

Static and Dynamic Measures of Active Music Listening as Indicators of Depression Risk

Aayush Surana¹, Yash Goyal¹, Vinoo Alluri¹

¹Cognitive Science Lab, International Institute of Information Technology Hyderabad

{aayush.surana, yash.goyal}@research.iiit.ac.in,
vinoo.alluri@iiit.ac.in

Abstract

Music, an integral part of our lives, which is not only a source of entertainment but plays an important role in mental well-being by impacting moods, emotions and other affective states. Music preferences and listening strategies have been shown to be associated with the psychological well-being of listeners including internalized symptomatology and depression. However, till date no studies exist that examine time-varying music consumption, in terms of acoustic content, and its association with users well-being. In the current study, we aim at unearthing static and dynamic patterns prevalent in active listening behavior of individuals which may be used as indicators of risk for depression. Mental well-being scores and listening histories of 541 Last.fm users were examined. Static and dynamic acoustic and emotion-related features were extracted from each user's listening history and correlated with their mental well-being scores. Results revealed that individuals with greater depression risk resort to higher dependency on music with greater repetitiveness in their listening activity. Furthermore, the affinity of depressed individuals towards music that can be perceived as sad was found to be resistant to change over time. This study has large implications for future work in the area of assessing mental illness risk by exploiting digital footprints of users via online music streaming platforms.

Index Terms: acoustic features, emotion dynamics, naturalistic music consumption, depression, lastfm

Depression is classified as a mood disorder. It may be described as feelings of sadness, loss, or anger that interfere with a person's everyday activities.

1. Introduction

Depression has become the leading disability in the modern world affecting around 4.4% of the population worldwide by 2015, according to reports from World Health Organization [1]. There has been a net increase of 18.4% in the number of people living with depression between 2005 and 2015 [2] and is only going to increase as a result of the COVID-19 pandemic [3, 4, 5, 6]. This has made depression an important topic of research associating it to different behavioral aspects of individuals with major implications on society. Research has demonstrated that during periods of depression, there is a strong reliance on listening to music impacting moods, emotions and altering affect states of an individual [7].

Musical engagement strategies have been linked to measures of ill-health including internalized symptomatology and depression [8, 9]. It has been found that using music for rumination, avoidance and mood worsening is an indicator of risk for depression [10]. To add to this the inability to cease from repeatedly listening to music which leads to such worsening has also been one of the representative engagement strategies of individuals at risk for depression [10]. In terms of musical content, people who are at-risk for depression demonstrate a liking for sad music [11]. The vast majority of these studies

have relied on self-reported musical consumption data and measures, typically using standard questionnaires, which may suffer from demand characteristics and/or social desirability bias. Therefore, in this age of big data, examining naturally occurring online consumption, especially using streaming platforms, provides a more ecologically valid reflection of their true preferences and behaviours [12] and has been proposed as the new frontier [13].

Studies have emerged that have examined mental well-being using multimedia content from social networking platforms such as Facebook, Twitter, Instagram [14, 15, 16, 17, 18]. Till date, Surana et. al [19] has been the only study that examined association between mental well-being and naturally occurring music listening habits. They examined online music consumption via the music streaming platform Last.fm¹ and associated attributes of the most listened to music tracks with the respective users' well-being scores. Subsequently, the social tags, specifically adjectives and adverbs, associated with the most frequently listened tracks per user were projected onto the standard 2-dimensional emotion space, that is Valence-Arousal (VA) model or the Russells Circumplex Model of Affect [19] where Valence represents pleasantness and Arousal represents energy or activity. In addition, a 3-dimensional emotion model (VAD) was also used with the addition of the Dominance or perceived intensity dimension and compared results. They found that individuals at-risk for depression predominantly listen to music which are socially tagged as being representative of low arousal and negatively valenced emotion, sadness, and its related terms, such as *dead*, *low*, *depressed*, *repetitive*, and *miserable*. However, solely focusing on social tags does not yield insights into the acoustic properties of the music. For example, a heavy metal piece of music with tags labelled as *transcendental* or *wonderful* and a sad love song or ballad with the same tags are different acoustically but would get clubbed into the same emotion category. Furthermore, a considerable amount of tracks on online platforms remain untagged resulting in loss of information. Hence it is important to examine acoustic features of music and the perceived emotions extracted from them thereof to get a better understanding of music consumption of individuals at-risk for depression.

Furthermore, emotions are dynamic processes that fluctuate in response to environmental demands and regulatory forces [20] and certain patterns or irregularities with which they fluctuate across time have been associated with psychological well-being. Higher levels of emotion variability and inertia have been linked to depression [21, 22]. These fluctuations in the internal emotion states of individuals might be reflected in their dynamic use of music. Hence, it is important to investigate time-varying musical consumption.

¹www.last.fm

Rumination (self talk) involves repetitive and passive thoughts focused on the causes and effects of a person's distress. However, these thoughts do not lead to the person engaging in active coping mechanisms or problem-solving strategies that would relieve distress and improve mood

Valence: This dimension refers to whether the emotion is positive or negative
Arousal: This represents the intensity of the emotion. It ranges from calm (low arousal) to highly agitated (high arousal).

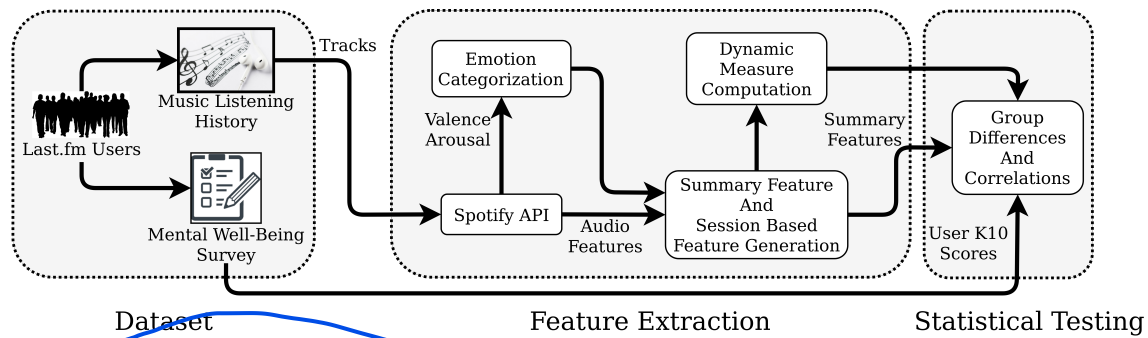


Figure 1: Methodology

Hypo

The aim of this study is to characterize online music consumption of individuals at-risk for depression by analyzing acoustic and emotion-related features and their dynamic variation. To this end, we analyze naturally occurring music listening behavior via listening histories of Last.fm users. We also examine time-varying properties of acoustic and emotion-related features and associate them with the respective users' well-being scores indicative of risk for depression. Based on previous literature we hypothesize the following Last.fm users who demonstrate high depression risk that is characterized by low well-being scores:

- greater overall consumption of music, especially music that is perceived as sad
- higher repetition in music tracks owing to ruminative and repetitive engagement.
- increased level of variability and/or inertia in terms of the acoustic and emotion information of music across time

The results are discussed in the context of potential underlying moods or states.

2. Methodology

Figure 1 encapsulates the procedure used in our study and is described as follows.

2.1. Dataset

Data was from the aforementioned Surana et al. [19] study comprising listening histories of 541 Last.fm users (82 females, mean age = 25.4 years, std = 7.3 years). The respective users' well-being was measured using standard diagnostic questionnaire, the Kesslers Psychological Distress Scale (K10) [23]. Additionally, the Healthy-Unhealthy music (HUMS) [10] scale was administered to measure musical engagement which describes music consumption behavior and indirectly indicates risk for depression, in addition to personality data. HUMS and personality information have not been used in the current study, but have been additionally used in the original study [19] to evaluate internal consistency, which was found to be high. K10 scores demonstrated high reliability (cronbach alpha = 0.91). The participants' listening histories were extracted for a duration of 6 months centered (3 months) around the time they filled the K10 questionnaire.

2.2. Feature Extraction

In order to extract acoustic features from a track we used the Spotify² public API. Spotify³ package was employed to search for each track in Spotify database and retrieve the values for 10 features that Spotify provides. Eight of these were audio features which comprised danceability, loudness, speechiness, acousticness, instrumentalness, liveness, tempo, and mode. All the features take continuous values except mode which is binary (0: major, 1: minor). Additional two emotion features representing the valence and energy/arousal were also retrieved. We further categorized each track into one of the four quadrants of the 2-dimensional VA space [24]. The first quadrant loosely translates to the emotion category representing happiness, the second quadrant to anger, the third to sadness and the fourth to tenderness, as shown in Figure 2[24]. For each quadrant we calculated a quadrant prevalence score (QPS) which is calculated as the proportion of tracks in the respective quadrants in each user's listening history. As mentioned before, individuals at risk for depression have an inclination for music representing sadness [19]. Hence, in order to get a bird's eye-view of the acoustic features associated with individuals at-risk when compared to no risk for depression, we calculated static features that comprise mean score for the eight audio features as well as the QPS over each users entire listening history which we term as summary features hereon.

2.2.1. Session-based feature generation

Since we aimed to capture the changes in participants' audio features and emotions over time, we defined time resolution based on sessions. A session was defined as a period of continuous listening activity, with a subsequent session occurring after an inactivity of at least two hours [25]. Further, in each session, audio features for all the tracks were averaged to denote the value of that feature for the respective session. Also, the QPS for the four quadrants were computed per session for each user. Thus, we obtained time-varying values for both, audio and QPS features, for each user over his/her 6-month listening period. The reason for using a smaller time interval/resolution based on sessions as opposed to larger intervals like days or weeks is due to the fact that emotions are more transient and short-lived than mood states which may span for days [26]. For example, continued music consumption across sessions with certain features may result in greater successive session-based QPS in certain quadrants and might reflect underlying moods. For exam-

²www.spotify.com

³developer.spotify.com

are we focussing of emotion or mood ?

ple, a user might continuously listen to sad music, the acoustic features of which might be predominant in say the third quadrant representing sadness, and hence may be indicative of underlying risk for depression based on aforementioned previous literature. Furthermore, we extracted additional measures that capture such dynamic behavior which provide a better understanding of users' mood shifts characterized by variability and inertia. These additional dynamic measures are explained in following section.

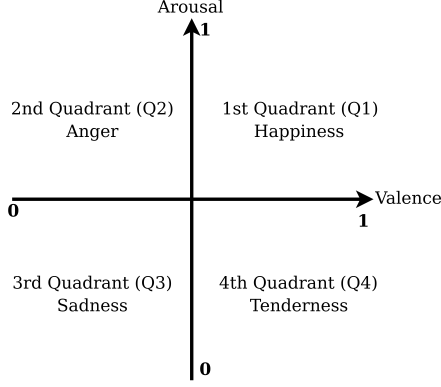


Figure 2: Four quadrants of Valence-Arousal space representing four emotional states.

Finally, since we had hypothesized that individuals at-risk for depression may engage in repetitive music listening behaviors, we defined a *repetitiveness index* (RI) per user, computed as shown in Equation 1. In order to capture a user's repetitiveness within a session, the proportion of tracks that were repeated is first calculated. Subsequently, RI is calculated as the average *repetitiveness* across all sessions.

$$RI_u = \frac{1}{N_u} \sum_{j \in N_u} \frac{\sum_{i \in T_{u,j}} (tr_{i,j} \times I(tr_{i,j} \geq 2))}{\sum_{i \in T_{u,j}} tr_{i,j}} \quad (1)$$

where,

N_u : number of sessions for user u

$T_{u,j}$: all unique tracks in j th session for user u

$tr_{i,j}$: frequency of track i in j th session

$I(condition)$: indicator function, 1 if *condition* evaluates to true, else 0

2.2.2. Dynamic Measures Computation

Previous studies have demonstrated the association of dynamic measures such as variability, instability, and inertia, with depression[27]. **Variability** represents the standard deviation of scores across sessions. **Inertia** is calculated as the autocorrelation coefficient (with time lag of 1 session) for the time-varying scores. It captures the predictability of a feature/emotion from one time-point to the next. **Higher inertia and extreme variability** have been found to be representative of ill-being and depressive moods, especially in negative affect states [21]. Instability is calculated as the mean squared successive difference (MSSD) for scores across sessions. It measures the deviation in scores from one time-point to the next. **A person with high instability implies higher fluctuations in music consumption with respect to the features/emotions.** Since, instability has been found to have mathematical dependency on the other two measures, that

is, it is directly proportional to variability and has inverse relation with inertia [20], we restrict our analyses to computing only variability and inertia. These two dynamic measures were calculated for all the session-based acoustic features and QPS.

2.3. Statistical Testing

We first performed tests of difference by dividing the users into No-Risk and At-Risk groups based on their K10 scores as done in Surana et al. [19]. **Users with K10 < 20 (n=193) fall under the No-Risk group while K10 > 29 (n=142) comprised the At-Risk group.** We performed a two-tailed Mann Whitney U (MWU) test on the static *summary features* between these groups. We further performed bootstrapping to account for Type I error and ensure that the observed differences are not due to chance using the same approach as Surana et al. Next we performed Spearman correlations between K10 and all 541 users' number of sessions, total playcount⁴, and RI. Similarly, correlation between K10 and each of the dynamic measures of audio and emotion features was computed. We additionally performed permutation testing to account for multiple comparisons due to the number of correlations performed and hence account for Type I errors. For each correlation, permutation testing was performed with 10,000 iterations. At each iteration, the K10 scores were randomly assigned (with replacement) to each user. The correlation value for each iteration was calculated. **The significance of the original statistic (Spearman correlation) was estimated from the distribution of the correlation statistics thus obtained.**

3. Results

There were about 8,09,000 unique tracks in the dataset, out of which the audio features were available for 5,73,000 (71%) tracks on Spotify. No significant differences were observed between the No-Risk and At-Risk groups using MWU test based on the static *summary features*. A borderline significant difference ($p = 0.06$) was observed for the acoustic feature *speechiness* having a higher median for the At-Risk group. Figure 3 displays the boxplots of the QPS *summary feature* for both groups. A borderline group difference ($p = 0.1$) for the QPS in the fourth quadrant (Q4) representing *Tenderness*. However, we do observe that the median in the third quadrant (Q3) representing *Sadness* was higher for the At-Risk group.

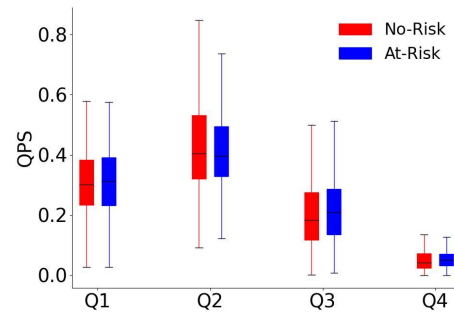


Figure 3: Boxplot of QPS summary feature (for each of the four quadrants) for No-Risk and At-Risk groups.

On an average, each user had 292 sessions. Significant

⁴Here, the total playcount refers to the total number of tracks(including the repetitions) listened by the user.

positive correlations were observed between K10 and number of sessions ($r = 0.16, p = 0.0003$), total playcount ($r = 0.11, p = 0.01$) and *RI* ($r = 0.16, p = 0.0002$). These correlations were found to be significant post permutation testing (all $p < .01$).

Audio Feature	Variability	Inertia
Danceability	-0.03	0.03
Loudness	-0.01	0.05
Speechiness	0.05	0.09*
Acousticness	0.02	0.05
Instrumentalness	-0.11**	0.02
Liveness	-0.05	0.05
Tempo	-0.05	0.03
Mode	-0.05	0.07

Table 1: Spearman correlations between users' dynamic measures represented by Variability and Inertia of audio features and K10 scores. (* $p < 0.05$, ** $p < 0.01$)

Table 1 displays the correlation results between the dynamic measures of the *audio features* and K10 scores. K10 had a negative correlation with variability in *instrumentalness* ($r = -0.11, p = 0.009$) which retained significance even after controlling for mean and inertia. Also, K10 correlated positively with inertia in *speechiness* ($r = 0.09, p = 0.03$) retaining significance after controlling for mean and variability. These correlations were also found to be significant after permutation testing.

Quadrant	Variability	Inertia
Q1	-0.06	0.03
Q2	-0.06	-0.03
Q3	0.02	0.1*
Q4	0.03	0.05

Table 2: Spearman correlations between users' dynamic measures, represented by Variability and Inertia of QPS and K10 scores. (* $p < 0.05$)

Table 2 presents the correlation results between the dynamic measures of QPS and K10. As expected, significant positive correlation was observed between inertia and K10 in the third quadrant (Q3) representing *sadness* ($p = 0.02$) retaining significance post permutation testing. In addition, this correlation was preserved after controlling for mean QPS ($r = 0.09, p = 0.03$).

4. Discussion

This study is the first to analyze the association between depression risk and music consumption via acoustic features. We analyse naturally occurring listening behavior of individuals on the online music streaming platform Last.fm as opposed to self-reported or lab-based studies. Further, this study is the first of its kind to look at the dynamic nature of music consumption to characterize risk for depression.

No significant differences were found between At-Risk and No-Risk groups in terms of audio features when they were averaged over the entire listening history of participants. This suggests that the kinds of music listened by individuals with varying depression risk has comparable features. However, a further inquiry into genre preferences might shed light on perceptual attributes that might be more relevant and representative

of At-Risk individuals' consumption. The borderline significant difference demonstrated in the fourth quadrant QPS and sub-threshold observable difference between the median QPS and respective ranges in Q3 (Figure 3) are very much in line with previous studies [11, 19]. Individuals with high K10 scores are characterized by states that are high in anxiety and arousal. When using music as a coping mechanism, a natural choice may be to listen to music that is low on arousal owing to their already heightened physiologically aroused states. Since, we take a subset of the participants to investigate group differences, future studies can aim to increase sample size to unearth more definitive trends.

Session-based analyses yielded several significant results which are in line with previous studies. The positive correlation observed between K10 and the total number of sessions and playcount suggests greater reliance on music for individuals with higher risk for depression. Higher musical engagement is representative of emotion focused-coping [28], especially during periods of depression [7]. Furthermore, the positive association between *Repetitiveness Index* and K10 reflects the ruminative coping style which is characteristic of individuals at-risk for depression [10] which may result in intensifying their negative moods [11, 29]. Whether this ruminatory behaviour with music as a virtual other is beneficial or maladaptive in the long-run remains to be seen.

The negative association found between K10 and variability in *instrumentalness*, an acoustic feature that indicates the instrumental nature or lower presence of spoken words in songs, suggests that the greater the risk of depression, the less the deviation in *instrumentalness*. This result coupled with higher median *instrumentalness* for At-Risk group than No-Risk is in line with the findings of Surana et al. [19] who report that At-Risk individuals gravitated towards neo-psychedelic-, avant garde-, dream-pop genres that are characterized by obscure vocals. The positive association between Inertia in *speechiness*, an acoustic feature that captures presence of spoken words in a track, and depression risk suggests that high risk individuals tend to be resistant to change when listening to certain music associated with specific levels of *speechiness* over time. One limitation is the of assignment an average estimate of an acoustic feature to a track, since it does not capture within-track variability. Hence, two tracks which sound perceptually different may have similar values for an acoustic feature as a result of averaging over the track. However, combining acoustic features with tag information in future may provide relevant perceptual information and hence better insight into describing aspects of music associated with individuals at risk.

Finally, the positive association between inertia in the Q3, representing *Sadness* in the VA space, and K10 supports our hypothesis that individuals At-Risk not only listen to music that is more reflective of their negatively valenced states, but also that this affinity is more resistant to change over time. Since it is known that individuals at risk for depression exhibit higher inertia in their own internal states, we can conclude that this behavior is reflected in their music consumption patterns. This is a novel finding that further corroborates the notion that music is indeed a mirror of the self, especially in terms of dynamic states.

In future, we intend to employ a multimodal approach of incorporating tags, acoustic features, and lyrics-based information such as sentiments, to gain a holistic understanding of music consumption characterizing depression risk. This would then pave way to designing early depression detection systems using digital music footprints.

5. References

- [1] W. H. Organization *et al.*, “Depression and other common mental disorders: global health estimates,” World Health Organization, Tech. Rep., 2017.
- [2] T. Vos, C. Allen, M. Arora, R. M. Barber, Z. A. Bhutta, A. Brown, A. Carter, D. C. Casey, F. J. Charlson, A. Z. Chen *et al.*, “Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the global burden of disease study 2015,” *The lancet*, vol. 388, no. 10053, pp. 1545–1602, 2016.
- [3] L. Sher, “The impact of the covid-19 pandemic on suicide rates,” *QJM: An International Journal of Medicine*, 2020.
- [4] U. Rehman, M. G. Shahnawaz, N. H. Khan, K. D. Kharshiing, M. Khurshed, K. Gupta, D. Kashyap, and R. Uniyal, “Depression, anxiety and stress among indians in times of covid-19 lockdown,” *Community mental health journal*, pp. 1–7, 2020.
- [5] S. Özdin and Ş. Bayrak Özdin, “Levels and predictors of anxiety, depression and health anxiety during covid-19 pandemic in turkish society: The importance of gender,” *International Journal of Social Psychiatry*, p. 0020764020927051, 2020.
- [6] C. Mazza, E. Ricci, S. Biondi, M. Colasanti, S. Ferracuti, C. Napoli, and P. Roma, “A nationwide survey of psychological distress among italian people during the covid-19 pandemic: Immediate psychological responses and associated factors,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 9, p. 3165, 2020.
- [7] J. Stewart, S. Garrido, C. Hense, and K. McFerran, “Music use for mood regulation: self-awareness and conscious listening choices in young people with tendencies to depression,” *Frontiers in psychology*, vol. 10, p. 1199, 2019.
- [8] S. Garrido, T. Eerola, and K. McFerran, “Group rumination: Social interactions around music in people with depression,” *Frontiers in psychology*, vol. 8, p. 490, 2017.
- [9] K. S. McFerran, S. Garrido, and S. Saarikallio, “A critical interpretive synthesis of the literature linking music and adolescent mental health,” *Youth & Society*, vol. 48, no. 4, pp. 521–538, 2016.
- [10] S. Saarikallio, C. Gold, and K. McFerran, “Development and validation of the healthy-unhealthy music scale,” *Child and adolescent mental health*, vol. 20, no. 4, pp. 210–217, 2015.
- [11] S. Garrido and E. Schubert, “Music and people with tendencies to depression,” *Music Perception: An Interdisciplinary Journal*, vol. 32, no. 4, pp. 313–321, 2015.
- [12] G. Nave, J. Minxha, D. M. Greenberg, M. Kosinski, D. Stillwell, and J. Rentfrow, “Musical preferences predict personality: evidence from active listening and facebook likes,” *Psychological Science*, vol. 29, no. 7, pp. 1145–1158, 2018.
- [13] D. M. Greenberg and P. J. Rentfrow, “Music and big data: a new frontier,” *Current opinion in behavioral sciences*, vol. 18, pp. 50–56, 2017.
- [14] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, “Predicting depression via social media,” in *Seventh international AAAI conference on weblogs and social media*, 2013.
- [15] G. Coppersmith, M. Dredze, and C. Harman, “Quantifying mental health signals in twitter,” in *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, 2014, pp. 51–60.
- [16] M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, “Characterizing and predicting postpartum depression from shared facebook data,” in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 2014, pp. 626–638.
- [17] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar, “Discovering shifts to suicidal ideation from mental health content in social media,” in *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2016, pp. 2098–2110.
- [18] A. G. Reece and C. M. Danforth, “Instagram photos reveal predictive markers of depression,” *EPJ Data Science*, vol. 6, no. 1, pp. 1–12, 2017.
- [19] A. Surana, Y. Goyal, M. Shrivastava, S. Saarikallio, and V. Al-luri, “Tag2risk: Harnessing social music tags for characterizing depression risk,” in *To appear in proceedings of the 21st International Society for Music Information Retrieval Conference*, 2020.
- [20] P. Koval, M. L. Pe, K. Meers, and P. Kuppens, “Affect dynamics in relation to depressive symptoms: variable, unstable or inert?” *Emotion*, vol. 13, no. 6, p. 1132, 2013.
- [21] P. Kuppens and P. Verduyn, “Emotion dynamics,” *Current Opinion in Psychology*, vol. 17, pp. 22–26, 2017.
- [22] M. Houben, W. Van Den Noortgate, and P. Kuppens, “The relation between short-term emotion dynamics and psychological well-being: A meta-analysis,” *Psychological bulletin*, vol. 141, no. 4, p. 901, 2015.
- [23] R. C. Kessler, G. Andrews, L. J. Colpe, E. Hiripi, D. K. Mroczek, S.-L. Normand, E. E. Walters, and A. M. Zaslavsky, “Short screening scales to monitor population prevalences and trends in non-specific psychological distress,” *Psychological medicine*, vol. 32, no. 6, pp. 959–976, 2002.
- [24] J. A. Russell, “A circumplex model of affect,” *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.
- [25] K. Gupta, N. Sachdeva, and V. Pudi, “Explicit modelling of the implicit short term user preferences for music recommendation,” in *European Conference on Information Retrieval*. Springer, 2018, pp. 333–344.
- [26] J. Rottenberg, “Mood and emotion in major depression,” *Current Directions in Psychological Science*, vol. 14, no. 3, pp. 167–170, 2005.
- [27] E. M. Seabrook, M. L. Kern, B. D. Fulcher, and N. S. Rickard, “Predicting depression from language-based emotion dynamics: longitudinal analysis of facebook and twitter status updates,” *Journal of medical Internet research*, vol. 20, no. 5, p. e168, 2018.
- [28] D. Miranda, P. Gaudreau, R. Debrosse, J. Morizot, and L. J. Kirmayer, “Music listening and mental health: Variations on internalizing psychopathology,” *Music, health, and wellbeing*, pp. 513–529, 2012.
- [29] S. Garrido and E. Schubert, “Moody melodies: Do they cheer us up? a study of the effect of sad music on mood,” *Psychology of Music*, vol. 43, no. 2, pp. 244–261, 2015.