them to tag more.

A detailed overview of the application can be found at **https://youtu.be/x2KZ2z0s4Uk** .

2.1 Initialization

The initial set a user is asked to label is based on k-means clustering of Spotify's audio features. The user is asked to label either 10 or 30 of their own tracks, so k is set to 10 or 30. This theoretically gives tracks which represent the major clusters of songs in Spotify's audio feature space and so should cover the majority of associations a user can have to songs in their library. Also a "reset labels" button on the home page allows the user to clear all the label data the user has provided. This way, the initialization process can be repeated for a fresh start.

2.2 Bulk labeling

Once the initialization process has been completed, if the user wants to label more songs, he/she can select multiple songs and labels and tag that selected group of songs with a color in one go. The user can also label their own saved Spotify playlists in one go, labeling each song with the chosen color.

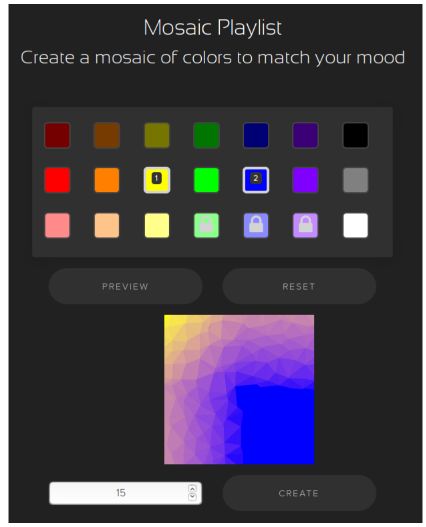


Figure 1: Mosaic playlist generation

2.3 Playlist generation

Spotivibes allows users to create their vibes-based playlists in three different ways: a gradient playlist, a single color playlist and a mosaic playlist.

One color. The single color playlist is pretty self-explanatory. The user will select a single color and will receive a playlist containing songs with a high label value for the selected color.

Gradient. The gradient playlist generation works by selecting two different colors. A new playlist will be generated with a gradual change in song vibe from start to finish. For example, the user selects the colors yellow and blue, for the first and second colors respectively. The first songs in the playlist will contain a higher "yellow" label assigned to it and gradually change to songs with that contain more "blue".

Mosaic. The mosaic pattern works by selecting multiple colors, the user can also select the same color multiple times. As shown in Figure 1, if a user selects two blue and one yellow, a playlist will be generated containing songs with more blue than yellow, but should also contain yellow.

Editing and Exporting Playlists. Once a playlist has been generated, the user can give feedback on each song by updating its color labels. Songs can also be removed, which gives negative feedback for future playlist generation. After creating and editing a playlist, a user can choose to export the playlist to their Spotify library. They can give it any custom name and later listen to it on their own Spotify account.

2.4 Statistics

As a part of their Spotivibes experience, users can get insight into their labeling behavior on a "Statistics" page. This page provides some basic information about the users, including the number of songs the user has labeled, and tracks in the library. More detailed statistics listed in the subsections below can be viewed by selecting a color from the color picker pop-up window on the left side. For

Figure 2: The different statistic representations of the user's labelling behavior and playlists.

displaying the different statistics plots shown by Figure 2, the data is calculated by the classifiers or retrieved from the Spotify API.

Landscape. The "Landscape" statistic is a detailed 3d plot providing information about the songs labeled with the selected color. The x, y, and z-axis of the 3d plot indicate tempo, valence, and loudness respectively. Each song labeled with the selected color will be displayed as a dot on this 3d plot, its size corresponding to the certainty with which we have classified it to be that color, as shown in the upper left part of Figure 2. For example, if a user associates yellow songs with high-tempo numbers, a cluster of larger dots will appear on the higher end of the tempo axis. The plot is interactive: it can be dragged to be viewed from different angles and when the user hovers over a dot, they can see the title and artist of the song it represents.

Influencers. The "Influencers" section, displayed in the upper right part of Figure 2, is a bar plot showing the 3 most influential artists within a color. The metric used to measure "influence" is simply the sum of the likelihood of all the songs of the artist within that color. In this way, influence indicates the likelihood of an artist being associated with the currently selected color, depending on how many of this artist's songs are classified as that color.

Decades. The "Decades" tile displayed in the lover left part of Figure 2 shows a histogram of the number of tracks in decades that belong to the selected color, weighted by their likelihood to be correctly classified.

Genres. The "Genre" tile displayed in the lover right part of Figure 2 shows a radial histogram of genres classified within the selected color.

2.5 Associating songs with colors

The algorithm that learns correspondences between songs and color tags is the heart of Spotivibes' functionality, yet is almost invisible to users. Since color labels are so personal, we do not make use of any inter-user information. This means that classifiers need to be trained for each individual user, yielding user-dependent correspondences between audio features and (categorically modeled) color tags.

Our color label predictor consists of an ensemble of classifiers and regressors from the scikit-(multi)learn [13] and XGBoost [4] python packages.

The label predictor must find underlying audio features that are strongly correlated with the labels that users give to songs. This, of course, means that the predictor is strongly influenced by how a user labels tracks. If a user chooses to use a color as a label for something completely uncorrelated with the audio features, no meaningful connections will be found, but this will also show accordingly in the "Statistics" overviews.

[15] found that, in the context of multi-label classification of music by emotion, the random k-label sets (RAKEL) ensembling method worked best on datasets of a similar size to most user Spotify libraries (less than 15,000 songs). The RAKEL algorithm works by training multiple naive predictors on subsets of the total label set and then deciding on the final output by voting among the predictors. Here we used scikit-multilearn's `RAkELo` module. Based on a set of training features and training labels, the algorithm outputs a list of features that is most descriptive for a color. To allow for some tolerance in the predictions, the RAKEL's binary classification was combined with a regression of the label values, for which we made use of scikit-multilearn's `MultiOutputRegressor` module. This means a song can e.g. be 30% green and 70% blue. Depending on need, the fractional scores can be thresholded in different post-processing steps. For example, labels with a score higher than 0.5 score are currently shown to users in the front-end (this gives a nice balance between showing many labels, while not just returning all the colors), and calculating "influencer" artists for the statistics page only incorporates the most certain predictions.

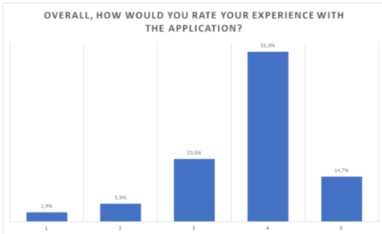
3 EVALUATION: USER STUDY

A user study was conducted to assess the usability of the system and quality of recommendations (measured by user satisfaction). The study was conducted with 34 participants recruited via personal connections and among computer science students of our university. They all had to freely explore the application on their own, and fill in a questionnaire afterwards. All users had to go through the longer setup which made them tag 30 of their Spotify favorite songs. The experiment lasted around 20 minutes per participant.

The questionnaire was composed of 17 questions. The answers to the main questions are shown in Figure 3. They were designed to measure the tediousness of the initial set-up process, user satisfaction in the recommendations, the perceived usefulness of the color-based tagging system and the usability of the interface. Other notable questions were about the user satisfaction of Spotivibes.

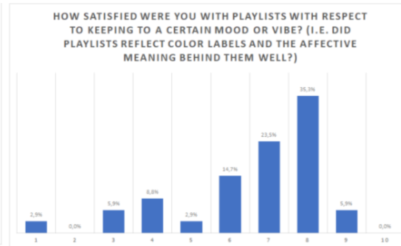
The user study concludes that overall, the participants had a good/satisfactory experience with the application (3.74 average on a 5 points scale), but were less satisfied with the services provided by Spotivibes (3.41 average on a 5 points scale), as shown in Figure 3a and Figure 3b .

Initialization Process. One thing that emerged during this study is that the song labeling process was on the edge of tediousness as shown in Figure 3c. The results shows an even split of 1/3rd of users agreeing the process was tedious, 1/3rd being neutral, and 1/3rd disagreeing that the process was tedious. We might consider going back to include the short labeling process in further user

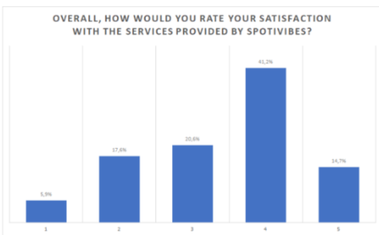


(a) Overall, how would you rate your experience with the application?

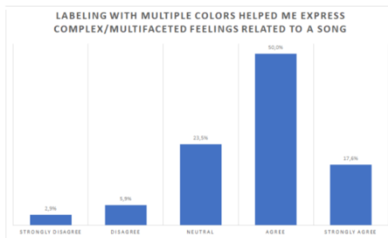
(c) The labeling process was tedious (Likert-scale)



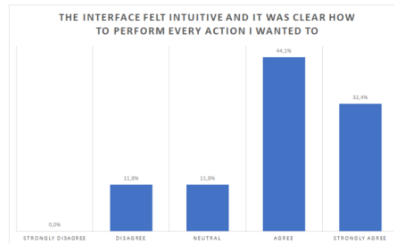
(e) How satisfied were you with playlists with respect to keeping to a certain mood or vibe? (10-point scale)



(b) Overall, how would you rate your satisfaction with the services provided by Spotivibes?



(d) Labeling with multiple colors helped me express complex/multifaceted feelings related to a song. (Likert-scale)



(f) The interface felt intuitive and it was clear how to perform every action I wanted to. (Likert-scale)

Figure 3: Bar charts with answers to six of the main questions about the users' satisfaction in Spotivibes, its functionalities and perceived usefulness.

experiments, but this can result in a decrease in playlist satisfaction. Perhaps a quick initialization process with better bulk labeling features could be included during initialization to improve user-friendliness as well as data for the classifier.

Playlist Generation. The user study was sadly inconclusive on the value of Spotivibes' color-based playlist generation, as shown by Figure 3e. Users were asked to rate their satisfaction with Spotify and Spotivibes on a 1 to 10 scale in terms of keeping to a given vibe or emotion. Disregarding low scores for Spotivibes related to a couple of users for which the initialization process failed due to bugs in the data management model, the minimum, lower quartile, median, upper quartile, and maximum were identical and the mean score for Spotivibes was 0.2 lower (not statistically significant). This might be affected by our choice of splitting the rating of Spotify and Spotivibes into different parts of the user study. In-person feedback from a couple of users indicated that they did not realize they were rating the two services against each other. Perhaps placing those two questions next to each other in the survey would have given a better view of how users actually felt about recommendations.

Colour Associations. Users were, however, generally satisfied with the use of colors as labels for emotions, as shown by Figure 3d. Half agreed that colors made it easy to represent emotions, a quarter were neutral, and a quarter disagreed. When asked whether using multiple colors helped express complex or multi-faceted feelings 65% agreed and only 10% disagreed. This does point towards the usefulness of colors as abstract labels for emotions in music. An interesting point to note was that users that gave negative feedback on the intuitiveness of the color labeling process (regarding their difficulty relating colors to songs or not knowing multiple labels could be used) also had lower satisfaction with the quality of playlist generation. This suggests that our classifier does actually pick up on patterns in the user's color labels and functions better when users label meaningfully.

4 CONCLUSION

Spotivibes is an innovative color-based tagging system that allows users to tags songs in a personal, intuitive and abstract way in order to get personalized playlists that supports their unique experience and needs of music. The current version of Spotivibes still is an early, but functional prototype, on which initial user studies have been performed. In future work, deeper research into the merit of the color-based tagging is planned to be performed, also including larger-scale user studies.

ACKNOWLEDGMENTS

We would like to thanks Bernd Kreynen for his valuable feedback throughout the project.

REFERENCES

[1] Anna Aljanaki, Frans Wiering, and Remco C Veltkamp. 2016. Studying emotion induced by music through a crowdsourcing game. *Information Processing & Management* 52, 1 (2016), 115–128.
[2] Anna Aljanaki, Yi-Hsuan Yang, and Mohammad Soleymani. 2017. Developing a benchmark for emotional analysis of music. *PloS one* 12, 3 (2017), e0173392.
[3] Michael A. Casey, Remco C. Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney. 2008. Content-Based Music Information Retrieval: Current Directions and Future Challenges. *Proc. IEEE* 96, 4 (April 2008), 668–696. https://doi.org/10.1109/JPROC.2008.916370
[4] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 785–794.

[5] Andrew Demetriou, Martha Larson, and Cynthia C. S. Liem. 2016. Go with the flow: When listeners use music as technology. (2016).

[6] J. Stephen Downie, Donald Byrd, and Tim Crawford. 2009. Ten years of ISMIR: Reflections on challenges and opportunities. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*. 13–18.

[7] Mohsen Kamalzadeh, Dominikus Baur, and Torsten Möller. 2012. A survey on music listening and management behaviours. (2012).

[8] Youngmoo E Kim, Erik M Schmidt, and Lloyd Emelle. 2008. Moodswings: A collaborative game for music mood label collection.. In *Ismir*, Vol. 2008. 231–236.

[9] Youngmoo E Kim, Erik M Schmidt, Raymond Migneco, Brandon G Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A Speck, and Douglas Turnbull. 2010. Music emotion recognition: A state of the art review. In *Proc. ISMIR*, Vol. 86. 937–952.

[10] Edith Law and Luis von Ahn. 2009. Input-agreement: A New Mechanism for Collecting Data Using Human Computation Games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1197–1206. https://doi.org/10.1145/1518701.1518881

[11] Cynthia C. S. Liem, Andreas Rauber, Thomas Lidy, Richard Lewis, Christopher Raphael, Joshua D Reiss, Tim Crawford, and Alan Hanjalic. 2012. Music information technology and professional stakeholder audiences: Mind the adoption gap. In *Dagstuhl Follow-Ups*, Vol. 3. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

[12] Adrian C North, David J Hargreaves, and Jon J Hargreaves. 2004. Uses of music in everyday life. *Music Perception: An Interdisciplinary Journal* 22, 1 (2004), 41–77.

[13] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12, Oct (2011), 2825–2830.

[14] Markus Schedl, Arthur Flexer, and Julián Urbano. 2013. The Neglected User in Music Information Retrieval Research. *Journal of Intelligent Information Systems* 41, 3 (2013), 523–539.

[15] Konstantinos Trohidis, Grigorios Tsoumakas, George Kalliris, and Ioannis P Vlahavas. 2008. Multi-label classification of music into emotions.. In *ISMIR*, Vol. 8. 325–330.

[16] Karthik Yadati, Cynthia C. S. Liem, Martha Larson, and Alan Hanjalic. 2017. On the Automatic Identification of Music for Common Activities. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*. ACM, 192–200.

[17] Yi-Hsuan Yang and Homer H. Chen. 2012. Machine Recognition of Music Emotion: A Review. *ACM Trans. Intell. Syst. Technol.* 3, 3, Article 40 (May 2012), 30 pages. https://doi.org/10.1145/2168752.2168754