EmoMTB: Emotion-aware Music Tower Blocks

Alessandro B. Melchiorre
Johannes Kepler University Linz and
Linz Institute of Technology

David Penz
Johannes Kepler University Linz and
TU Wien

Christian Ganhör
Johannes Kepler University Linz

Oleg Lesota
Johannes Kepler University Linz and
Linz Institute of Technology

Vasco Fragoso
Johannes Kepler University Linz

Florian Fritzl
Salzburg University of Applied
Sciences

Emilia Parada-Cabaleiro
Johannes Kepler University Linz and
Linz Institute of Technology

Franz Schubert
University of Applied Arts Vienna
and St. Pölten UAS

Markus Schedl
Johannes Kepler University Linz and
Linz Institute of Technology

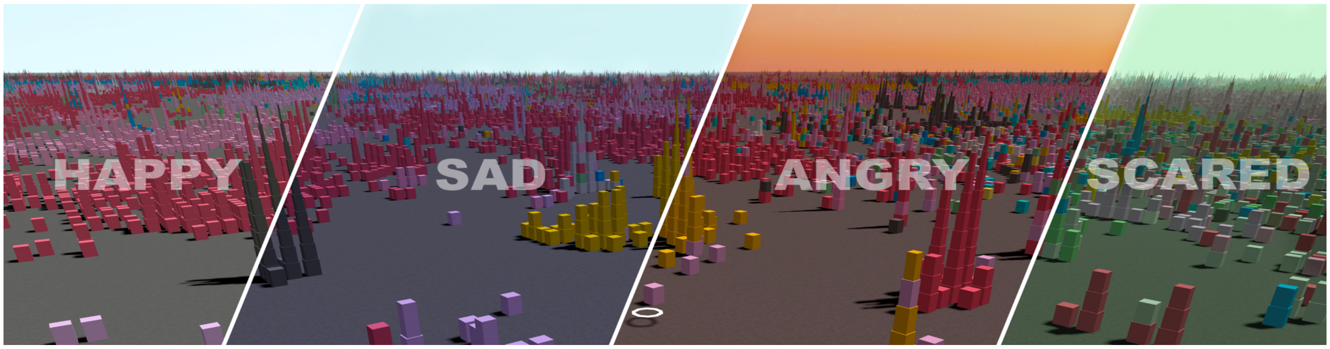


Figure 1: EmoMTB's landscape for the four emotional themes.

ABSTRACT

We introduce Emotion-aware Music Tower Blocks (EmoMTB), an audiovisual interface to explore large music collections. It creates a musical landscape, by adopting the metaphor of a city, where similar songs are grouped into the same building and nearby buildings form neighborhoods of particular genres. In order to personalize the user experience, an underlying classifier monitors textual user-generated content, by predicting their emotional state and adapting the audiovisual elements of the interface accordingly. EmoMTB enables users to explore different musical styles either within their comfort zone or outside of it. Besides, tailoring the results of the recommender engine to match the affective state of the user, EmoMTB offers a unique way to discover and enjoy music. EmoMTB supports exploring a collection of circa half a million streamed songs using a regular smartphone as a control interface to navigate in the landscape.

CCS CONCEPTS

• **Information systems** → **Users and interactive retrieval**; • **Human-centered computing** → *Human computer interaction (HCI)*; **Visualization systems and tools**.

ICMR '22, June 27–30, 2022, Newark, NJ, USA.
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9238-9/22/06.
https://doi.org/10.1145/3512527.3531351

KEYWORDS

music exploration, intelligent user interface, clustering, emotion recognition

ACM Reference Format: Alessandro B. Melchiorre, David Penz, Christian Ganhör, Oleg Lesota, Vasco Fragoso, Florian Fritzl, Emilia Parada-Cabaleiro, Franz Schubert, Markus Schedl. 2022. EmoMTB: Emotion-aware Music Tower Blocks. In *Proceedings of the 2022 Int'l Conference on Multimedia Retrieval (ICMR '22), June 27–30, 2022, Newark, NJ, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3512527.3531351

1 INTRODUCTION

Music exploration and discovery represent an essential aspect of nowadays' online music streaming services, especially considering that available music collections usually comprise several tens of million music pieces[1,2]. Users' navigation of these large musical assortments is usually aided by text-based search and recommendation systems which return a list of tracks which might be tedious to explore. Moreover, there is evidence that the main motivation for everyday music consumption is the underlying emotional component gained by such an experience [6]. Therefore, developing emotion-aware music exploration systems is an important research endeavor.

[1] https://newsroom.spotify.com/company-info/
[2] https://www.deezer.com/en/company/press

In this context, we introduce Emotion-aware Music Tower Blocks (EmoMTB), an audiovisual interface that enhances music exploration and discovery in large music collections by taking into account users' emotional states. EmoMTB follows the general idea of MTB [15], adopting the metaphor of a city to represent a music collection, where each building (tower block) is constituted of several cubes, each representing a track. However, EmoMTB provides substantially novel functionalities, technical architectures, and possibilities of interactions compared to MTB. The major novelties include: (i) Users can navigate and interact with EmoMTB's world by connecting to the interface through their personal phones and using game-like controllers; thus, providing a more immersive experience. (ii) EmoMTB fetches a personalized recommendation list for the user and enables the user to locate and 'jump' to the recommended blocks in the landscape; thus, encouraging users' exploration and discovery of the surrounding tracks beyond 'traditional' recommendations. (iii) The whole audiovisual interaction is driven by the user's emotional state (either manually chosen or automatically inferred); thus, making music discovery a unique personalized experience.

EmoMTB has been showcased at the Ars Electronica Festival 2021.[3] A teaser video of the exhibit is available from https://www.youtube.com/watch?v=JKgAlWObc-0.

2 RELATED WORK

For a comprehensive survey of intelligent user interfaces for music discovery, please refer to Knees et al. [7]. Most related music user interfaces create a spatial arrangement of musical collections. Early approaches include *Islands of Music* [11] and *nepTune* [8]. Both interfaces reorganize the music tracks of a collection according to their audio features, where similar tracks are clustered together forming 'islands' that raise from the ocean (sparse zones).

Similarly, *Music Galaxy*, an adaptive user interface that visualizes a music collection as a music galaxy, has been proposed by Stober and Nürnberger [16]. The arrangement of the stars (music tracks) is dictated by a distance metric over the audio features, which can be manipulated and adapted to the taste of the users.

The concept of mood-focused exploration of music collections is exploited by Vad et al. [17], whose work employs mood-related descriptors extracted from the music audio. Differently, in EmoMTB we assign an emotion to a track based on the overall sentiment extracted from user-generated of the music social network Last.fm [4] on the specific track extracted from text, which provides a novel emotional dimension for music exploration.

Similarly to MTB [15], EmoMTB is mostly built on user-generated data, such as genre-assigned tags and emotions inferred from microblogs. However, EmoMTB extends MTB by (1) providing a list of recommendations tailored to the users' taste and helping them to travel the vast city; and (2) by adapting the interface and the recommendations to the current emotion of the user. These two features make EmoMTB distinct from previous works as it brings together the concepts of spacial and mood-aware music exploration.

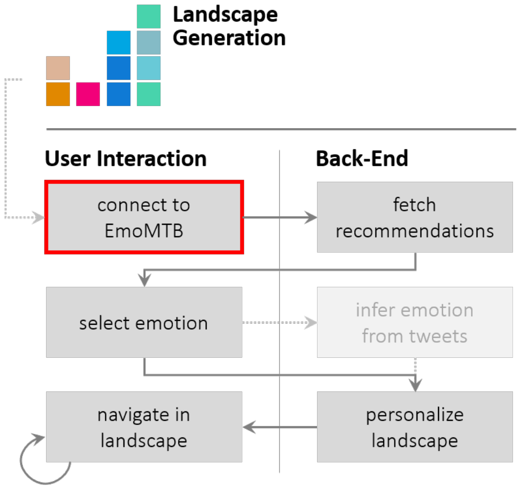


Figure 2: Onboarding procedure of the user to enjoy EmoMTB

3 FUNCTIONALITY AND IMPLEMENTATION

The EmoMTB interface provides an immersive experience of music exploration and discovery. Our interface offers two interacting channels to the users: (1) a large screen to showcase the main landscape built from our data; and (2) the user's mobile phone for options and controls to navigate through the landscape (see Section 3.2 for more details). With the mobile phone as the main source of interaction, the user follows an initial setup procedure to get started. This user flow is shown in Figure 2.

Once the landscape has been generated and the user has connected to EmoMTB, our application proceeds to fetch personalized recommendations using the Spotify API. Subsequently, the user has the choice to select one of four emotions that will influence the user experience (see Section 3.3) or decide to use the emotion inferred from Twitter.

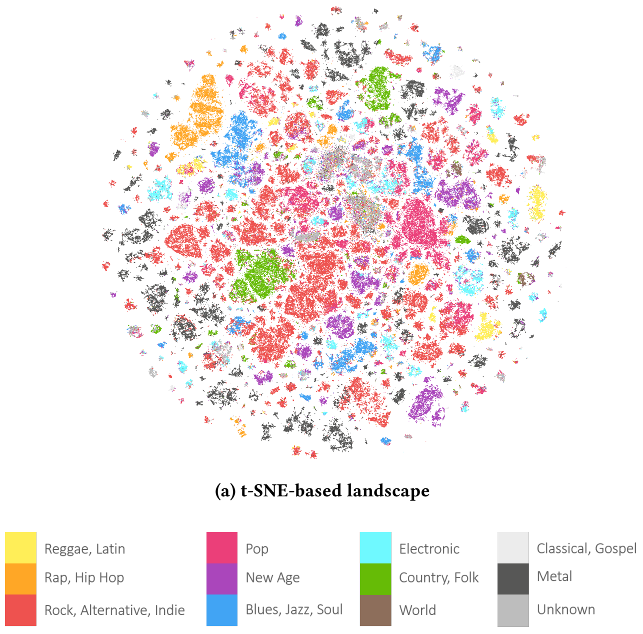
The EmoMTB interface is created based on the LFM-2b dataset [9, 14] containing information about circa 51 million tracks including metadata and community-assigned tags. We further enhance these tracks with acoustic features (see Section 3.1) from Spotify's APIs.[5] After this step, we use the 436,064 tracks successfully matched to Spotify to build our interface.

3.1 Landscape Generation

EmoMTB's landscape is comprised of colorful track-blocks clustered using the t-student distributed stochastic neighborhood embedding (t-SNE) [18]. To assign a color to each track/block, we use its genre and delineate a genre-color mapping based on the results from the user study presented by Holm et al. [5]. The resulting landscape and the color-mapping are shown in Figure 3a and 3b, respectively. As we can see from the landscape, the tracks form several neighborhoods delineated by the different colors/genres, and in some cases, they overlap e. g., rock (red) and metal (black) clusters.

As input for the t-SNE algorithm, each track is represented by both fine-grained genres and audio features. As for the former, we extract the genre information of the tracks from their Last.fm

[3] https://ars.electronica.art/newdigitaldeal/de/music-tower-blocks/
[4] https://www.last.fm/
[5] https://developer.spotify.com/documentation/web-api/



(a) t-SNE-based landscape

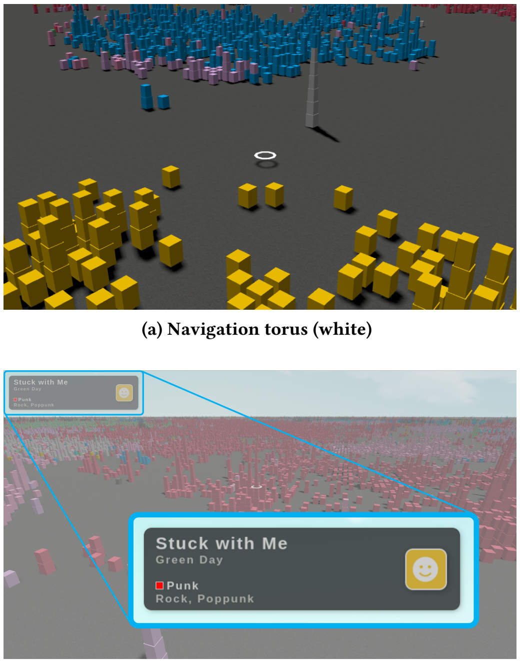
(b) Genre-color mapping

Figure 3: EmoMTB map

community-assigned tags which are available in the LFM-2b dataset. To do so, we filter the tags by matching them against the large EveryNoise[6] genre collection, resulting in 2,374 fine-grained genres. Each track is then represented by a TF-IDF vector using as term frequency the Last.fm tag weights and as document frequency the number of tracks that the genre tag was assigned to.

Similar to MTB [15], we further enrich the available tracks by fetching audio feature from Spotify. In particular, we use: *Acousticness* (probability that a song is acoustic), *Energy* (intensity and activity), *Speechiness* (presence of spoken words), *Instrumentalness* (probability of not containing vocals), and *Valence* (probability of the track conveying positiveness). This leads to a total of 2,379 features per track (TF-IDF genre weights and acoustic features).

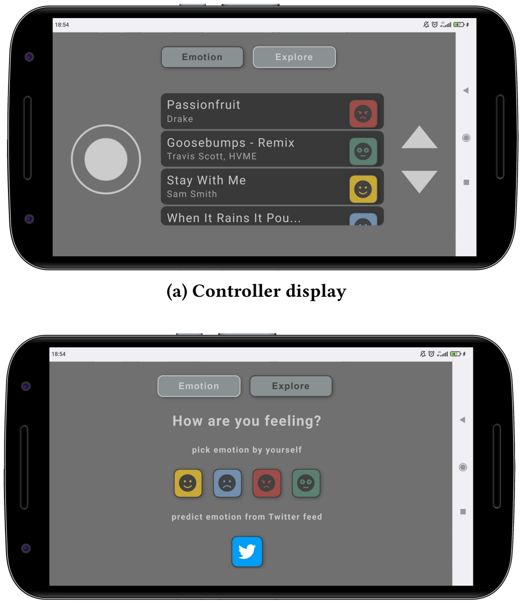
We then apply Principle Component Analysis (PCA) before using t-SNE to generate the final projection. We do so by selecting a number of components (i. e., 405) that covers 95% of the explained variance, resulting in compacted representations of the tracks. Finally, we compute t-SNE (using a perplexity of 45) based on the PCA-processed features, which projects the tracks to a 2-dimensional coordinate space. We further discretize these coordinates and represent each track as a colored box in the interface. Moreover, tracks placed at highly similar coordinates are stacked on top of each other while being sorted based on their popularity, with the most popular being on top. This resembles the metaphor of more important people in a company building occupying offices on higher floors.



(a) Navigation torus (white)

(b) Track information

Figure 4: Landscape visualization



(a) Controller display

(b) Emotion selection

Figure 5: User's phone interface.[7]

3.2 Navigation and Interaction

To navigate EmoMTB, the user controls a white hovering torus (cf. Figure 4a) which enables exploring the landscape and selecting tracks to play. While the user's torus is hovering over a track/block, the track's name, artist, corresponding genre, and predicted emotion (see Section 3.3) is displayed in the upper left corner of the visualization (cf. Figure 4b). By staying on a block for a moment, the playback of its corresponding track starts and continues until

[6] https://everynoise.com/

[7] Android Emulator Skins - Copyright 2012, Ming Chen

the user either moves the torus to an empty space of the landscape or onto another block.

To explore the landscape, a controller interface, specifically designed to run on the browser of the user's smartphone, is used (cf. Figure 5a). Besides navigating the landscape, the smartphone interface also allows to manually select an emotion (cf. Figure 5b). On the left, the navigation display presents a joystick control that allows both moving the torus and rotating the visualization. In order to allow the user to easily navigate the vast map of EmoMTB, the controllers are enhanced with linear acceleration, i. e., if the user moves ahead for a while, the torus will begin to accelerate until it reaches a maximum velocity. On the right side of the controller, there are two arrow buttons, one pointing up and the other down (elevator-like). They offer the possibility to move vertically along the towers made of blocks. Finally, in the middle of the screen, the user's personalized recommendations are shown in a scrollable list. With a click on a recommended track, the visualization transports the user to the position of the track in the landscape, through a smooth animation.

3.3 Emotion Selection

EmoMTB's components for emotion extraction and matching are based on Ekman's 'Big Six', a model that defines emotions as discrete classes [4]. A categorical model was preferred over a dimensional one [13] due to the higher familiarity of the general public with emotional words, such as joy, than with dimensional concepts, such as valence. From the six categories identified by Ekman, four were selected: happiness, sadness, anger, and fear. Surprise and disgust, since not unequivocally accepted as basic emotions [10], were discarded, thus reducing the complexity of the user interaction. Despite the limitations of referring to basic emotions when investigating affect in music [3], the same four emotions used to model participants' affect were also adopted to create the emotional mapping of the tracks. This was considered the best solution as the domain-specific models tailored to assess emotions in music [19] are different to those typically used to model human affect [4].

Users' state could be manually chosen or automatically inferred. For the manual selection, users would choose one of four emojis in the mobile user interface (cf. Figure 5b). When the users' state is automatically inferred, this is considered as the global emotion of the 'crowd' extracted from the visitors' posts on the EmoMTB Twitter feed. The tracks' underlying emotion is instead inferred from the Last.fm user-generated tags associated with them. To determine the crowd's state, the emotion is selected based on the predominant category of the last 3 tweets.

To automatically infer users' and tracks' emotions, and match them to the four target categories, a Multilayer Perceptron Classifier, trained on several datasets of labeled social media posts for sentiment analysis [1], is used. By using the standard emotion extraction tool VADER from NLTK module,[8] and the emotional lexicon ANEW[2] (Affective Norms for English Words), we extract feature vectors from both the tweets and the tags in a Bag-of-Words fashion. These feature vectors are used to automatically infer users' and tracks' emotions, respectively.

Once the system detects the users' emotion (either automatically inferred or manually selected), a matching between users' and tracks' emotions is performed by reordering the recommendation list (cf. Section 3.4). Besides affecting the users' recommendations, the resulting emotion also influences the appearance of the generated landscape, i. e., the theme. Next to the sky (whose color reflects the target emotion is selected according to previous research outcomes [12]), also the light's intensity is adapted. As displayed in Figure 1, while happiness is associated with a bright blue sky (like a nice summer day), fear leads to an eerie landscape with poor light (like in a scene of a horror movie).

3.4 Recommendations

In the mobile user interface of EmoMTB, the user's music exploration is further supported by a list of personalized track recommendations. The user is able to click on the displayed recommendations to move their position within the EmoMTB landscape to the respective song and play it.

To obtain the personalized recommendations, we require the user's permission to access the respective Spotify profile. We then fetch up to 200 recommended tracks using the respective API, based on the top tracks the user listened to. This list of recommendations is further mapped to the tracks/blocks in our dataset. Initially, the recommendations are sorted based on the order returned by the Spotify API. Once an emotion is identified by the system (either automatically inferred from Twitter or manually selected by the user), the list of recommended tracks is sorted according to the emotional state, tracks with the selected/predicted emotion going on top of the list.

4 CONCLUSION

We introduced EmoMTB, an immersive audiovisual interface that enables users to discover new music by exploring a city-like landscape of tracks/blocks organized into buildings and neighborhoods (districts). The users' experience is further enhanced by an affect-driven component that adapts the interface and tailors the recommendations to the users' (inferred or selected) emotions. EmoMTB was presented at the Ars Electronica Festival 2021 and received very positive feedback from the visitors, along with some suggestions for further extensions. Among these, several visitors asked whether the layout of the city could be tailored to their own music preferences and if the blocks/neighborhoods might be moved or modified by users. Following this, a promising direction to pursue in the future would be to further develop the interaction capabilities in order to allow users to change the landscape or even to build up their own very personal cities. In addition, turning EmoMTB into a *collaborative* music exploration platform where different users are depicted as avatars and see each others' actions might make EmoMTB even more appealing. So would integrating support to create playlists, e. g., visualized by tramway tracks in the city.

ACKNOWLEDGMENT

EmoMTB received financial support by the State of Upper Austria and the Federal Ministry of Education, Science, and Research, through grant LIT-ARS-2020-015; and by the Austrian Science Fund (FWF): P33526. We further thank Peter Knees and Michael Mayr for their help with the first version of the interface.

[8] www.nltk.org/_modules/nltk/sentiment/vader.html

REFERENCES

[1] Francisca Adoma Acheampong, Chen Wenyu, and Henry Nunoo-Mensah. 2020. Text-based emotion detection: Advances, challenges, and opportunities. *Engineering Reports* 2, 7 (2020), e12189.

[2] Margaret M Bradley and Peter J Lang. 1999. *Affective norms for English words (ANEW): Instruction manual and affective ratings.* Technical Report. The center for research in psychophysiology.

[3] Julian Cespedes-Guevara and Tuomas Eerola. 2018. Music communicates affects, not basic emotions–a constructionist account of attribution of emotional meanings to music. *Frontiers in psychology* 9 (2018), 215.

[4] Paul Ekman. 1999. Basic emotions. In *Handbook of cognition and emotion*, T. Dalgleish and M. J. Power (Eds.). Vol. 98. John Wiley & Sons Ltd, New York, NY, USA, 45–60.

[5] Jukka Holm, Antti Aaltonen, and Harri Siirtola. 2009. Associating colours with musical genres. *Journal of New Music Research* 38, 1 (2009), 87–100.

[6] Patrik N Juslin and Petri Laukka. 2004. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of new music research* 33, 3 (2004), 217–238.

[7] Peter Knees, Markus Schedl, and Masataka Goto. 2019. Intelligent User Interfaces for Music Discovery: The Past 20 Years and What's to Come.. In *ISMIR*. 44–53.

[8] Peter Knees, Markus Schedl, Tim Pohle, and Gerhard Widmer. 2007. Exploring music collections in virtual landscapes. *IEEE multimedia* 14, 3 (2007), 46–54.

[9] Alessandro B Melchiorre, Navid Rekabsaz, Emilia Parada-Cabaleiro, Stefan Brandl, Oleg Lesota, and Markus Schedl. 2021. Investigating gender fairness of recommendation algorithms in the music domain. *Information Processing & Management* 58, 5 (2021), 102666.

[10] Andrew Ortony and Terence J Turner. 1990. What's basic about basic emotions? *Psychological Review* 97 (1990), 315–331.

[11] Elias Pampalk, Andreas Rauber, and Dieter Merkl. 2002. Content-based organization and visualization of music archives. In *Proceedings of the tenth ACM international conference on Multimedia*. 570–579.

[12] Matevž Pesek, Gregor Strle, Alenka Kavčič, and Matija Marolt. 2017. The Moodo dataset: Integrating user context with emotional and color perception of music for affective music information retrieval. *Journal of New Music Research* 46, 3 (2017), 246–260.

[13] J. A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39 (1980), 1161–1178.

[14] Markus Schedl, Stefan Brandl, Oleg Lesota, Emilia Parada-Cabaleiro, David Penz, and Navid Rekabsaz. 2022. LFM-2b: A Dataset of Enriched Music Listening Events for Recommender Systems Research and Fairness Analysis. In *Proceedings of the 7th ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR 2022)*.

[15] Markus Schedl, Michael Mayr, and Peter Knees. 2020. Music Tower Blocks: Multi-Faceted Exploration Interface for Web-Scale Music Access. In *Proceedings of the 2020 International Conference on Multimedia Retrieval*. 388–392.

[16] Sebastian Stober and Andreas Nürnberger. 2010. MusicGalaxy–an adaptive user-interface for exploratory music retrieval. In *Proc. of 7th Sound and Music Computing conference (SMC'10)*.

[17] Beatrix Vad, Daniel Boland, John Williamson, Roderick Murray-Smith, and Peter Berg Steffensen. 2015. Design and evaluation of a probabilistic music projection interface. (2015).

[18] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).

[19] Marcel Zentner, Didier Grandjean, and Klaus R Scherer. 2008. Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* 8, 4 (2008), 494.