Emotion Based Music Recommendation System

Mikhail Rumiantcev, Oleksiy Khriyenko
University of Jyväskylä
Jyväskylä, Finland
mikhail.rumiantcev, oleksiy.khriyenko@jyu.fi

Abstract— Nowadays, music platforms provide easy access to large amounts of music. They are working continuously to improve music organization and search management thereby addressing the problem of choice and simplify exploring new music pieces. Recommendation systems gain more and more popularity and help people to select appropriate music for all occasions. However, there is still a gap in personalization and emotions driven recommendations. Music has a great influence on humans and is widely used for relaxing, mood regulation, destruction from stress and diseases, to maintain mental and physical work. There is a wide range of clinical settings and practices in music therapy for wellbeing support. This paper will present the design of the personalized music recommendation system, driven by listener feelings, emotions and activity contexts. With a combination of artificial intelligence technologies and generalized music therapy approaches, a recommendation system is targeted to help people with music selection for different life situations and maintain their mental and physical conditions.

I. INTRODUCTION
A lot of research has been done with respect to music-driven influence on the physiological and emotional state of a human. Humans perceive a variety of feelings from different types of music and from ancient times considered music influence in the formation of a personal character and ability to treat diseases [1]. Music listening has a significant impact on human feelings, thoughts and as a result, it influences mental and physical health, and the topic of music wellbeing support is gaining popularity. A lot of measurements and research has been conducted to understand the impact of music on brain activity [2]. Musliu et al. [3] explain the impact of music on memory. Music therapy is considered as an effective enhancement to standard care in the treatment of depression [4], [5].

Along with rational decision-making, emotions aspect makes a significant impact on driving decisions [6]. Emotions aware recommendation system would be able to better understand people’s requirements and feelings and select appropriate music pieces according to the emotional context. Music-related emotions are classically considered from two main perspectives: emotions which can be observed in music (cognitivist perspective) and emotions that are perceived from music (emotivist perspective). Vempala and Russo in their research [7] compared correlations between music and these two alternative kinds of emotions. They trained neural networks for both perspectives with music features using as inputs, outputs of these models were arousal and valence-based emotions which were observed by music analyses and by psychological feedback received from research participants. Results showed that in both cases networks showed similar results for arousal, for valence the cognitivist perspective networks performed better. Even taking into account the fact that it is impossible to separate features, which influence more to emotions perceived from music from others that correlate more with emotions in music, the research clearly illustrates the possibility of using this method and possibilities for further development.

All processes in the human body are closely interrelated, therefore emotions, psychical and psychophysical conditions might have an impact on each other. The cardiovascular system is significantly influenced by stress [8]. Some sorts of music effects on heart rate, blood pressure and other psychophysical conditions as well [9]. Ellis and Thayer [10] draw our attention to the fact that different music attributes such as tempo or beat level can trigger emotional, psychophysiological and behavioral effects.

Music recommendation can be applied in different areas such as support of intellectual and physical work, studying, sports, relaxing, stress and tiredness destruction, music therapy and many others.

In this work we present the design of the personalized emotion-driven music recommendation system. Principal purposes of the recommender are: addressing the choice problem, exploring new music pieces, support mental and physical wellbeing and support in improving working processes. The design involves a combination of artificial intelligence techniques and generalized, music recommendation and therapy approaches. This paper clarifies approach of applying emotion-driven personalization while music recommendation process.

The following section describes application cases of the recommendation system and the problem domain clarification. The third section includes the system performance description and approaches used at the system architecture. The fourth section illustrates the experimental prototype of the recommendation system. Research results, limitations and further work are presented in the final part of this paper.

II. USE CASES
Taking into account research referred to in the previous section, sets of music-related attributes and their values are...
recognized concerning the human emotional and physiological states they might cause or lead to. In this section, we are going to review the use cases in which the music influence on human well-being can be applied.

1) Intellectual work and study: Maintaining energy, vitality, spirits, and freshness of the brain and sharpness of attention is important for the productivity of the intellectual work and studying. Studies [11] have shown that almost half of the research respondents believe that music improves their concentration during studying, while others highlighted that music helps them to keep their mind calm and prevent from sleepiness. When the person is awake and cheerful, the work proceeds more efficiently. Music listening in this particular case should be directed to increase arousal and help to refresh form tiredness faster. Key criteria of the personal condition evaluation during and after the working process are tiredness, arousal, satisfaction from the process, productivity results. Our approach implies capturing indicators of mentioned criteria before, during and after listening session and match them with attributes of listened music. Of course, many other factors might be presented which have influence on people while doing intellectual work, to support this process with music efficiently, the system needs to be aware of other factors rather than music curation distracted from performing tasks at a sufficient level.

2) Physical work and sport: Activities might vary in requirements such as speed and stamina. Psychological and physical conditions define human wellbeing and performance. In sports, there are long-standing stable practices of health support that involve deep medical analyses, measurements and wellbeing monitoring. In adopting these practices, our approach of personal support through music listening can significantly enhance regulation methodologies in sport. Listening of the preferred music does not bring much effect in performance during highly intensive repeated sprint exercises; however, it boosts the motivation and decreases overexertion [12]. Music listening has a significant effect on performance but does not decrease perceived exertion on 1.5 miles running exercises [13]. Taking into account the results of these studies we can find out that music has a different influence in heterogeneous sport activities. People prefer fast tempo music for anaerobic exercises and slower music for exercises targeted to strength and stamina training, at the same time individual choice factor is important of the music selection [14].

3) Personal safety: Everyday people face situations when they need to keep the attention sharpened on critically important things to avoid life and health threats. For example, drivers have to be focused on traffic, drowsiness or sickness might lead to damaging and other implications. Jeon in his research [15] describes how music can be used to mitigate affective effects while driving. Pedestrians are in the risk group, particularly when they cross the street. In both cases, nothing should distract people from traffic, including music. The recommendation system has to help to be revived and feel full of energy and prevent destruction from the surrounding environment.

4) Music therapy: If Music is widely used to support well-being, treat stress and distract people from their diseases [4]. Nowadays, a wide range of clinical settings and comprehensive music therapy practices is used to support mental and physical health [16]. Music listening promotes relaxation in daily life and is efficient in mood regulation [17], [18]. Listening styles and music preferences might reflect the personality, life difficulties, and stress perceptions [19]. Listening to appropriate music can help with blood pressure stabilization and stress treatment [20]. With respect to the external supervision of medical experts, we aim to drive our recommendation system to perform music curation and help to detect mental disorders at early stages.

III. SYSTEM PERFORMANCE DESCRIPTION

Music influence on human wellbeing can be used across various areas; the use cases presented in the previous section demonstrate the possibility of the practical application of music influence on humans. To address these and other possible cases there is a need of having a generalized system with reflect personal features of each particular user. This section targets design the solution that is aimed to recognize personal aspects of the physical and emotional influence of music-related features in various contexts and combine them with well-known generalized approaches.

Regardless of an actual purpose, whether there is a need to change an emotional state of a user or maintain it and keep it the same, the main function of the system is to search for the nearest (closest) music tracks to the abstract etalon one, which is defined by a certain set of music related attributes.

We may highlight several action modes of our system. The first one is a straightforward approach when we just specify the final destination - the desired point in the emotional space of a user. In the music-driven emotion treatment context, this point has a corresponding vector of various music, person and context related features with their values. In this case, using corresponding distance measurement function, the system searches for nearest/closest music tracks. Another, advance mode of the system applies smoother, less aggressive transition towards the desired point, taking into account the current position of a user in the emotional space. In this case, systems step by step reduce the distance between feature vectors that represent current and desired points. This mode supposes to be less stressful and obtrusive for a user. At the same time, it requires more sophisticated algorithm for multidimensional (multi-featured) “delta” distance (step) calculation, taking into account the personality of each particular user. Therefore, we have to capture Personalized Emotion Transformation Model (PETM) for each user during data collection and model creation stage. In both cases we have to maintain the model during the further performance of the system to always fit an actual current personality of a user. Thus, continuous data collection and processing user feedback will support model training on the fly and enrich personalization in recommendations.

At the first stage, the system builds a general user profile (GUP) based on data obtained via the initial survey. Further, based on the GUP system builds an initial Music-driven
Emotional Model (MEM) of a user. This model could also be aggregated with an average model that presents the nearest cluster of other users, which are clustered based on their GUPs.

Since we aim to design a more efficient model, we emphasize on personalization and adaptation of the system. Thus, the second stage of the system is a stage of further personalization/tuning of the initial Music-driven Model of a user.

To perform transitions of the personal conditions in different contexts each personalized model must determine what types of music can lead to specific state changes. For this reason, items with particular sets of music attributes should be labeled with values that describe the most probable impact from the listening experience. These labels and music pools have to be tuned and unique for each personal profile. The method implies the learning of the model how to recognize music attributes and how they influence to a particular person.

A. Data collection

To further explore the relationships between human feelings and music, the system requires comprehensive information about the listening experience and music nature. In this section we describe what kind of data is needed and by which possible ways it can be collected (see Fig 1.).

Fig. 1. Recommendation System Ecosystem

1) MuPsych [21] research tool: This data collection tool is an application that captures and collects samples of music listening experience on mobile devices. It is targeted to investigate the psychological transitions of a person perceived from listening to music [22]. The mobile application is integrated with music platforms such as Spotify, it captures metadata and attributes of listened music tracks and triggers surveys during the listening session to ask listeners about their listening reasons, activities, feelings and perceived emotions. MuPsych provides extensive information reflecting music emotions that is a solid ground for the emotion-driven music classification particularly in early stages when the recommendation system faces the interaction gap with listeners.

2) General Personal Data: Context independent data that could be collected from the user directly via MuPsych application (MuPsych App) based surveys or via gathering data from various external services (e.g. Social Network personal profile, etc.). In this case, MuPsych App is considered as a data collection module that populates Personal Data Space of a user with corresponding information and, at the same time, gets permission from a user to access external services with the corresponding authentication.

3) Sensor-based Personal Data is collected by various sensors and devices with respect to the psychological and physiological parameters of a user. It could be GPS based positioning, body movement, face and voice-based mood pattern. Contemporary monitoring systems can track heart and
respiratory rate variability, oxygen arterial saturation, and consumption, liquid intake, sleep quality, number of steps, distance traveled, type of physical activities and many others. That could be achieved by analyses of the data obtained by various sensors such as accelerometers, gyroscopes, magnetic field, galvanic skin response and temperature sensors. Monitoring devices are presented in various forms, smartphones smartwatches and wrist fitness trackers are becoming ordinary objects for most people, some devices are attached to the body, for example, chest belts, contemporary technologies allow to design clothes with in-built textile-based sensors. Devices interconnected with nearby processing units can share the data for deeper health observations. Continuous monitoring can be performed remotely, ergonomically and in a relaxed atmosphere from one side, from the other it can help to diagnose various health problems in early stages. [23] Since, mobile phones are equipped with some sensors, mobile application of the recommendation system could also play a role in sensor-based data collection tools. In most of the cases, sensor-based data is a raw data that could be used to reason about more upper-level data such as activity, surrounding environment, emotional state, etc. Therefore, to make our system smarter and less obtrusive, we have to increase the usage of externally and automatically collected data.

4) External Supervisor driven Data: describes a used from the perspective of external supervisors, such as parents or family members, teachers, therapists, etc. To collect this data, we need an extra front-end application for survey-based data collection, or to utilize a mediator that allows data collection via existing communication channels (Social Network-based application or other external services).

5) Upper-level Data Reasoning: modules populate Personal Data Space with application-specific data taking into account all suitable data that is available in the data space and data, which is accessible via DBPedia [24] or other information services. At the same time, these modules might initiate retrieval of extra data, if a confidence level of the reasoned value is low due to lack of fresh data, or due to specific conditions that minimize the influence of available data on the final outcome of the module. For example, a module that reasons about the current emotional state of a user takes into account relevant data provided by the user via self-reporting channel, user’s heartbeat rate, and emotional state description provided by external supervisor (teacher). It might happen, that in some circumstances (based on some contextual information), the algorithm applies conditions that minimize the trust level of user’s self-reported data (lowering influence of this source of information). At the same time, the system might not have updated value from more reliable (in this context) external supervisor, and its old value does not make any sense. Additionally, we might also be not able to fully rely on high heartbeat measurement influence, since we might be unconfident with the current activity of a person (e.g. whether (s)he is doing some sports or not). Finally, emotion state reasoning algorithm might return us a possible value with low confidence level and it will require additional retrieval of necessary data (e.g. to submit a request to external supervisor) or/and retrieval of extra data that, for example, cannot be constantly updated in the user’s Personal Data Space due to some limitations (e.g. limited amount of API based requests to the remote service). An example of such extra data could be an analysis of emotional footprints of recent messages sent by users via social communication channels (e.g. Facebook), or emotion-oriented face and voice recognition applied on top of recently captured image/video/sound.

6) Under an Upper-level Data: we might consider various contexts such as Activity, Surrounding Environment, etc. Values for these contextual parameters might be gathered from the user directly, as well as reasoned based on relevant data. For example, speed (calculated based on GPS data) might indicate that the user is traveling. Depending on the speed range and map data, the system may also reason about the type of vehicle user is traveling by. Certain speed range and temporal data (season) in combination with heartbeat measurements might indicate that the user is doing some sport activities (running, biking, rowing, skiing, etc.).

7) Music-related Features and Metadata: are retrieved by feature detection module itself or via external music/audio feature detection tools (e.g. The Echo Nest [25], etc.), and collected from music content providers (e.g. Spotify[26], Pandora, Google Play, Beats Music, SoundCloud, etc.) or other external music-related information services (e.g. MusicBrainz [27], BDTune [28], etc.). This data includes the physical features of an audio track and the meaning of corresponding lyrics. At the same time, taking into account an emotional specific of the offered system, indirectly related information, such as the meaning of an associated video track (video clip(s)) or some event(s) closely associated with the audio track, could be also considered as personalized track related data. In this case, we definitely will require a more advanced and sophisticated data-retrieving tool.

8) Personal Data Space: is constantly populated with updated values of raw data as well as upper-level data provided by corresponding reasoners. These data can be updated automatically in the background, according to a defined scheduling logic, and/or could be refreshed on-demand. Therefore, direct consumption of data available in Personal Data Space may speed up and facilitate the performance of the system [29]. To be easily compatible and interoperable with other systems, Personal Data Space will follow commonly accepted W3C standards with respect to Semantic Web [30] and Linked Data [31], [32] principles. The internal data model will be built as an extension of the widely used ontologies (Music Ontology [33], Audio Features Ontology [34], Emotion Ontology [35], etc.). Being organized as a remote service and supported with the corresponding API, Personal Data Space could be utilized by other services and applications. It will facilitate the dissemination of some of the project’s results, as well as, could be considered as an extra outcome of the project.

9) Interactive user feedback: Listening behavioral factors during music listening sessions play a significant role in determining preferences and emotional states. Explicit interactive feedback involves liking or disliking and music track rating. Along with that, we can get the feedback through an implicit way by analyzing actions performed during the listening session: track listening duration (was it listened fully or partly), replaying, skipping, forwarding, back warding and downloading. Music search history and strategy allows building a personal
preferences picture. Such listening experience analyzes drive the system decision about user preferences and help to avoid or at least to reduce relying on obtrusive surveys and feeling of feedback forms. In such a way the system looks more natural and does not look like an imposed treatment mechanism.

B. Classifying

The primary objective of the system is selecting music tracks to change and maintain the psychophysical states of humans. Music tracks should be classified with labels that describe emotional, psychological and psychophysiological states and transitions. Therefore, we have to solve the problem of classifying for the content-based filtering (see Fig 2.). Of course, to achieve more accuracy in deep neural network performance the datasets should be large enough. Other models such as k-nearest neighbors and random forest can also be considered for this task and used for the cross-validation purpose. Random forest sows more accuracy in music genre classifying based on music features in comparison with k-nearest neighbors. For more precise judgments and selection of the particular model the system should be tested with large enough real data sets.

C. Incremental learning

The system needs to adapt to the dynamic nature of the real-world settings and extend the knowledge without retraining of the existing model. To achieve deeper personalization, more accuracy and perform real-time recommendations Reinforcement Learning (RL) can be used as a tool to support incremental learning of the neural network model. This approach is grown in use to enrich recommendations of multimedia items in the interactive environments, as an example let’s consider the emotion aware video recommendation system [37], RL is applicable to the dynamic music playlist generation [38], [39].

The operation principle of the RL is performing actions based on policies that have been built with respect to the environment and the operational history. Markov Decision Process is a generalized model of the RL, which includes the following components: a set of states within the environment, actions which expected to be performed by the agent against the environment, which drives the ongoing actions and the feedback or reward function which tunes the policy and directs actions. In our case the contextual listening experience with music corpus can be considered as an environment. The recommendation system takes the agent role and performs actions on the selection of music items in the model considered as states. The reward function is driven by the sensor data, surveys and the user behavioral feedback (like, dislike, skip, forward, backward, rating, listening time, etc.)

D. Real-time data processing

The recommendation system involves a lot of analyses to select useful objects with respect to appropriate context in large space of options. The system needs to be aware of listening experience events, their duration, preferences, feedback, and physical and psychophysical fluctuations. To achieve deep personalization, the system has to collect and process large amounts of individualized events in real-time. There is a need for a distributed system to support large data streams with a growing number of users. Apache Kafka is the message stream platform which can be used as a tool to bind client applications and data processing units, it has high throughput and low latency in real-time data streaming.

IV. EXPERIMENTAL PROTOTYPE

To validate the recommendation model, we elaborated the trial prototype of the system. The core of the system is a web service which receives feedback about listening sessions and music attributes, and classifies music tracks for further recommendations. The mobile application consists of the music player integrated with MuPsych tool and the Spotify platform. Users can listen to music with the player or in Spotify, it does not affect the data collection and recommendation process.

At first stages the model training process of the experimental prototype is mainly relying on the data collected by the MuPsych tool. Users can keep full anonymity, the
application only offers basic information and personality surveys to bind a person to an appropriate user group to tune more accurately further recommendations. The contextual data of the music listening is based on emotional state, listening to reason and the activity. The mobile application provides questions related to the mood and the listening reasons when the user starts playing music, after 5 minutes and at the end of listening. At the same time, the MuPsych tool captures music attributes of listened music. All the data with the interactive user feedback is collected by the data processing engine. Based on the datasets already collected by the MuPsych and continuously updated data from the user side, the music recommendation system performs classification of music tracks with respect to their properties, user clusters, contextual data and listening reasons.

At the second stage, the system provides generalized playlists which are created with respect to the mood and activity context for an appropriate audience group. When the system learns more about the user, it continuously classifies tracks and tunes the personalized recommendation model. Fig. 3 shows the graphic user interface of the mobile application, it involves interactive listening with the feedback, which effects on further playlist creation. Trial experiments showed that the core music attributes such as energy, valence, tempo, and loudness in generated playlists have sufficient matching with attributes of music tracks from MuPsych datasets for the particular mood and activity contexts. Therefore, we can conclude that the system selects music with similar attributes to those, which were explored during emotional state transitions during listening. There are needed further elaborations, comprehensive experiments and validations of the system to judge how recommendations effect mood and match preferences.

V. CONCLUSION AND FUTURE WORK

Paying attention to various factors, such as particular context, personal parameters, feelings and emotions, is highly important to a decision-making process of recommendations. Contemporary music recommendation systems face the gap in personalization, human feelings, contextual preferences and emotional factors while suggesting music. In this paper, we proposed emotion-driven recommendation system with respect to personalized preferences and particular life and activity contexts. The approach presented in this study is targeted to provide maximum benefits for people from the music listening experience. It is important to make the system aware of how it is doing the recommendations, to continuously improve the music selection. By feeding the data from various sources, the system is aimed to listen to each particular user and understand their purposes of listening, feelings and contextual preferences to select the best-suited music pieces for them. We observed what kind of data is needed for the recommendation system and how it can be fetched. Main data processing tools are clarified in the scope of this paper and the experimental prototype has been elaborated. However, to achieve maximum accuracy in predictions and make them more or less relevant, machine learning systems require a large amount of the data to train the models. At this moment the data collection is in active process. At the same time this kind of system requires significant clinical research and collaboration with psychologists to tune and test the model for real recommendations and reduce possible associated risks. Further work on the implementation and testing of the recommendation engine, empirical experiments and impact evaluations are considered for the next step when the appropriate amount of the data will be collected. Music creation by artificially intelligent systems with particular music attributes to move states of human emotions can be considered as the further elaboration work in this context.
REFERENCES