

# Affective Idiosyncratic Responses to Music

Sky CH-Wang<sup>◦</sup> Evan Li<sup>◦</sup> Oliver Li<sup>◦</sup> Smaranda Muresan<sup>◦•</sup> Zhou Yu<sup>◦</sup>

<sup>◦</sup>Department of Computer Science, Columbia University

<sup>•</sup>Data Science Institute, Columbia University

skywang@cs.columbia.edu

{el3078, al4143, smara, zy2461}@columbia.edu

## Abstract

Affective responses to music are highly personal. Despite consensus that idiosyncratic factors play a key role in regulating how listeners emotionally respond to music, precisely measuring the marginal effects of these variables has proved challenging. To address this gap, we develop computational methods to measure affective responses to music from over 403M listener comments on a Chinese social music platform. Building on studies from music psychology in systematic and quasi-causal analyses, we test for musical, lyrical, contextual, demographic, and mental health effects that drive listener affective responses. Finally, motivated by the social phenomenon known as 网抑云 (wǎng-yì-yún), we identify influencing factors of platform user self-disclosures, the social support they receive, and notable differences in discloser user activity.

## 1 Introduction

Music can evoke powerful emotions in listeners (Meyer, 1956). However, our emotional reactions to it are not universal— affective responses to music are highly *personal*. Just as you may wonder why your friend is sobbing to a song that you only feel ambivalent about, a listener’s emotional response to music not only varies with inherent audio or lyrical features (Hevner, 1935; Webster and Weir, 2005; Van der Zwaag et al., 2011), but also with other factors such as a listener’s demographics, mental health conditions, and surrounding environment (Krugman, 1943; Robazza et al., 1994; Gregory and Varney, 1996; Juslin and Västfjäll, 2008; Saarikallio et al., 2013; Garrido et al., 2018). As a result of this idiosyncrasy, it has been extremely difficult to precisely measure the marginal effects of these variables on a listener’s affective response (Yang et al., 2007; Beveridge and Knox, 2018). This difficulty is further compounded especially when examining how a collection of these

factors influence individual affective reactions in combination (Gómez-Cañón et al., 2021).

Music psychology has long focused on identifying the relationships between human affect and music, both in those that are perceived and those that are felt. Perceiving and feeling emotions in music, while highly related, are not identical (Hunter and Schellenberg, 2010). Examining the latter has proved challenging, as in addition to insufficient scale for finding significance, measuring felt emotions in participatory studies often interferes with the experience itself (Gabrielsson and Lindström, 2010). While recent computational studies have attempted citizen science approaches for annotation (Gutiérrez Páez et al., 2021), reliability remains an issue; annotator confusion persists between the concepts of perceived and induced emotions (Juslin, 2019). Our work expands this line of research by examining affective responses to music in a natural setting: an online social music platform.

We test for differences in affective responses to music by computationally measuring expressed emotions from a massive study of over 403M listener comments on one of China’s largest social music platforms, Netease Cloud Music. Our paper offers the following three contributions. First, we reveal several nuances in listener affective responses against a host of musical, lyrical, and contextual factors, showing evidence of emotional contagion. Second, in a multi-modal quasi-causal analysis, we show that listeners of different genders and ages vary in their reactions to musical stimuli and identify specific features driving demographic effects on affective responses. Third, motivated by the social phenomenon known as 网抑云,<sup>1</sup> we systematically study self-disclosures of mental health disorders on the platform, identifying driving fac-

<sup>1</sup>网抑云, pronounced as wǎng-yì-yún, is a pun on the platform name 网易云 that refers to the outpour of emotional and personal comments on the social music platform especially late at night and under sad songs. 抑, here, is the first character of the word for depression—抑郁.

tors of this behavior, the social support they receive, and differences in discloser user activity.

## 2 Data

Our work is drawn from one of the largest music streaming services in China, Netease Cloud Music, and focuses on Chinese-language user content.

**Netease Cloud Music.** 网易云音乐 (wǎng-yì-yún-yīn-yuè) has over 185 million monthly active users (Dredge, 2022). Unlike mainstream music streaming services in the United States such as Spotify and Apple Music, Netease Cloud Music is a *social music platform* (Zhou et al., 2018; Wang and Fu, 2020). Here, among other unique features, each song, album, and playlist have comment sections that serve as discussion boards, where users can post top-level comments as well as reply to or up-vote existing ones. These platform interactions serve as a natural setting on listener responses, where users are able to post freely<sup>2</sup> in the comment sections of what they are currently listening to. Users are required to create an account to access most of the platform’s features; when doing so, users optionally input personal demographic information like age, gender, and location, which they can then choose to display as public or private.

**Dataset Collection.** To collect a representative sample of public platform commenting activity, we adapt traditional snowball sampling (Atkinson and Flint, 2001) across multiple random seeds to build an exhaustive list of user, song, album, and playlist entity ids on the platform. We then uniformly sample from the set of entities that have at least one public comment posted. Data was collected from all public content ranging from the platform’s inception, 2013, to 2022, totaling over 455K albums, 2.87M songs, 1.36M playlists, 29.9M users, and 403M comments. A detailed breakdown of our data and a view of the interaction interface of the platform are shown in Appendix Section A. The study and data collection were approved by an Institutional Review Board as exempt.

## 3 Measuring Affective Response

We measure affective responses to music as expressed in comments posted under their comment sections. Since not all comments are indicative of a user’s emotional response, we sample a subset of user content and examine both the experiencer

<sup>2</sup>See our [Limitations](#) section for a discussion on platform moderation and censorship.

---

我只想和你一个人做那些浪漫到极致的事

*Translation:* I just want to do the most romantic things with you alone

果然不该来的。混蛋老爸，气死我了！

*Translation:* I shouldn’t have come.

Asshole dad, pisses me off!

太棒了好听太治愈了我莫名有点想哭

*Translation:* It’s so good, it’s so healing,

I feel like crying for some reason

---

Table 1: Example top-level comments indicating an affective response.

of the emotion and its expressed stimulus, before conducting our analysis.

**Emotion Experiencer.** Two annotators first manually annotated 1000 comments selected uniformly at random to identify the experiencer of the emotion expressed in top-level comments. Top-level comments were chosen to limit dyadic interaction effects and are used to measure affective responses later on. With an initial Cohen’s  $\kappa$  of 0.80 and with disagreements resolved via discussion, similar to (Mohammad et al., 2014), we find that the experiencer of the emotion expressed in the comment is often the commenter themselves (99.1%); we thus maintain this assumption in our later experiments. Selected examples and annotation guidelines are shown in Table 1 and Appendix Section B, respectively.

**Affective Stimulus.** Next, annotators were tasked with identifying what caused the emotional response in the comment itself. Annotators labeled for comments containing emotions that could explicitly be said to *not* originate from music—under the BRECVEMA framework of music-induced emotions (Juslin, 2013), emotions are evoked in listeners via a combination of mechanisms related to aesthetic appreciation, entrainment, visual imagination, and emotional associations with past experiences, among other factors. A listener’s emotional state also has an effect on their music choice; for example, listeners often use music for mood regulation, or as a coping mechanism (Stewart et al., 2019; Schäfer et al., 2020). Here, we make no explicit causal assumptions of music choice and seek only to measure comment affective responses. With an initial Cohen’s  $\kappa$  of 0.76 and with disagreements resolved via discussion, we only find a few instances (3.3%) where affective stimuli may be

explicitly attributed elsewhere. There are a few patterns among these irrelevant comments: namely, that they primarily relate to album images, quotations, and easily identifiable spam messages, i.e. “沙发” (meaning “first comment”). Aiming for high precision, we create simple regular expressions and redundancy filtering to increase the relevance of comments with affective content, achieving a precision of 98.8% on a held-out test set of the same size. Specific annotation guidelines and filtering methods are shown in Appendix Sections B and D, respectively.

**Measuring Affective Response.** We characterize emotions across a 2-dimensional plane of valence and arousal following the Russell model of emotions (Russell, 1980), representing the degree of positivity and emotion intensity, respectively. Specifically, we employ a lexicon-based approach to measure valence and arousal in music comments, using one of the largest crowd-sourced datasets for the Chinese language—Chinese EmoBank (Yu et al., 2016)—containing 5512 words annotated for their valence and arousal. In the following sections, these measures of expressed emotions in comments are what we define as listener affective response.

## 4 Variations in Affective Response

Computing comment-level valence and arousal by averaging across word-level scores,<sup>3</sup> we analyze variations in listener affective responses to (1) musical and (2) lyrical features, (3) contextual factors, and (4) user demographic variables.

### 4.1 Musical and Lyrical Features

In prior work, while much emphasis is placed on identifying the causes of perceived emotions in music, less emphasis has been placed on emotional responses, which are highly influenced by extramusical and contextual factors in listeners (Gómez-Cañón et al., 2021). Recent work has attempted to use physiological signals and self-reported emotions to measure emotional responses in listeners (Hu et al., 2018), though this has proved challenging partly due to a high degree of intercorrelations and confounds, causing the number of trials needed to measure such effects to be intractable relative

<sup>3</sup>To confirm that our findings were robust against high-volume sentiment terms, following a distribution analysis of comment valence and arousal scores, we drop the top three most frequent terms—好听, 好, and 喜欢, which roughly translate to “sounds good”, “good”, and “like”, respectively—and recompute our experiments, obtaining similar results.

to typical experiment scale (Eerola et al., 2013). Using our data, we test for the marginal effects of musical and lyrical features on affective responses.

**Methods.** To understand the marginal contributions of these variables on affective responses, we fit separate multivariable linear regression models on response valence and arousal, including the features described below as regressors. As affective responses are highly idiosyncratic (Juslin and Västfjäll, 2008), we further control for listener demographics, namely age, gender, and location. We then test for multicollinearity by computing the variance inflation factor (VIF) for each variable and iteratively remove collinear variables in our regression that have a VIF greater than 5. In our analyses, we stratify continuous variables (i.e. tempo) into fixed-length category indicator variables (i.e., 80-90 BPM, 90-100 BPM, and so on) and measure the average marginal effects (AME) on valence and arousal of each stratum, using the first of such categories as the reference group (i.e., the AME of 90-100 BPM, and so on, relative to 80-90 BPM).

**Musical Features.** We use *librosa* (McFee et al., 2015), *pydub* (Robert et al., 2018), and *timbral\_models* of the AudioCommons project (Font et al., 2016) to derive song file musical features. We extract (1) **tempo** and (2) **tempo standard deviation**, both in beats per minute (BPM) (Ellis, 2007); (3) **loudness**, measured as the average decibels relative to full scale (dBFS) value of the entire song; (4) **mode**, namely, major or minor, and (5) **key**, i.e. C# minor (Krumhansl, 2001); as well as eight additional timbral features, or the perceived sound qualities of a piece of music. They are, specifically, (6) **depth**, related to the emphasis of low frequency content, (7) **brightness**, a measure that correlates both with the spectral centroid and the ratio of high frequencies to the sum of all energy of a sound, and (8) **warmth**, often created by low and midrange frequencies and associated with lower harmonics (Pearce et al., 2017); (9) **roughness**, a sound’s buzzing, harsh, and raspy quality (Vassilakis and Fitz, 2007); (10) **sharpness**, measuring high frequency content (Zwicker and Fastl, 2013); (11) **hardness**, the amount of aggression (Pearce et al., 2019); (12) **reverberation**, a sound or echo’s persistence after it is initially produced (Jan and Wang, 2012); and (13) **boominess**, a sound’s deep resonant quality as measured by the booming index (Hatano and Hashimoto, 2000). Here, reverberation is classed as a binary variable, while all other

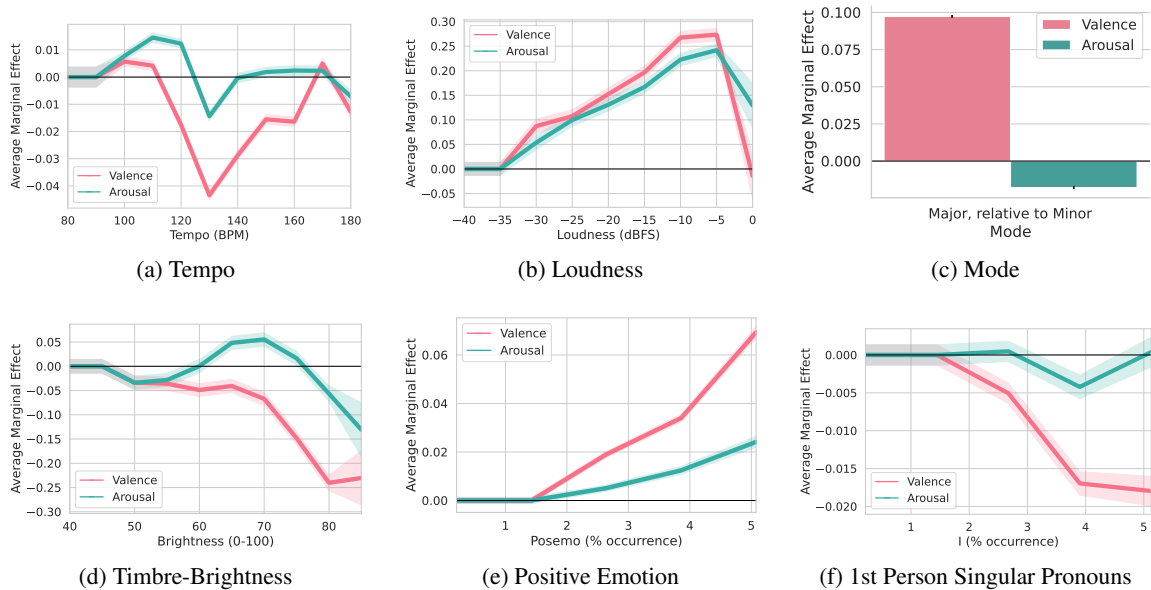


Figure 1: Average marginal effects of musical and lyrical features on listener affective responses with other features and listener demographics as controls. Standard errors are shown; valence in red, arousal in blue. The complete set of figures for musical and lyrical features are shown in Appendix Section E, Figures 13-18.

timbral features are measured and clipped to values between 0 and 100 following their regression models’ training data domain considerations.

**Psycholinguistic Lyrical Features.** Similar to Mihalcea and Strapparava (2012) while limiting our analysis to Chinese language songs, we extract coarse psycholinguistic lexical features of lyrics. Specifically, we preprocess lyrics with regular expressions to remove extraneous information as shown in Appendix Section C and use the Simplified Chinese version (Huang et al., 2012) of the Linguistic Inquiry and Word Count (LIWC) lexicon (Pennebaker et al., 2015) to create normalized counts of tokenized word semantic classes.

**Topic-wise Lyrical Features.** To capture thematic trends across song lyrics, we train a 20-topic LDA model on preprocessed song lyrics and manually label each topic with its prominent theme, i.e. Nationalism/China or Hometown/Childhood. Labeled lyrical topics and the top words associated with each are shown in Appendix Section C.

**Results.** Figure 1 reveals five important trends in affective responses across a variety of musical and lyrical features.

First, tempo exhibits a bimodal distribution relative to both valence and arousal; listeners are most intensely positive for tempos of around 110 BPM and 160 BPM, with the former eliciting greater arousal. Higher tempo variation also sees similar

increases in affective responses, although tempo standard deviations of around 35-40 BPM produce the opposite effect, with arousal peaking earlier than valence. Our findings are consistent with prior work on listener self-ratings and measured physiological responses that have used coarse categorizations of tempo, i.e. “fast” tempo (Liu et al., 2018), or the presence and absence of tempo variation (Kamenetsky et al., 1997), as opposed to the continuous measures we use here.

Second, consistent with prior work (Schubert, 2004; Gomez and Danuser, 2007), loudness generally produces a strong positive correlation with more intensely positive reactions; changes in loudness also see a greater change in AME than that of tempo. However, this trend is reversed for songs that are loudest (i.e. between -5 and 0 dBFS)—while unexplored in prior work within music psychology, this observation intuitively follows neural downregulation responses to excessively loud or unpleasant sounds (Hirano et al., 2006; Koelsch, 2014).

Third, consistent across all keys (Appendix Figure 14), major mode in songs has a greater valence and a lower arousal than minor mode. This observation extends prior work investigating the interaction effects between mode and affective responses (Van der Zwaag et al., 2011) in a western tonal context, suggesting that associations of sadness



and happiness by way of musical mode are also consistent in Chinese listeners.

Fourth, increases in most timbral characteristics see similar increases in the intensity and positivity of reactions up until a point of extremity, whereafter the opposite effect is observed. The only exceptions are roughness and warmth, in which both valence and arousal see monotonically decreasing and increasing trends, respectively. Our results for brightness specifically provide nuance into how, when exploring an expanded set of timbral characteristics and moving beyond only varying timbre through different instruments (Hailstone et al., 2009; Eerola et al., 2012; Wallmark et al., 2018), excess of a timbral feature can produce the opposite initial effect on affective responses.

Fifth, listener affective reactions mirror the psychological states expressed in lyrics. Changes in response valence and arousal closely match the proportion of LIWC category terms for affective processes. Greater use of positive emotion terms sees greater response positivity ( $r=0.93, p<0.05$ ), while the opposite is true for negative emotion terms ( $r=-0.92, p<0.05$ ), and both saw rises in response intensity with their increased use. Furthermore, increases in first-person pronouns also see decreases in valence ( $r=-0.94, p<0.05$ ), mirroring work on the depressed psychological states reflected through their increased use (Pennebaker, 2011). Interpreted together with our findings on musical features, these observations mirror emotional contagion (Juslin, 2013), where the recognition of emotions expressed in music evokes similar emotions in listeners.

These findings, compared to prior work, highlight the importance of using finer-grained measurements on an extended set of features and controls to provide a more nuanced analysis of emotional responses to musical and lyrical stimuli. Expanded results with the full list of figures are shown in Appendix Figures 13-18.

## 4.2 Contextual Factors

Extramusical factors such as listening context (e.g., listening to music when grabbing coffee vs. when exercising) also influence the emotional effects of music (Sloboda and O’neill, 2001; Greasley and Lamont, 2011; Vuoskoski and Eerola, 2015). Prior work has primarily utilized experience sampling methods (Csikszentmihalyi and LeFevre, 1989) to study musical experiences in everyday contexts—

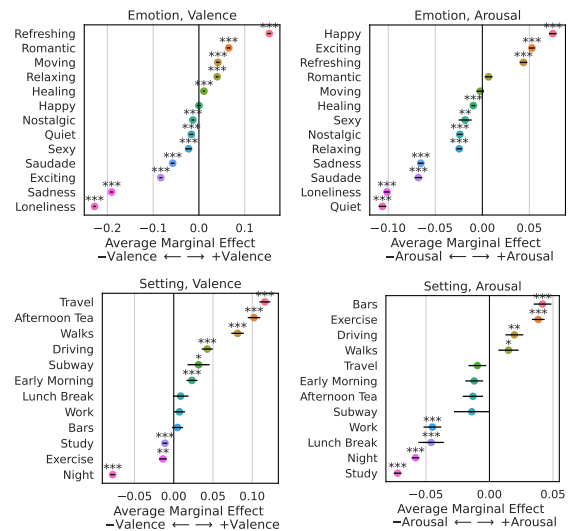


Figure 2: Averaged marginal effects of contextual choices in emotion-tagged (top) and setting-tagged (bottom) playlists on listener affective responses; standard errors are shown.

where participants are polled at random intervals during the day—though generalizations to the population at large have proved difficult with small sample sizes (Sloboda et al., 2001; Juslin et al., 2008). While we are unable to obtain information about the physical setting a user was in (i.e. that a user was exercising when listening to a song), here, using our data on playlists and treating the choices of users in listening to playlists of specific types as context, we tease out the marginal effects that these choices have on affective responses.

**Choice as Context.** We obtain context variables on 1.36M playlists through their tags, used by creators to label individual playlists. Tags consist of a set of physical setting (e.g., afternoon tea), emotional (e.g., nostalgic), and thematic (e.g., video game music) categories, in addition to language (e.g., Chinese) and stylistic (e.g., jazz) labels. As users primarily discover new playlists within the platform by browsing specific tags, we treat these tags as implicit signals of *choice* with these listening contexts, aiming to identify those that may differ on the emotional responses produced—i.e. that a user chose to listen to an exercise tagged playlist rather than an afternoon tea tagged one—noting that we do not make explicit causal assumptions behind the factors that led to these user choices.

**Methods.** To identify the marginal effects of contextual choices on affective responses, we fit separate mixed-effect multivariable linear regression models on response valence and arousal, includ-

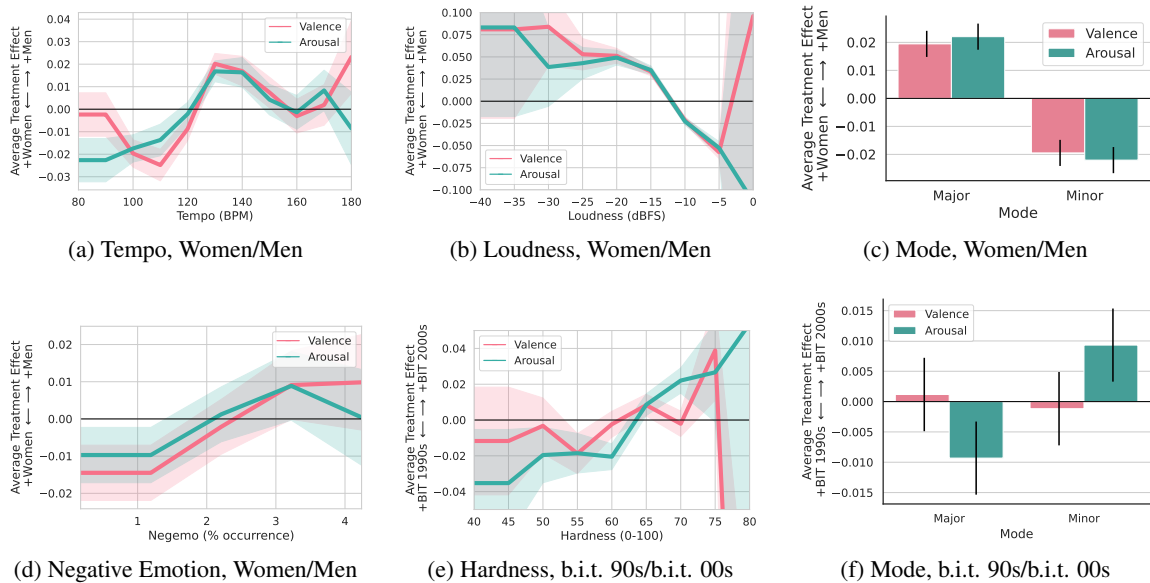


Figure 3: Relative average treatment effects of gender (women/men) and age-based (born in the (b.i.t.) 1990s/b.i.t. 2000s) demographic groups on listener response valence and arousal against musical and lyrical features. Standard errors are shown; valence in red, arousal in blue. The complete set of figures for musical and lyrical features are shown in Appendix Section E, Figures 24-26.

ing tagged category indicator variables as features and control for listener demographics. To further control for differences between playlist songs, we include them as random effects; for computational tractability, we include only random effects for songs that are labeled with 10 or more unique tags. **Results.** Cultivating affect is a driving reason behind why users create playlists (DeNora, 2000; Siles et al., 2019), and our results point to how playlists created by users are also generally successful at cultivating these affects among the general user population as well. As shown in Figure 2, playlists tagged by leisurely activity categories corresponded to the highest positivity in responses, and are consistent with prior work on stress levels in everyday situations (Västfjäll et al., 2012), while arousal trends mirror diurnal shifts in emotion and physical activity (Golder and Macy, 2011). Expanded results for all tagged categories are shown in Appendix Section E, Figures 19-23.

### 4.3 Demographic Variations

Individuality is a driving factor in how listeners experience musically-evoked emotions (Yang et al., 2007; Juslin and Västfjäll, 2008; Gómez-Cañón et al., 2021). However, measuring *how* individual differences affect emotional responses to music has proved challenging, with many researchers citing the insufficiency of typical experiment scale as a

primary reason (Juslin et al., 2008; Lundqvist et al., 2009; Cameron et al., 2013), especially in the presence of confounders. For listener demographics, prior work has seen conflicting observations of how demographic effects modulate affective responses against musical features. For example, some observe that age and gender modulate emotional responses against tempo, mode, volume, and pitch (Webster and Weir, 2005; Chen et al., 2020), while others report the absence of such demographic effects or even contrasting observations (Robazza et al., 1994; Cameron et al., 2013). These contrary results might be due to variable experimental setups between studies, wherein the method of measurement will often interfere with the experience itself (Gabrielsson and Lindström, 2010). This raises the importance of studying emotional reactions in a natural setting when analyzing affective responses to music in everyday situations. Here, we test for demographic differences in affective responses in relation to song features using our data.

**Demographic Variables.** Our analysis focuses on two main demographic variables, namely listener gender<sup>4</sup> and age.<sup>5</sup> We operate within the constraints of platform-provided choices in user

<sup>4</sup>Platform-provided options for gender are limited to binary options (男 man and 女 woman); see [Ethical Considerations](#).

<sup>5</sup>In the form of "XX后", e.g. 00后 refers to individuals that are born between 2000 and 2010.

registration for our variable categories and use only publicly displayed user data in our analysis.

**Methods.** To test for differences between pairs of demographic groups in their affective responses to musical and lyrical features, we formalize alternations between groups as *treatments* and compute average treatment effects (ATE). In order to account for covariates and reduce bias due to confounding variables, we construct a multi-modal stratified propensity score matching (PSM) model as a quasi-causal analysis of demographic effects. Here, we formalize comments as *subjects*; the propensity score, defined traditionally as the likelihood of being assigned to a treatment group based on observed characteristics of the subject (Rosenbaum and Rubin, 1983), is thus a scaled estimate of the likelihood of a commenter being of a demographic group  $g_i$  given a set of song features  $f_i$ , or  $P(g_i|f_i)$ . We estimate this probability—the propensity score—via logistic regression on a song’s musical and lyrical features, and match data points within stratified deciles of this score to mitigate confounding bias (Rosenbaum and Rubin, 1984; Paul, 2017). Within these matched and stratified deciles, we fit separate linear regression models on response valence and arousal against specific song features, weighting and pooling stratum-specific estimated treatment effects to estimate the ATE (Imbens, 2004) and its variance (Lunceford and Davidian, 2004). Consistent with prior work in musical emotions (Kamenetsky et al., 1997) and in social psychology on how cultural constructions of gender may account for differences in emotional display (Bem, 1974), we observe that response valence and arousal by demographic groups differ in their distributions—for example, as shown in Appendix Section A.7, comments made by women are on average higher in both valence and arousal than those made by men. Therefore, we test specifically for standardized *change* in affective responses across song features within demographic groups. Finally, as in Section 4.1, we stratify continuous variables in our analyses into fixed-length categories and estimate the ATE of each stratum.

**Results.** Shown in Figure 3, we find that listener age and gender both modulated affective responses to statistically significant degrees across a series of musical and lyrical features. Compared to men, women had more intensely positive affective reactions for songs that were louder ( $>-12$  dBFS), of lower tempo ( $<120$  BPM), of minimal tempo stan-

dard deviation ( $<5$ ), of minor mode, and that had reverb; though gender differences often diminished (i.e. tempo  $>160$ BPM) or became statistically insignificant at feature extremities. Lyrically, women were affected more negatively with a greater proportion of negemo terms, while men were affected more positively for posemo terms, consistent with observed gendered responses in other mediums (Bradley et al., 2001; Fernández et al., 2012). Compared to women, men had more intensely positive affective responses to darker, flatter, softer, smoother, and warmer timbral characteristics. Age effects were much less pronounced, with those born in the 2000s reacting more positively to harder and rougher timbral features as well as minor modes than those born in the 1990s. Full results are shown in Appendix Section E, Figures 24 to 26.

Affective experiences are central reasons for music consumption (DeNora, 1999), though music choices often misalign with intended well-being outcomes (Stewart et al., 2019). We hope our work further facilitates more effective and more intentional music choices in daily consumption to achieve these desired results. Under appraisal theory, affective responses are learned and conditioned through individual lived experiences rather than innate to certain biological factors (Brody and Hall, 2010). These findings on demographic effects should then be interpreted to be *products* of the social norms, values, and lived experiences (de Boise, 2016) of those who may self-identify under the broad demographic groups in question; with the platform, song, and comment board being part of the context in which these emotions are deployed.

## 5 Disclosures of Mental Health Disorders

In the context of social media, given the frequent benefits of anonymity (De Choudhury and De, 2014) and social connectedness (Bazarova and Choi, 2014), self-disclosures of personal details can be a method to find social support, advice, and belonging (Ernala et al., 2018; Yang et al., 2019). This phenomenon gives life to “网抑云,” which refers to the outpour of emotional and personal comments on the social music platform, especially late at night and under sad songs. While known colloquially and in popular culture,<sup>6</sup> the mechanisms behind self-disclosure phenomena in the context of social music platforms are not well understood.

<sup>6</sup>“网抑云” was termed one of 2020’s top ten internet buzzwords, <https://baike.baidu.com/item/2020十大网络热词/>

Motivated to better understand disclosures of mental health disorders in a musically-situated social environment, we frame them as affective responses (Ho et al., 2018) and test for factors driving this behavior, the social support they receive, and differences in discloser user activity. Addressing these unknowns will help us understand how users may use social music platforms for therapeutic purposes (Schriewer and Bulaj, 2016) and guide us to better support vulnerable and at-risk individuals.

**Dataset Collection.** In the absence of clinically-aligned user data (Harrigian et al., 2020), we source disorder terms from the DSM-5-TR<sup>7</sup> (American Psychiatric Association, 2022) and utilize regular expressions to identify disclosures of self-reported statements of diagnosis (Coppersmith et al., 2014, 2015; Cohan et al., 2018) for mental health disorders in music comments. Two Chinese native speakers further manually filter for genuine statements of disclosure (i.e., excluding jokes, quotes, and clearly disingenuous statements), resulting in 1133 users with self-reported mental health disorders. We find that, out of all disclosers, most disclose depression (81.2%), anxiety (19.9%), and bipolar (18.5%) disorders; additionally, most users (60.6%) self-identify as women, consistent with constituent gender differences of affective disorders in national studies (Huang et al., 2019). Disclosers show greater platform usage (Kolmogorov-Smirnov on user levels,  $p < 0.01$ ), insomnia-aligned diurnal user activity consistent with disorder symptoms (Taylor et al., 2005; Harvey, 2008), increased engagement with playlists of sadder natures, e.g., as shown in Figure 4, loneliness (+302%), sadness (+158%), and night (+50.1%), and decreased engagement with playlists of more active natures, e.g., exercise (-51.7%), compared to typical users. These observations mirror affective disorder activity trends (Cooney et al., 2013) and suggest that people with affective disorders are more likely to use music reflective of negative emotions than positive emotions to manage feelings of sadness and depression (Stewart et al., 2019). A detailed breakdown of our data, comorbidities, and our specific regular expressions are described in Appendix Section F.

**Affective Response.** Treating the act of self-disclosure as an affective response, we test for factors driving this behavior. Statements of self-

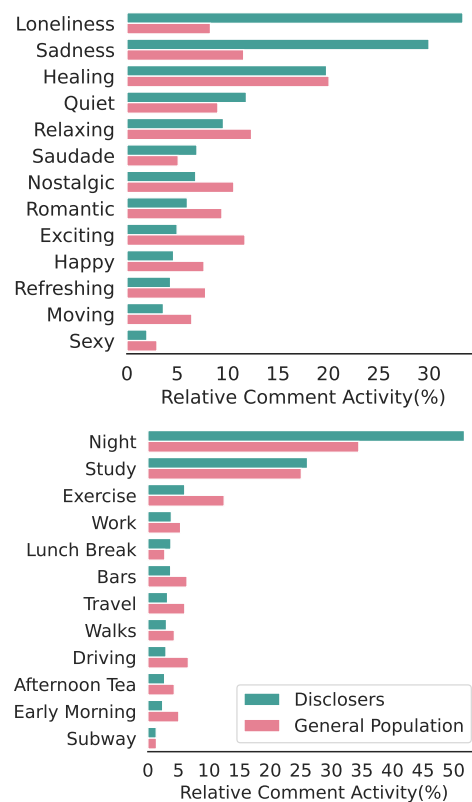


Figure 4: Relative tagged playlist commenting activity between disclosers and the set of all users on emotion (top) and setting (bottom) tagged playlists. Note that as each playlist may have up to three unique tags, relative tag percentages do not add up to 100%. The complete set of figures for all playlist tag categories are shown in Appendix Section F, Figure 6.

disclosure are more likely to appear as top-level comments (78.1%) than as replies (21.9%); top-level disclosures are biased towards songs with features generally associated with sadness (Juslin and Laukka, 2004) ( $p < 0.01$  for all features other than tempo, Kolmogorov-Smirnov), i.e. softer songs with minor modes, and towards playlists with tags of the same nature, while disclosures in reply occur to comments that indicate emotional distress, or that are themselves replies to existing comments made by the discloser (“怎么了”, meaning “what’s wrong”). Responses to the first are split in function, with disclosers either expressing their diagnosis in empathy for encouragement (“...我一年前也确诊了, 事情会好起来的”, meaning “...I was diagnosed a year ago, things will be better”), or to commiserate (“我也确诊了, 活着好难...”, meaning “I was diagnosed too, living is so hard...”), showing evidence of resonance (Miller, 2015; Rosa, 2019) and high person-centered condolence (High and Dillard, 2012).

<sup>7</sup>See Limitations for a discussion on method caveats and standards of diagnosis in clinical psychology.



**Social Support.** Characterizing audience engagement around self-disclosure comments in their content, we identify supportive comments according to the four major classes of social support around health concerns—prescriptive, informational, instrumental, and emotional support—from established literature (Turner et al., 1983; George et al., 1989; De Choudhury and De, 2014) and label the main type of support each comment falls under. We then fit logistic regression models on the dependent variable of receipt, aiming to identify *where* users are more likely to receive a supportive comment in response to disclosure; including song features as independent variables and song popularity, comment length, user demographics, and comment LDA topic distributions as controls. We observe that emotional (52%, e.g., “加油, 事情一定会好起来的我保证”, meaning “good luck, everything will be better I promise”) and prescriptive support (31%, e.g., “听一些令人振奋的歌曲吧”, meaning “listen to heart raising songs”) largely exceeds informational (9%, e.g., “...治疗可能会有帮助, 两年治疗后我...”, meaning “...therapy could help, after two years of therapy I...”) and instrumental (8%, e.g., “...你可以私聊我”, meaning “...you can private message me”) forms in response to disclosures. Several psycholinguistic lyrical features proved statistically significant ( $p < 0.05$ ) in predicting if a disclosure comment to a song would receive a supportive reply; the rate of terms in lyrics relating to social processes, specifically friend (+2.23) and ingest (+0.97), positively predict this prosocial behavior, and negative emotion terms (-2.04) do so negatively, mirroring negative correlations between sadness and prosocial tendencies (Ye et al., 2020). For musical features, only reverberation did so positively (+0.90). While past work has studied the prosocial effects of music, most have only used a limited set of author-chosen songs (Greitemeyer, 2009; Kennedy, 2013) or crowd-sourced prosocial perceptions (Ruth, 2017); here, we specifically identify what makes for prosocial songs and situate our study in the context of social support to mental health self-disclosures. Taken together, these observations not only provide ample pointers for music therapists on musical and dyadic conversational means for more successful emotion-focused interventions (Jensen, 2001) but also guide users on how to effectively find social support on the platform when needed (De Choudhury and Kiciman, 2017).

## 6 Discussion and Future Work

In this work, we sought to examine the driving factors behind variations in emotional reactions to music, via a large-scale computational study of a Chinese social music platform. Our analyses here reveal several nuances in how idiosyncratic variables elicit emotional responses, with a degree of precision that prior studies have often lacked thus far. In a case study of mental health self-disclosures in music comments, we characterized a type of discourse in the context of a popular social phenomenon, demonstrated the importance of posting location in determining the social support disclosures would receive, and revealed several factors driving the prosociality of music in this context. We see our present work situated in the broader context of studying emotionality in music and in the design of platforms to promote healthier interactions more centered on user well-being. Here, we highlight a few limitations and directions for future work; models, code, and anonymized data are made available at <https://github.com/skychwang/music-emotions>.

The music we listen to has a strong effect on our moods (McCraty et al., 1998). The integration of emotional response analysis into music recommendation systems could promote healthier recommendations (Konstan and Riedl, 2012; Singh et al., 2020) more cognizant of listener well-being outcomes. No one size fits all, and more sophisticated analyses could better capture more factors that explain emotional response variations towards creating more personalized music emotion recommendation systems.

While our work measures the effects of demographic variables on emotional responses, there remains a bio-psycho-social question on identifying the causes behind *why* this variation exists as it relates to song features. Lived experiences condition our emotions (Brody, 1997); future work could aim, through significant theoretical and qualitative study, to better identify the relationships and causes behind these variation outcomes.

Several open questions also remain as to whether *risk* may be qualified in this context in relation to well-being. Specifically, it would be interesting to study how recommendation interactions may disproportionately affect those afflicted with mental health disorders, and how we may design platforms, in the context of well-being outcomes, under normative goals of equity and distributive justice (Rawls, 2001).

## Ethical Considerations

**Data Release.** For user comments, taking user privacy considerations into account, we release the set of comment ids used in our analyses—which researchers are able to use in conjunction with the Netease API to obtain original comment content—mirroring Twitter data release guidelines for academic research.

**Identity Affiliation.** In studying demographic effects, we examine only the aggregate behavior of users who make public their demographic self-identification *choices* during registration under platform constraints. In particular, we note that platform choices for gender are limited only to binary options—男 men and 女 women. These choices should not be interpreted to have taken into account gender fluidity considerations or the multidimensional spectrum of gender identities (Larson, 2017).

## Limitations

**Measuring Affective Response.** In particular, we mirror the concerns by Mohammad (2020); notably, that (1) emotion lexicons are limited in coverage and do not include all possible terms in a language, and that (2) as languages and, in particular, our perceptions of words in them are by nature entities of change that inherently possess socio-cultural variations, emotion scores for words are not immutable, neither longitudinally nor socio-culturally. As such, while we have attempted to mitigate for this limitation by (1) choosing the largest Chinese emotion lexicon annotated for words sourced from the domain of social media and (2) comparing our findings to that of previous smaller-scale in-person studies that use varying methods to measure emotion when possible—even as no “gold standard” measure of emotional response exists, physiological, behavioral, or otherwise (Mauss and Robinson, 2009)—we encourage future work to further examine these phenomena in a greater variety of contexts. Further, our study does not make explicit causal claims around factors of music choice and user predisposition, i.e. what caused users to choose to listen to a specific song, or what their states of mind were prior to making this choice. While our work shows evidence of variations in affective responses correlated with musical, lyrical, demographic, and mental health factors, like the quasi-causal results estimating demographic effects on listener affective responses, we do not argue that these alone explain the entirety of the associated variations. In moving

towards truly causal studies (Feder et al., 2021), we encourage further direct participatory work to further examine these observations in larger, more controlled, and even cross-cultural contexts.

**Censorship and Moderation.** Users are able to report comments that violate platform rules,<sup>8</sup> and active moderation of user content exists on the platform. As we use only public posts on the platform, it is thus important to interpret our findings in the context of internet censorship in China (Vuori and Paltemaa, 2015). In particular, as noted by previous studies on mental health postings in Chinese social media (Cui et al., 2022), comments that go against certain government objectives—such as the “stability and unity for a harmonious society” (Wang, 2012), which mental health-related postings may go against—are often censored (Paltemaa et al., 2020). While pilot tests matching regular expressions on such phrases within platform comments still yielded significant quantities, the degree of censorship that these types of comments receive remains unclear.

**Statements of Diagnosis.** As we study users with self-reported statements of diagnosis, our method only potentially captures a sub-population of each disorder—those who choose to disclose a diagnosis on a public platform under the option of anonymity. While we have attempted to increase the precision of identifying individuals who are diagnosed with specific disorders through significant manual annotation, in the lack of clinically-aligned user data, we nonetheless are unable to verify if genuine-appearing disclosures of mental health disorder diagnoses are ultimately truthful. However, as noted by (Coppersmith et al., 2014), given the stigmas often associated with mental illnesses, it seems unlikely that users would disclose that they were diagnosed with a condition they do not possess. Individuals who may be diagnosed with affective disorders undoubtedly also remain in the set of all users that we compare disclosers against and, as such, our results on platform user activity differences should only be interpreted in the context of discovering broad themes—not as ground truths of comparisons between those who are diagnosed and those who aren’t. Finally, we also note concerns in clinical psychology on the heterogeneity of psychiatric diagnoses, which remains contentious in current literature. Notably, that standards of diagnosis all use different decision-making

<sup>8</sup><http://music.163.com/m/topic/13336053>

rules, that significant overlaps exist in symptoms between diagnoses, and that they may instead mask the complex underlying causes of human distress with potentially scientifically meaningless labels (Allsopp et al., 2019).

## Acknowledgements

We thank David Jurgens, Michelle Cohn, Xiang Zhou, Maximillian Chen, Kexin Fan, and the anonymous reviewers for their helpful comments, thoughts, and discussions. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-2036197.

## References

- Kate Allsopp, John Read, Rhiannon Corcoran, and Peter Kinderman. 2019. Heterogeneity in psychiatric diagnostic classification. *Psychiatry research*, 279:15–22.
- American Psychiatric Association. 2022. *Diagnostic and statistical manual of mental disorders, fifth edition, text revision*. American psychiatric association Washington, DC.
- Rowland Atkinson and John Flint. 2001. Accessing hidden and hard-to-reach populations: Snowball research strategies. *Social research update*, 33(1):1–4.
- Syeda Beenish Bareeqa, Syed Ijlal Ahmed, Syeda Sana Samar, Waqas Yasin, Sani Zehra, George M Monese, and Robert V Gouthro. 2021. Prevalence of depression, anxiety and stress in china during covid-19 pandemic: A systematic review with meta-analysis. *The International Journal of Psychiatry in Medicine*, 56(4):210–227.
- Natalya N Bazarova and Yoon Hyung Choi. 2014. Self-disclosure in social media: Extending the functional approach to disclosure motivations and characteristics on social network sites. *Journal of Communication*, 64(4):635–657.
- Sandra L Bem. 1974. The measurement of psychological androgyny. *Journal of consulting and clinical psychology*, 42(2):155.
- Scott Beveridge and Don Knox. 2018. Popular music and the role of vocal melody in perceived emotion. *Psychology of Music*, 46(3):411–423.
- Margaret M Bradley, Maurizio Codispoti, Dean Sabatinelli, and Peter J Lang. 2001. Emotion and motivation ii: sex differences in picture processing. *Emotion*, 1(3):300.
- Leslie R Brody. 1997. Gender and emotion: Beyond stereotypes. *Journal of Social issues*, 53(2):369–393.
- Leslie R Brody and Judith A Hall. 2010. Gender, emotion, and socialization. In *Handbook of gender research in psychology*, pages 429–454. Springer.
- Michaëlle Cameron, Julie Baker, and Mark Peterson. 2013. Waiting for service: The effects of music volume and gender. *Services Marketing Quarterly*, 34(4):257–273.
- Xuqian Chen, Shengqiao Huang, Xueting Hei, and Hongyuan Zeng. 2020. Felt emotion elicited by music: are sensitivities to various musical features different for young children and young adults? *The Spanish Journal of Psychology*, 23.
- Arman Cohan, Bart Desmet, Andrew Yates, Luca Soldaini, Sean MacAvaney, and Nazli Goharian. 2018. Smhd: a large-scale resource for exploring online language usage for multiple mental health conditions. In *27th International Conference on Computational Linguistics*, pages 1485–1497. ACL.
- Gary M Cooney, Kerry Dwan, Carolyn A Greig, Debbie A Lawlor, Jane Rimer, Fiona R Waugh, Marion McMurdo, and Gillian E Mead. 2013. Exercise for depression. *Cochrane database of systematic reviews*, (9).
- Glen Coppersmith, Mark Dredze, and Craig Harman. 2014. Quantifying mental health signals in twitter. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, pages 51–60.
- Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015. From adhd to sad: Analyzing the language of mental health on twitter through self-reported diagnoses. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, pages 1–10.
- Mihaly Csikszentmihalyi and Judith LeFevre. 1989. Optimal experience in work and leisure. *Journal of personality and social psychology*, 56(5):815.
- Jesse Cui, Tingdan Zhang, Dandan Pang, Kokil Jaidka, Garrick Sherman, Vinit Jakhethiya, Lyle Ungar, and Sharath Chandra Guntuku. 2022. Social media reveals urban-rural differences in stress across china. *ICWSM*.
- Sam de Boise. 2016. Contesting ‘sex’ and ‘gender’ difference in emotions through music use in the uk. *Journal of Gender Studies*, 25(1):66–84.
- Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Eighth international AAAI conference on weblogs and social media*.
- Munmun De Choudhury and Emre Kiciman. 2017. The language of social support in social media and its effect on suicidal ideation risk. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, pages 32–41.

- Tia DeNora. 1999. Music as a technology of the self. *Poetics*, 27(1):31–56.
- Tia DeNora. 2000. *Music in everyday life*. Cambridge University Press.
- Stuart Dredge. 2022. Netease cloud music reveals its revenues grew 43% in 2021.
- Tuomas Eerola, Rafael Ferrer, and Vinoo Alluri. 2012. Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Perception: An Interdisciplinary Journal*, 30(1):49–70.
- Tuomas Eerola, Anders Friberg, and Roberto Bresin. 2013. Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Frontiers in psychology*, 4:487.
- Daniel PW Ellis. 2007. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60.
- Sindhu Kiranmai Ernala, Tristan Labetoulle, Fred Bane, Michael L Birnbaum, Asra F Rizvi, John M Kane, and Munmun De Choudhury. 2018. Characterizing audience engagement and assessing its impact on social media disclosures of mental illnesses. In *Twelfth international AAAI conference on web and social media*.
- Amir Feder, Katherine A Keith, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar, Zach Wood-Doughty, Jacob Eisenstein, Justin Grimmer, Roi Reichart, Margaret E Roberts, et al. 2021. Causal inference in natural language processing: Estimation, prediction, interpretation and beyond. *arXiv preprint arXiv:2109.00725*.
- Cristina Fernández, Juan C Pascual, Joaquim Soler, Matilde Elices, Maria J Portella, and Enrique Fernández-Abascal. 2012. Physiological responses induced by emotion-eliciting films. *Applied psychophysiology and biofeedback*, 37(2):73–79.
- Frederic Font, Tim Brookes, George Fazekas, Martin Guerber, Amaury La Burthe, David Plans, Mark d. Plumbley, Meir Shaashua, Wenwu Wang, and Xavier Serra. 2016. audio commons: bringing creative commons audio content to the creative industries. *journal of the audio engineering society*.
- Alf Gabrielsson and Siv Lindström. 2010. Strong experiences with music. *Handbook of music and emotion: Theory, research, applications*, pages 547–574.
- Sandra Garrido, Catherine J Stevens, Esther Chang, Laura Dunne, and Janette Perz. 2018. Music and dementia: individual differences in response to personalized playlists. *Journal of Alzheimer's Disease*, 64(3):933–941.
- Linda K George, Dan G Blazer, Dana C Hughes, and Nancy Fowler. 1989. Social support and the outcome of major depression. *The British Journal of Psychiatry*, 154(4):478–485.
- Scott A Golder and Michael W Macy. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881.
- Patrick Gomez and Brigitta Danuser. 2007. Relationships between musical structure and psychophysiological measures of emotion. *Emotion*, 7(2):377.
- Juan Sebastián Gómez-Cañón, Estefanía Cano, Tuomas Eerola, Perfecto Herrera, Xiao Hu, Yi-Hsuan Yang, and Emilia Gómez. 2021. Music emotion recognition: Toward new, robust standards in personalized and context-sensitive applications. *IEEE Signal Processing Magazine*, 38(6):106–114.
- Alinka E Greasley and Alexandra Lamont. 2011. Exploring engagement with music in everyday life using experience sampling methodology. *Musicae Scientiae*, 15(1):45–71.
- Andrew H Gregory and Nicholas Varney. 1996. Cross-cultural comparisons in the affective response to music. *Psychology of Music*, 24(1):47–52.
- Tobias Greitemeyer. 2009. Effects of songs with prosocial lyrics on prosocial behavior: Further evidence and a mediating mechanism. *Personality and Social Psychology Bulletin*, 35(11):1500–1511.
- Nicolás Felipe Gutiérrez Páez, Juan Sebastián Gómez-Cañón, Lorenzo Porcaro, Patricia Santos, Davinia Hernández-Leo, and Emilia Gómez. 2021. Emotion annotation of music: A citizen science approach. In *International Conference on Collaboration Technologies and Social Computing*, pages 51–66. Springer.
- Julia C Hailstone, Rohani Omar, Susie MD Henley, Chris Frost, Michael G Kenward, and Jason D Warren. 2009. It's not what you play, it's how you play it: Timbre affects perception of emotion in music. *Quarterly journal of experimental psychology*, 62(11):2141–2155.
- Keith Harrigan, Carlos Aguirre, and Mark Dredze. 2020. On the state of social media data for mental health research. *arXiv preprint arXiv:2011.05233*.
- Allison G Harvey. 2008. Sleep and circadian rhythms in bipolar disorder: seeking synchrony, harmony, and regulation. *American journal of psychiatry*, 165(7):820–829.
- Shigeo Hatano and Takeo Hashimoto. 2000. Booming index as a measure for evaluating booming sensation. In *Proc. Inter-Noise*, 233, pages 1–6.
- Kate Hevner. 1935. Expression in music: a discussion of experimental studies and theories. *Psychological review*, 42(2):186.
- Andrew C High and James Price Dillard. 2012. A review and meta-analysis of person-centered messages and social support outcomes. *Communication Studies*, 63(1):99–118.



- Yoshiyuki Hirano, Masafumi Fujita, Kazuko Watanabe, Masami Niwa, Toru Takahashi, Masayuki Kanematsu, Yasushi Ido, Mihoko Tomida, and Minoru Onozuka. 2006. Effect of unpleasant loud noise on hippocampal activities during picture encoding: an fmri study. *Brain and cognition*, 61(3):280–285.
- Annabell Ho, Jeff Hancock, and Adam S Miner. 2018. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, 68(4):712–733.
- Xiao Hu, Fanjie Li, and Tzi-Dong Jeremy Ng. 2018. On the relationships between music-induced emotion and physiological signals. In *ISMIR*, pages 362–369.
- Chin-Lan Huang, Cindy K Chung, Natalie Hui, Yi-Cheng Lin, Yi-Tai Seih, Ben CP Lam, Wei-Chuan Chen, Michael H Bond, and James W Pennebaker. 2012. The development of the chinese linguistic inquiry and word count dictionary. *Chinese Journal of Psychology*.
- Yueqin Huang, YU Wang, Hong Wang, Zhaorui Liu, Xin Yu, Jie Yan, Yaqin Yu, Changgui Kou, Xiufeng Xu, Jin Lu, et al. 2019. Prevalence of mental disorders in china: a cross-sectional epidemiological study. *The Lancet Psychiatry*, 6(3):211–224.
- Patrick G Hunter and E Glenn Schellenberg. 2010. Music and emotion. In *Music perception*, pages 129–164. Springer.
- Guido W Imbens. 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and statistics*, 86(1):4–29.
- Tariquillah Jan and Wenwu Wang. 2012. Blind reverberation time estimation based on laplace distribution. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 2050–2054. IEEE.
- Kaja L Jensen. 2001. The effects of selected classical music on self-disclosure. *Journal of music therapy*, 38(1):2–27.
- Patrik N. Juslin. 2013. From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews*, 10(3):235–266.
- Patrik N Juslin. 2019. *Musical emotions explained: Unlocking the secrets of musical affect*. Oxford University Press, USA.
- 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):217–238.
- Patrik N Juslin, Simon Liljeström, Daniel Västfjäll, Gonçalo Barradas, and Ana Silva. 2008. An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion*, 8(5):668.
- Patrik N Juslin and Daniel Västfjäll. 2008. Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and brain sciences*, 31(5):559–575.
- Stuart B Kamenetsky, David S Hill, and Sandra E Trehub. 1997. Effect of tempo and dynamics on the perception of emotion in music. *Psychology of Music*, 25(2):149–160.
- Patrick Kennedy. 2013. The relationship between prosocial music and helping behaviour and its mediators: An irish college sample. *Journal of European Psychology Students*, 4(1).
- Stefan Koelsch. 2014. Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, 15(3):170–180.
- Joseph A Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. *User modeling and user-adapted interaction*, 22(1):101–123.
- Herbert E Krugman. 1943. Affective response to music as a function of familiarity. *The Journal of Abnormal and Social Psychology*, 38(3):388.
- Carol L Krumhansl. 2001. *Cognitive foundations of musical pitch*, volume 17. Oxford University Press.
- Brian Larson. 2017. Gender as a variable in natural-language processing: Ethical considerations. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 1–11, Valencia, Spain. Association for Computational Linguistics.
- Ying Liu, Guangyuan Liu, Dongtao Wei, Qiang Li, Guangjie Yuan, Shifu Wu, Gaoyuan Wang, and Xingcong Zhao. 2018. Effects of musical tempo on musicians’ and non-musicians’ emotional experience when listening to music. *Frontiers in Psychology*, page 2118.
- Jared K Lunceford and Marie Davidian. 2004. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in medicine*, 23(19):2937–2960.
- Lars-Olov Lundqvist, Fredrik Carlsson, Per Hilmersson, and Patrik N Juslin. 2009. Emotional responses to music: Experience, expression, and physiology. *Psychology of music*, 37(1):61–90.
- Iris B Mauss and Michael D Robinson. 2009. Measures of emotion: A review. *Cognition and emotion*, 23(2):209–237.
- Rollin McCraty, Bob Barrios-Choplin, Michael Atkinson, and Dana Tomasino. 1998. The effects of different types of music on mood, tension, and mental clarity. *Alternative therapies in health and medicine*, 4(1):75–84.

- Brian McFee, Colin Raffel, Dawen Liang, Daniel P Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. 2015. librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, volume 8, pages 18–25. Citeseer.
- Leonard B Meyer. 1956. *Emotion and meaning in music*. University of Chicago Press.
- Rada Mihalcea and Carlo Strapparava. 2012. Lyrics, music, and emotions. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 590–599.
- Vincent Miller. 2015. Resonance as a social phenomenon. *Sociological Research Online*, 20(2):58–70.
- Saif Mohammad, Xiaodan Zhu, and Joel Martin. 2014. Semantic role labeling of emotions in tweets. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 32–41, Baltimore, Maryland. Association for Computational Linguistics.
- Saif M Mohammad. 2020. Practical and ethical considerations in the effective use of emotion and sentiment lexicons. *arXiv preprint arXiv:2011.03492*.
- National Bureau of Statistics of China. 2021. Main data of the seventh national population census.
- Lauri Paltemaa, Juha A Vuori, Mikael Mattlin, and Jouko Katajisto. 2020. Meta-information censorship and the creation of the chinanet bubble. *Information, Communication & Society*, 23(14):2064–2080.
- Michael Paul. 2017. Feature selection as causal inference: Experiments with text classification. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 163–172.
- Andy Pearce, Tim Brookes, and Russell Mason. 2017. First prototype of timbral characterisation tool for semantically annotating non-musical. *Audio Commons project deliverable D, 5*.
- Andy Pearce, Tim Brookes, and Russell Mason. 2019. Modelling timbral hardness. *Applied Sciences*, 9(3):466.
- James W Pennebaker. 2011. The secret life of pronouns. *New Scientist*, 211(2828):42–45.
- James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015. Technical report.
- John Rawls. 2001. *Justice as fairness: A restatement*. Harvard University Press.
- Claudio Robazza, Cristina Macaluso, and Valentina D’Urso. 1994. Emotional reactions to music by gender, age, and expertise. *Perceptual and Motor skills*, 79(2):939–944.
- James Robert, Marc Webbie, et al. 2018. *Pydub*.
- Hartmut Rosa. 2019. *Resonance: A sociology of our relationship to the world*. John Wiley & Sons.
- Paul R Rosenbaum and Donald B Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.
- Paul R Rosenbaum and Donald B Rubin. 1984. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association*, 79(387):516–524.
- James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.
- Nicolas Ruth. 2017. “heal the world”: A field experiment on the effects of music with prosocial lyrics on prosocial behavior. *Psychology of Music*, 45(2):298–304.
- Suvi Saarikallio, Sirke Nieminen, and Elvira Brattico. 2013. Affective reactions to musical stimuli reflect emotional use of music in everyday life. *Musicae Scientiae*, 17(1):27–39.
- Katharina Schäfer, Suvi Saarikallio, and Tuomas Eerola. 2020. Music may reduce loneliness and act as social surrogate for a friend: evidence from an experimental listening study. *Music & Science*, 3:2059204320935709.
- Karl Schriewer and Grzegorz Bulaj. 2016. Music streaming services as adjunct therapies for depression, anxiety, and bipolar symptoms: convergence of digital technologies, mobile apps, emotions, and global mental health. *Frontiers in public health*, 4:217.
- Emery Schubert. 2004. Modeling perceived emotion with continuous musical features. *Music perception*, 21(4):561–585.
- Ignacio Siles, Andrés Segura-Castillo, Mónica Sancho, and Ricardo Solís-Quesada. 2019. Genres as social affect: Cultivating moods and emotions through playlists on spotify. *Social Media+ Society*, 5(2):2056305119847514.
- Ashudeep Singh, Yoni Halpern, Nithum Thain, Konstantina Christakopoulou, E Chi, Jilin Chen, and Alex Beutel. 2020. Building healthy recommendation sequences for everyone: A safe reinforcement learning approach. In *Proceedings of the FAccTRec Workshop, Online*, pages 26–27.
- John A Sloboda, Susan A O’Neill, and Antonia Ivaldi. 2001. Functions of music in everyday life: An exploratory study using the experience sampling method. *Musicae scientiae*, 5(1):9–32.
- John A Sloboda and Susan A O’Neill. 2001. Emotions in everyday listening to music. *Music and emotion: Theory and research*, 8:415–429.

- Joanna Stewart, Sandra Garrido, Cherry Hense, and Katrina McFerran. 2019. Music use for mood regulation: Self-awareness and conscious listening choices in young people with tendencies to depression. *Frontiers in psychology*, 10:1199.
- Daniel J Taylor, Kenneth L Lichstein, H Heith Durrence, Brant W Reidel, and Andrew J Bush. 2005. Epidemiology of insomnia, depression, and anxiety. *Sleep*, 28(11):1457–1464.
- R Jay Turner, B Gail Frankel, and Deborah M Levin. 1983. Social support: Conceptualization, measurement, and implications for mental health. *Research in community & mental health*.
- Marjolein D Van der Zwaag, Joyce HDM Westerink, and Egon L Van den Broek. 2011. Emotional and psychophysiological responses to tempo, mode, and percussiveness. *Musicae Scientiae*, 15(2):250–269.
- Pantelis N Vassilakis and K Fitz. 2007. Sra: A web-based research tool for spectral and roughness analysis of sound signals. In *Proceedings of the 4th Sound and Music Computing (SMC) Conference*, pages 319–325.
- Daniel Västfjäll, Patrik N Juslin, and Terry Hartig. 2012. Music, subjective wellbeing, and health: The role of everyday emotions. *Music, health, and wellbeing*, pages 405–423.
- Juha Antero Vuori and Lauri Paltemaa. 2015. The lexicon of fear: Chinese internet control practice in sina weibo microblog censorship. *Surveillance & society*, 13(3/4):400–421.
- Jonna K Vuoskoski and Tuomas Eerola. 2015. Extramusical information contributes to emotions induced by music. *Psychology of Music*, 43(2):262–274.
- Zachary Wallmark, Marco Iacoboni, Choi Deblieck, and Roger A Kendall. 2018. Embodied listening and timbre: Perceptual, acoustical, and neural correlates. *Music Perception: An Interdisciplinary Journal*, 35(3):332–363.
- Han Wang and RongRong Fu. 2020. Exploring user experience of music social mode-take netease cloud music as an example. In *International Conference on Applied Human Factors and Ergonomics*, pages 993–999. Springer.
- Shaojung Sharon Wang. 2012. China’s internet lexicon: Symbolic meaning and commoditization of grass mud horse in the harmonious society. *First Monday*.
- Gregory D Webster and Catherine G Weir. 2005. Emotional responses to music: Interactive effects of mode, texture, and tempo. *Motivation and Emotion*, 29(1):19–39.
- Diyi Yang, Zheng Yao, Joseph Seering, and Robert Kraut. 2019. The channel matters: Self-disclosure, reciprocity and social support in online cancer support groups. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–15.
- Yi-Hsuan Yang, Ya-Fan Su, Yu-Ching Lin, and Homer H Chen. 2007. Music emotion recognition: The role of individuality. In *Proceedings of the international workshop on Human-centered multimedia*, pages 13–22.
- Yingying Ye, Tingting Long, Cuizhen Liu, and Dan Xu. 2020. The effect of emotion on prosocial tendency: the moderating effect of epidemic severity under the outbreak of covid-19. *Frontiers in psychology*, 11:588701.
- Liang-Chih Yu, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Robert Lai, and Xuejie Zhang. 2016. Building Chinese affective resources in valence-arousal dimensions. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 540–545, San Diego, California. Association for Computational Linguistics.
- Zhenkun Zhou, Ke Xu, and Jichang Zhao. 2018. Homophily of music listening in online social networks of china. *Social Networks*, 55:160–169.
- Eberhard Zwicker and Hugo Fastl. 2013. *Psychoacoustics: Facts and models*, volume 22. Springer Science & Business Media.

## Overview of Appendix

We provide, as supplementary material, additional information about our dataset, annotation guidelines, preprocessing details, and expanded results across all experiments.

### A Data

This section describes summary statistics of our data, as well as a view of the platform’s user interface.

#### A.1 Platform Interface

Users are able to interact with the platform through their browsers, native OS applications, and phone apps. Screenshots of a song’s interface are shown in Figure 7, as is a view of the iOS application’s commenting page for a song.

#### A.2 Users

User age, gender, and region distributions (Figure 8) show that the majority of users are young men that hail from major metropolitan areas. The top 20 regions that users hail from are, in descending order, Beijing (4.71%), Guangzhou (4.22%), Shanghai (3.80%), Chengdu (3.47%), Shenzhen (2.58%), Chongqing (2.56%), Nanjing (2.51%), Wuhan (2.43%), Hangzhou (2.21%), Changsha (1.95%), Xian, (1.91%), Overseas-Other (1.77%), Zhengzhou (1.54%), Hefei (1.30%), Tianjin (1.28%), Suzhou (1.22%), Kunming (1.20%), Urumqi (1.19%), Jinan (1.03%), Fuzhou (0.99%), Qingdao (0.91%), Nanning (0.89%), Nanchang (0.88%), Shenyang (0.83%), Harbin (0.83%), Foshan (0.77%), Dongguan (0.74%), Guiyang (0.74%), Shijiazhuang (0.73%), and Ningbo (0.68%). General trends for user gender and region taken in the context of user ages mirror population trends indicated by the Chinese Census (National Bureau of Statistics of China, 2021).

#### A.3 Songs

Song comment and comment token distributions are shown in Figure 9; lyric preprocessing and topic modeling details are in Appendix Section C.

#### A.4 Playlists

Playlist comment distributions and comment token distributions are shown in Figure 10. The top 20 most popular tags used for playlists are, in descending order, Language-Western (21.5%), Style-Pop (19.2%), Language-Chinese (15.1%),

Group	Valence (m./std.)	Arousal (m./std.)
Women	5.93/1.33	5.23/1.09
Men	5.72/1.38	5.14/1.13
10后	5.76/1.37	5.17/1.12
05后	5.79/1.40	5.21/1.10
00后	5.85/1.38	5.22/1.10
95后	5.80/1.36	5.17/1.11
90后	5.76/1.34	5.12/1.11
85后	5.74/1.36	5.13/1.12
80后	5.79/1.32	5.10/1.11
75后	5.75/1.32	5.08/1.10
70后	5.81/1.31	5.09/1.09
65后	5.81/1.33	5.13/1.11
60后	5.91/1.33	5.15/1.10
55后	5.81/1.36	5.18/1.11
50后	5.77/1.40	5.19/1.11

Table 2: Comment valence and arousal mean (m.) and standard deviations (std.) for demographic groups on gender and age.

Style-Electronic (13.4%), Emotion-Healing (7.66%), Theme-ACG (7.56%), Setting-Night (7.03%), Style-Soft (6.56%), Theme-Games (6.18%), Theme-Movies (6.10%), Emotion-Relaxing (5.87%), Emotion-Exciting (5.33%), Style-Rock (5.32%), Style-Rap (5.08%), Emotion-Nostalgic (4.91%), Emotion-Quiet (4.65%), Setting-Study (4.50%), Theme-Classics (4.11%), and Emotion-Sadness (4.09%). Note here that as playlists may each contain at most three tags, summing percentages for all tags exceeds 100%.

#### A.5 Albums

Album comment, comment token, and release date distributions are shown in Figure 11. Songs with at least one comment show exponential bias towards recently released music.

#### A.6 Artists

A distribution of the number of albums and songs per artist is shown in Figure 12.

#### A.7 Demographic Baselines

Users of different demographic groups have varying comment valence and arousal means and standard deviations. These statistics, stratified by demographic groups on gender and age, are shown in Table 2.



## B Emotion Annotation Guidelines

This section describes the annotation guidelines used by annotators in our pilot studies to determine, in top-level comments, (1) the emotion experiencer, or who was the primary experiencer of the emotions expressed in comments, and (2) the affective stimulus of the emotion expressed in the comment. Annotators consisted of two Chinese native speakers and were asked to annotate a set of 1000 randomly selected comments on the platform.

Annotators were first tasked with familiarizing themselves with the BRECHEMA framework of musically evoked emotions (Juslin, 2013) before being presented with the following questionnaires for annotation:

### Question 1: The Emotion Experiencer

**Comment:** 真特么的带感这曲子!

Q. Who was the primary experiencer of the emotion expressed in the comment?

- The commenter themselves.
- Someone other than the commenter themselves.
- This comment possesses no emotional content.

### Question 2: The Affective Stimulus

**Comment:** 真特么的带感这曲子!

Q. What was the primary affective stimulus of the emotion expressed in the comment?

- The song, album, or playlist.
- Something other than the song, album, or playlist.
- This comment possesses no emotional content.

As stated, annotators were asked to resolve initial annotation disagreements through discussion in order to come up with a set of annotations that both agreed on.

## C Lyric Topics and Preprocessing

This section describes our lyrical preprocessing methods and 20-topic LDA model results on song lyrics.

**Preprocessing.** We first identify instrumental music by matching lyric data on the substring 纯音乐, used by the platform to denote songs of this category. For non-instrumental pieces, we filter

out lines with song metadata (e.g. composers) by removing lines that match the following regex:

```
:|: | 《|》 |produced by|vocals  
by|recorded by|edited by|mixed  
by|mastered by| - | -
```

As repeated lyrics are denoted with overlapping time stamps, e.g.

```
[1:00.00] [2:00.00] 雨淋湿了天空
```

indicates that the line 雨淋湿了天空 is repeated at minutes 1 and 2), we further unfurl and reorder lines by timestamp, duplicating lines when necessary. Further tokenization details are shown in [Text Preprocessing](#).

**Topic Modeling.** We train a 20-topic LDA model on preprocessed song lyrics and manually label each lyric with its prominent theme. While some degree of variation exists for listener affective responses across songs of each topic, these topic distributions are primarily used as lyrical content controls in our regression models. Labeled topics and their top words are shown in [Table 3](#).

## D Text Preprocessing

This section describes our text preprocessing pipeline for all text data on the platform, namely (1) lyrics and (2) listener comments.

**Preprocessing.** We analyze only Chinese language content, using Google’s Compact Language Detector v3 (gclid3<sup>9</sup>) to detect text language and keep only Chinese language texts. We then convert all traditional Chinese characters to their simplified forms using `hanziconv`<sup>10</sup> to ensure consistency in our experiments—i.e. when calculating LIWC scores, for which we use the simplified Chinese version (Huang et al., 2012)—and finally tokenize with `jieba`.<sup>11</sup>

**Filtering for Affective Content.** Following annotations of listener comments, we filter out all comments that match the following regular expressions ([Table 4](#)), aiming to increase the precision of comments in our analysis that indicate an affective response. The following filters generally match with easily identifiable spam messages, i.e. “first comment”, album images, and quotations.

<sup>9</sup><https://github.com/google/cld3>

<sup>10</sup><https://pythonhosted.org/hanziconv/>

<sup>11</sup><https://github.com/fxsjy/jieba>

沙发,第一,第二,第三,第四,第五,第六,第七,第八,第九,第十,第1,第2,第3,第4,第5,第6,第7,第8,第9,一楼,留名,封面,没人,来晚了,板凳,求,前排,识曲,后排,一条,好少,不火,助攻,作者,评论,人呢,来了,“\*”,“\*”,“\*”,“\*”,《\*》,<\*>,: ,:,9+

Table 4: Regular expressions used to filter out irrelevant listener comments that do not indicate an affective response.

## E Expanded Results for Variations in Affective Response

Expanded results for variations in affective responses are shown in Figures 13 and 14 for musical features, Figures 15-18 for LIWC lyrical features, Figures 19-23 for settings and other playlist tags, and Figures 24-26 for demographic effects of gender and age.

## F Expanded Results for Disclosures of Mental Health Disorders

**Regular Expressions.** We source mental health disorders from the DSM-5-TR (American Psychiatric Association, 2022) and construct regular expressions to identify possible comments that self-disclose a diagnosis of a mental health disorder; specific regular expressions are shown below in Table 5. These regular expressions return 2319 matched comments in total.

**Manual Filtering.** Two Chinese native speakers then manually screened out comments that lacked a clear statement of self-disclosure. These comments primarily consisted of those that (1) described other people’s diagnoses, i.e. relatives, friends, or celebrities, (2) described recovery from a disorder, (3) were of speculative nature on the diagnosis, e.g. “我觉得...” (“I think...”), (4) described what a diagnosis was but not necessarily that the listener themselves was diagnosed, and (5) only used diagnosis terms to exaggerate sentiment. Examples of positive and false positive comments are shown below in Table 6. Following manual annotation, 46.9% of regular expression-matched comments were eliminated, leaving 1231 comments deemed as self-disclosures.

精神分裂,精分,人格分裂,妄想,思觉失调,强迫症,创伤后应激,情感障碍,情绪障碍,情绪失调,躁狂,狂躁,躁郁,双向,双相,抑郁,忧郁,重郁,轻郁,焦虑,社交焦虑,社焦,社交恐惧,人群恐惧,社恐,余恐,恐惧,恐慌,广场恐惧,分离焦虑,缄默,人格障碍,躯体变形,体象障碍,适应障碍,多重人格,人格解体,现实解体,创伤性失忆,解离性身份障碍,躯体症状,歇斯底里,转换症,转换障碍,做作性障碍,装病候群,代理性孟乔森,进食障碍,摄食障碍,神经性饮食失调,反刍,厌食,贪食,暴食,暴饮暴食,异食,失眠,睡眠障碍,嗜睡,睡眠相位后移,快动眼睡眠,睡瘫,睡眠瘫痪,梦魇症,易怒症,暴怒症,行为障碍,品行障碍,偏执,边缘性人格,边缘性人格,边缘型人格,边缘性格,做作.\*人格,自恋.\*人格,回避.\*人格,依赖.\*人格,精神病,心理疾病

Table 5: Mental health disorder condition name strings; these are prefixed/suffixed with the strings for “diagnosed” (“确诊.\*”) and “diagnosis” (“诊断.\*”), i.e. “确诊抑郁” for “diagnosed with depression”, to act as initial regular expression filters for users who self-disclose a diagnosis of a mental health disorder.

### Positives

今天,我被确诊为抑郁症了。  
去了大城市诊断的双相情感障碍  
524双子,确诊精分。  
诊断人格分裂。即使这样也要活下去啊!  
2017.12.25确诊焦虑症中度抑郁我很痛苦

### False Positives

医院等待医生上班确诊我是否患有抑郁症  
整的我有精神分裂症自我诊断上瘾了  
确诊中度抑郁症,现在已经走出来啦  
我已经几个朋友确诊抑郁了。  
我认知是抑郁症心境低落与其处境不相符

Table 6: Examples of positive and false positive self-disclosure statements of mental health disorders encountered in our manual labeling of comments matched with regular expressions. Partial comments are shown.

**Disorder Matches.** In total, 1133 users made self-disclosure statements. A breakdown of users by disorder class is shown in Table 7; note that the total users per class across all classes exceeds 1133 due to comorbidities.

Disclosed Disorder Class	Matched Users
Depressive	920
Anxiety	225
Bipolar and Related	201
Schizophrenia Spectrum and Other Psychotic	108
Sleep-Wake	35
Personality	18
Feeding and Eating	11
Obsessive-Compulsive and Related	5
Somatic Symptom and Related	4
Dissociative	1
Trauma- and Stressor-Related	1

Table 7: The number of users who self-disclose a mental health disorder, stratified over broad disorder classes.

**Diurnal User Activity.** Stratifying user activity across hours and measuring the relative comments made per stratum, we observe that disclosers show greater platform activity in the AMs (1-5 AM) and around 11AM-5PM compared to the set of all users. Shown in Figure 5 below, these observations are consistent with insomnia-aligned diurnal user activity, prevalent in individuals diagnosed with affective disorders (Taylor et al., 2005; Harvey, 2008). Note here that due to platform data limitations, while comment *dates* are available for all comments on the platform, only those made in the past year had *times* recorded and, as a result, are what we use in our analysis here on diurnal user activity. Thus, it is important to interpret these in the context of the COVID-19 pandemic, which has caused an increase in the prevalence of anxiety and depression worldwide (Bareeqa et al., 2021).

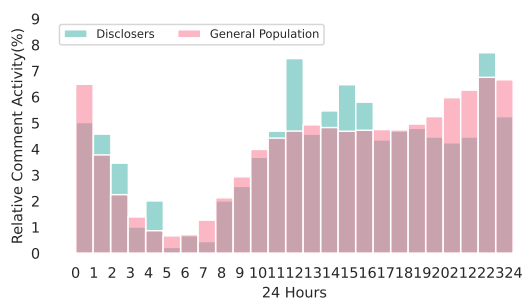


Figure 5: Diurnal commenting activity between disclosers and the set of all users.

**Playlist Engagement.** Relative tagged playlist engagements are shown in Figure 6 for disclosers and the set of all users; these are expanded figures as noted in Section 5 of the main paper. Notably, disclosers show greater engagement with emotion tagged playlists; within emotional and setting tags, disclosers show overwhelmingly greater engagement with tagged playlists of sadder nature, i.e. loneliness (+302%), sadness (+158%), and night (+50.1%), as well as decreased engagement with playlists of more active natures, e.g. exercise (-51.7%). These observations mirror affective disorder activity trends (Cooney et al., 2013) and suggest that people with affective disorders are more likely to use music reflective of negative emotions than positive emotions to manage feelings of sadness and depression (Stewart et al., 2019).

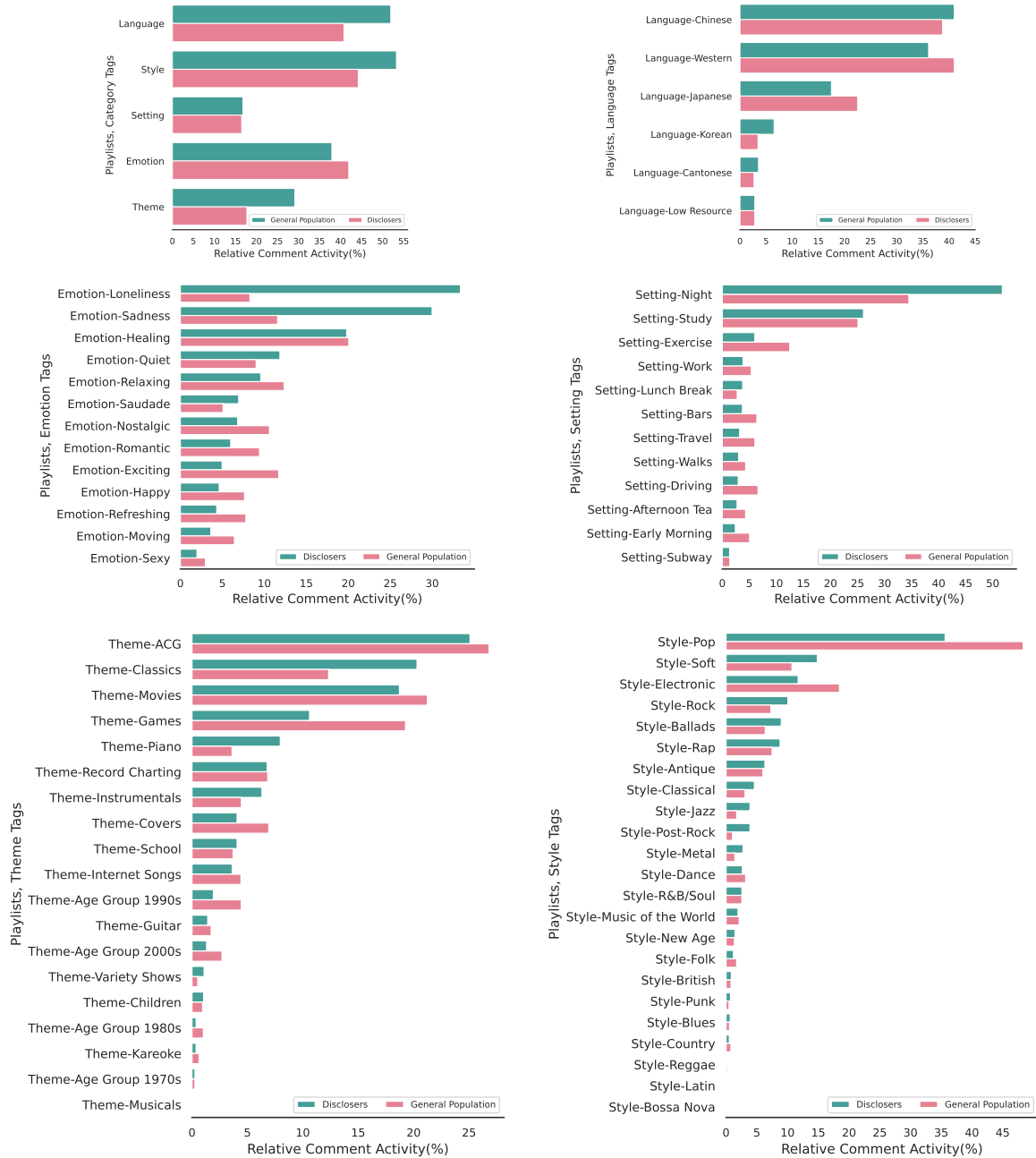


Figure 6: Relative tagged playlist commenting activity between disclosers and the set of all users. A breakdown of engagement with the five broad tag categories is shown on the **top left**, while other figures show each category's relative tag engagements. Note that as each playlist may have up to three unique tags, relative tag percentages do not add up to 100%.



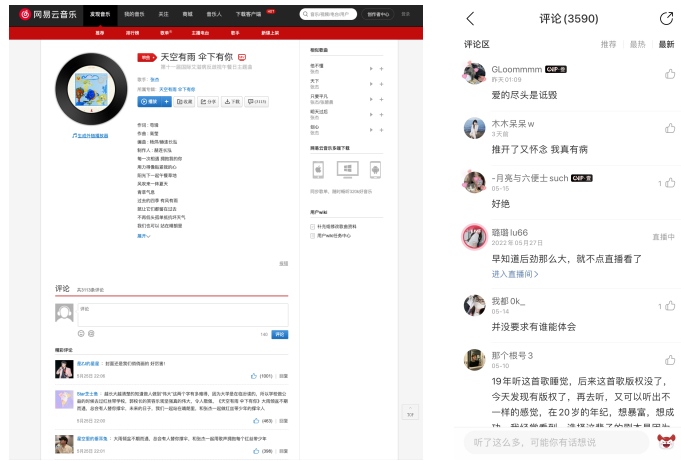


Figure 7: Screenshots of the platform’s in-browser web page interface (left), showing the description, lyrics, and comment board of a song, and iOS in-app interface (right), showing the comment board of a song.

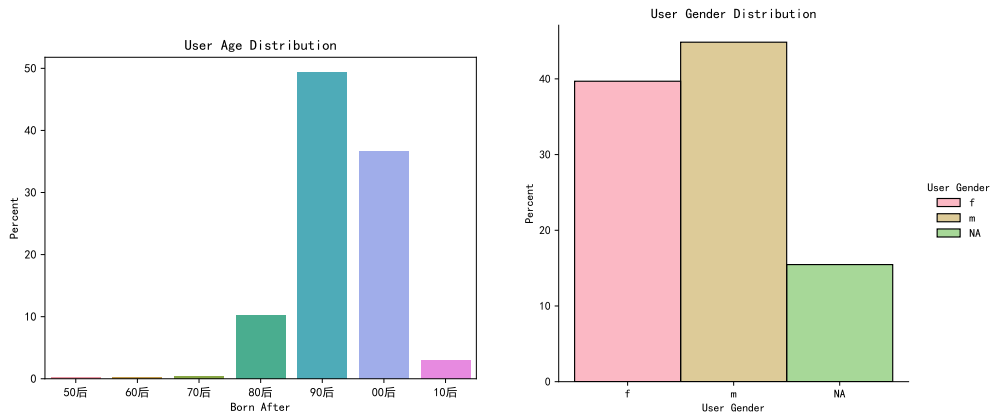


Figure 8: User age distributions (left) according to the decade of birth (i.e. 00后 for those born in the 2000s), and user gender distributions (right) across platform-available choices for gender. Here, NA implies that the user omitted to input gender information during registration.

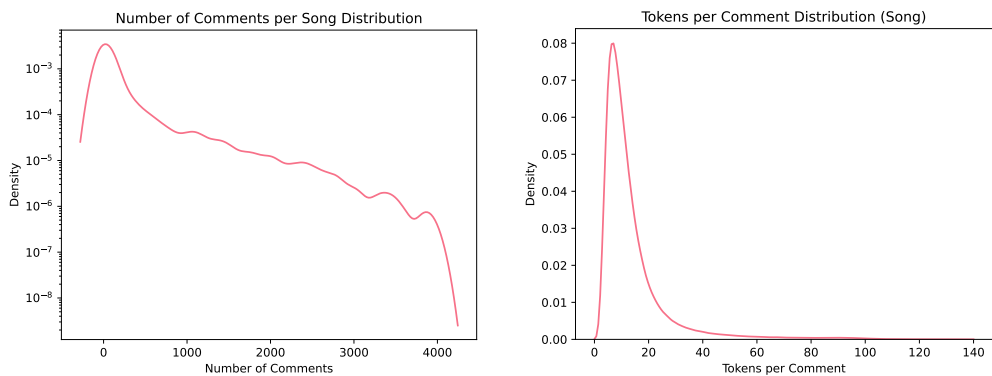


Figure 9: Comment (left) and comment token (right) distributions across all songs with at least one comment.

Topic Theme	Top Tokens
0. Romance/Sentiment (爱情/感性/伤感)	爱, 未, 今天, 便, 似, 一生, 令, 心, 没, 心中, 里, 太, 愿, 仍然, 想, 没法, 一起, 讲, 吻, 快乐
1. Youth/Hope/Warmth (青春/希望/阳光)	梦想, 希望, 地方, 梦, 世界, 身旁, 远方, 青春, 路, 模样, 时光, 方向, 未来, 流浪, 勇敢, 阳光, 带, 温暖, 生命, 心中
2. Transcendental (人生/社会/超俗)	人间, 江湖, 天地, 皆, 天下, 少年, 山河, 剑, 生死, 笑, 问, 间, 世间, 道, 万里, 便, 江山, 英雄, 合, 此生
3. Hometown/Childhood (故乡/童年)	花, 家, 牵挂, 长大, 噢, 回家, 记得, 回答, 说话, 画, 天涯, 挣扎, 走, 呐, 害怕, 变化, 落下, 傻, 年华, 故事
4. Friendship/Hedonism (享受/欲望/世俗)	吃, 不要, 没, 兄弟, 快, 音乐, 钱, 没有, 起来, 新, 听, 带, 买, 今天, 走, 站, 玩, 现在, 喝, 说唱
5. Love/Lust (恋爱/情欲)	喔, 女, 男, 合, 阮, 喝, 一杯, 一半, 讲, 耶, 人生, 爱, 伊, 酒, 甲, 拢, 啲呀, 啊啊啊, 心爱, 搁
6. Memories/Regret (从前/失望)	没有, 想, 没, 不会, 知道, 现在, 不想, 已经, 生活, 里, 太, 真的, 想要, 时间, 总是, 听, 曾经, 其实, 不能, 一直
7. Nature/Spring (阳光/故乡/自然/草原)	唱, 美丽, 姑娘, 飞, 月亮, 草原, 歌, 吹, 故乡, 春天, 开, 歌声, 轻轻, 歌唱, 亲爱, 太阳, 一片, 花儿, 远方, 阳光
8. Breakups/Sadness (分手/情感/失恋)	走, 没有, 爱, 手, 寂寞, 温柔, 快乐, 不要, 懂, 回头, 以后, 梦, 朋友, 难过, 自由, 不会, 最后, 记得, 沉默, 拥有
9. Nostalgia/Melancholy (桑感/忧愁/思念/思乡)	相思, 一曲, 醉, 听, 落, 岁月, 梦, 红尘, 明月, 桃花, 笑, 人间, 花, 叹, 不见, 故人, 春风, 似, 间, 清风, 见
10. Heartbreak/Loneliness (爱情/失恋/孤独/伤心)	爱, 心, 爱情, 眼泪, 哭, 太, 寂寞, 不要, 泪, 越, 女人, 心碎, 恨, 伤, 美, 幸福, 想, 错, 后悔, 不会
11. Wistful/Sentimental (思念/孤独)	梦, 一生, 情, 心, 愿, 心中, 今生, 难, 梦里, 往事, 雨, 问, 岁月, 泪, 匆匆, 人生, 如今, 相逢, 相思, 风雨
12. Family/Longing (家庭/思念)	妈妈, 唔, 哒, 想, 喵, 好想你, 爸爸, , 宝贝, 话, , 滴, 摇, 快, 系, 滴答, 讲, 一只, 哟, 笑
13. Celestial/Awe (孤独/渺小)	里, 风, 天空, 故事, 听, 记忆, 温柔, 城市, 雨, 回忆, 梦, 黑夜, 时光, 相遇, 声音, 风景, 夜空, 梦境, 流星
14. Love (爱情)	爱, 想, 爱情, 心, 忘记, 没有, 离开, 永远, 等待, 明白, 回忆, 不会, 未来, 不要, 我爱你, 相信, 一起, 不能, 愿意, 一次
15. Countryside/Family (乡村/山水)	哟, 哥哥, 嗨, 里, 妹妹, 哥, 走, 转, 白, 长, 嘞, 溜, 红, 山, 开, 水, 见, 笑, 送
16. Blossoming/Joy (爱情/幸福)	想, 一起, 喜欢, 陪, 爱, 知道, 想要, 笑, 世界, 微笑, 拥抱, 感觉, 慢慢, 眼睛, 听, 心, 心里, 我要, 带, 幸福
17. Nationalism/China (爱国)	中国, 恭喜, 新, 菩萨, 溜, 南无, 祖国, 祝, 北京, 人民, 英雄, 来来来, 新年, 吼, 平安, 东方, 历史, 阿弥陀佛, 祝福, 菩提
18. Time/Nihilism (时间/转瞬即逝)	一天, 时间, 再见, 身边, 永远, 脸, 世界, 改变, 思念, 从前, 想念, 明天, 远, 出现, 看见, 回忆, 昨天, 一点, 一遍, 一年
19. Being/Existential (存在/生命的意义)	世界, 无法, 灵魂, 现实, 需要, 不断, 成为, 黑暗, 继续, 命运, 生命, 身体, 内心, 像是, 保持, 自我, 有人, 每个, 孤独, 自由

Table 3: Manually labeled lyrical topics and their top tokens, as captured from a 20-topic LDA model trained on preprocessed song lyrics.

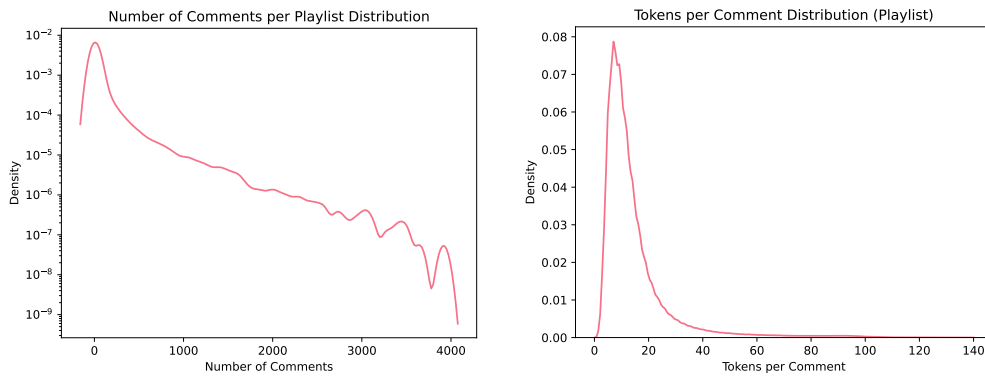


Figure 10: Comment (left) and comment token (right) distributions across all playlists with at least one comment.

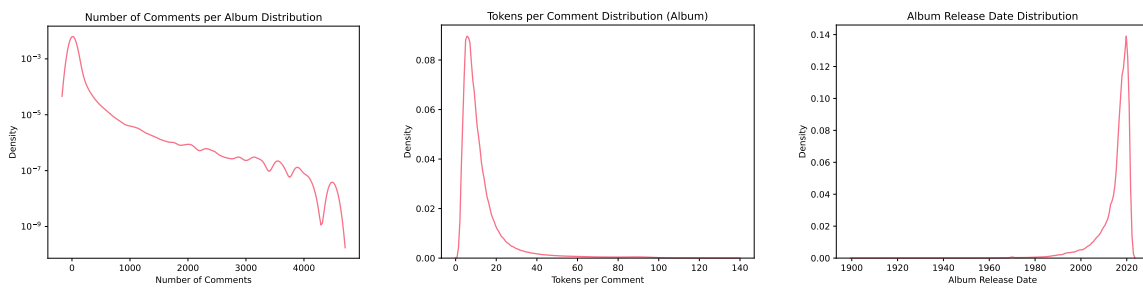


Figure 11: Comment (left) and comment token (middle) distributions across all albums with at least one comment, as well as album release date distributions (right).

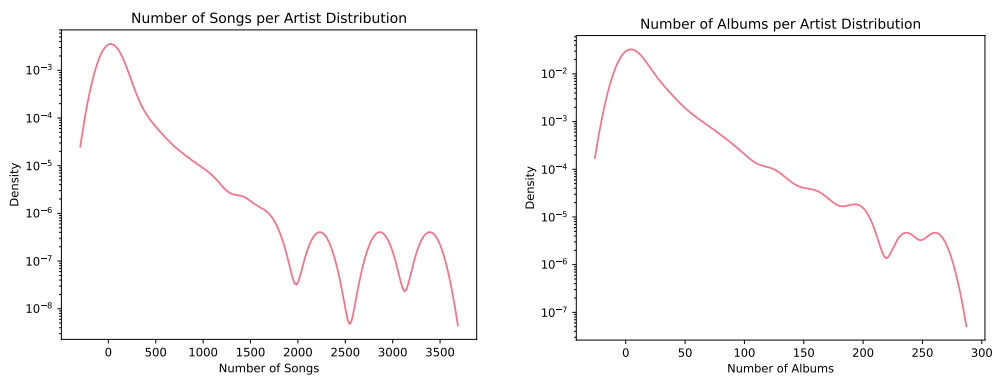


Figure 12: Song (left) and album (right) distributions per artist across all artists. Platform-listed artists with the highest amount of songs and albums are generic compilations of multiple artists, e.g. “华语群星” (“Chinese stars”).



Figure 13: Average marginal effects of **musical features** on listener affective responses, controlling for lyrical features and listener demographics. Standard errors are shown; **valence** in red, **arousal** in blue.

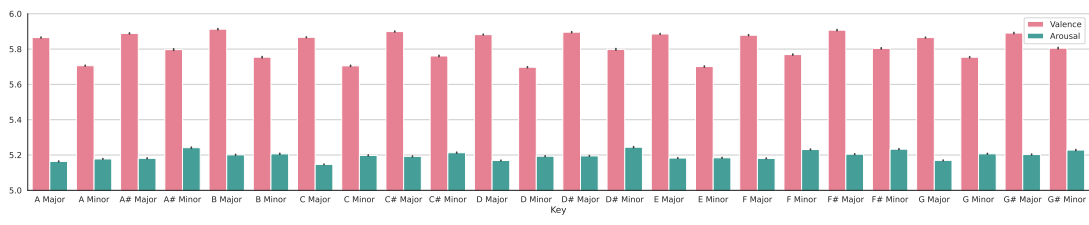


Figure 14: Raw valence and arousal scores for variations in listener affective responses with respect to **key**. Across all keys, valence response to major mode keys is consistently higher than that of their corresponding minor mode key, while the opposite relationship exists for arousal response. Standard errors are shown; **valence** in red, **arousal** in blue.





Figure 15: Average marginal effects of LIRC psycholinguistic lexical category **lyrical features** on listener affective responses, controlling for musical features and listener demographics. With the intent to reduce noise at the extremities, x-axis limits are capped at their 95% quantile values. Arranged in alphabetical order, standard errors are shown; **valence** in red, **arousal** in blue (Part 1/4).



Figure 16: Average marginal effects of LIRC psycholinguistic lexical category **lyrical features** on listener affective responses, controlling for musical features and listener demographics. With the intent to reduce noise at the extremities, x-axis limits are capped at their 95% quantile values. Arranged in alphabetical order, standard errors are shown; valence in red, arousal in blue (Part 2/4).



Figure 17: Average marginal effects of LIWC psycholinguistic lexical category **lyrical features** on listener affective responses, controlling for musical features and listener demographics. With the intent to reduce noise at the extremities, x-axis limits are capped at their 95% quantile values. Arranged in alphabetical order, standard errors are shown; **valence** in red, **arousal** in blue (Part 3/4).



Figure 18: Average marginal effects of LIWC psycholinguistic lexical category **lyrical features** on listener affective responses, controlling for musical features and listener demographics. With the intent to reduce noise at the extremities, x-axis limits are capped at their 95% quantile values. Arranged in alphabetical order, standard errors are shown; **valence** in red, **arousal** in blue (Part 4/4).

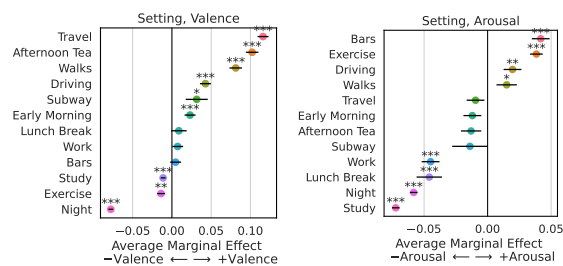


Figure 19: Average marginal effects of listening contexts in **setting-tagged** playlists on listener affective responses, controlling for songs and user demographic variables; standard errors are shown.



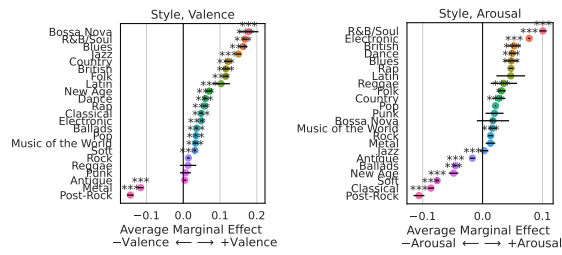


Figure 20: Average marginal effects of listening contexts in **style-tagged** playlists on listener affective responses, controlling for songs and user demographic variables; standard errors are shown.

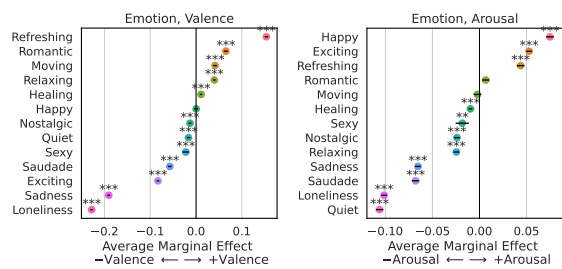


Figure 21: Average marginal effects of listening contexts in **emotion-tagged** playlists on listener affective responses, controlling for songs and user demographic variables; standard errors are shown.

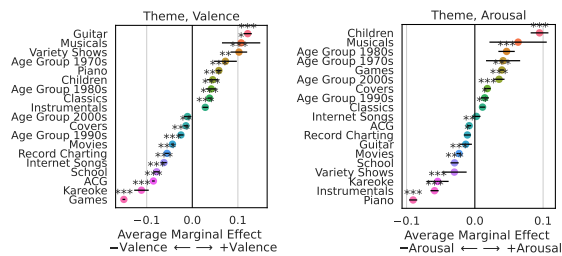


Figure 22: Average marginal effects of listening contexts in **theme-tagged** playlists on listener affective responses, controlling for songs and user demographic variables; standard errors are shown.

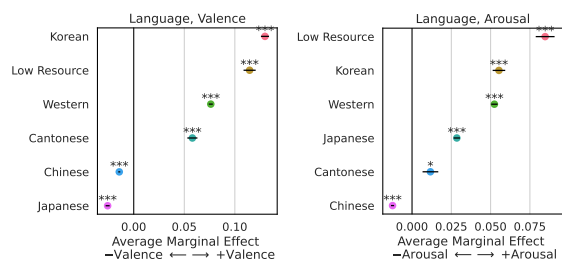


Figure 23: Average marginal effects of listening contexts in **language-tagged** playlists on listener affective responses, controlling for songs and user demographic variables; standard errors are shown.

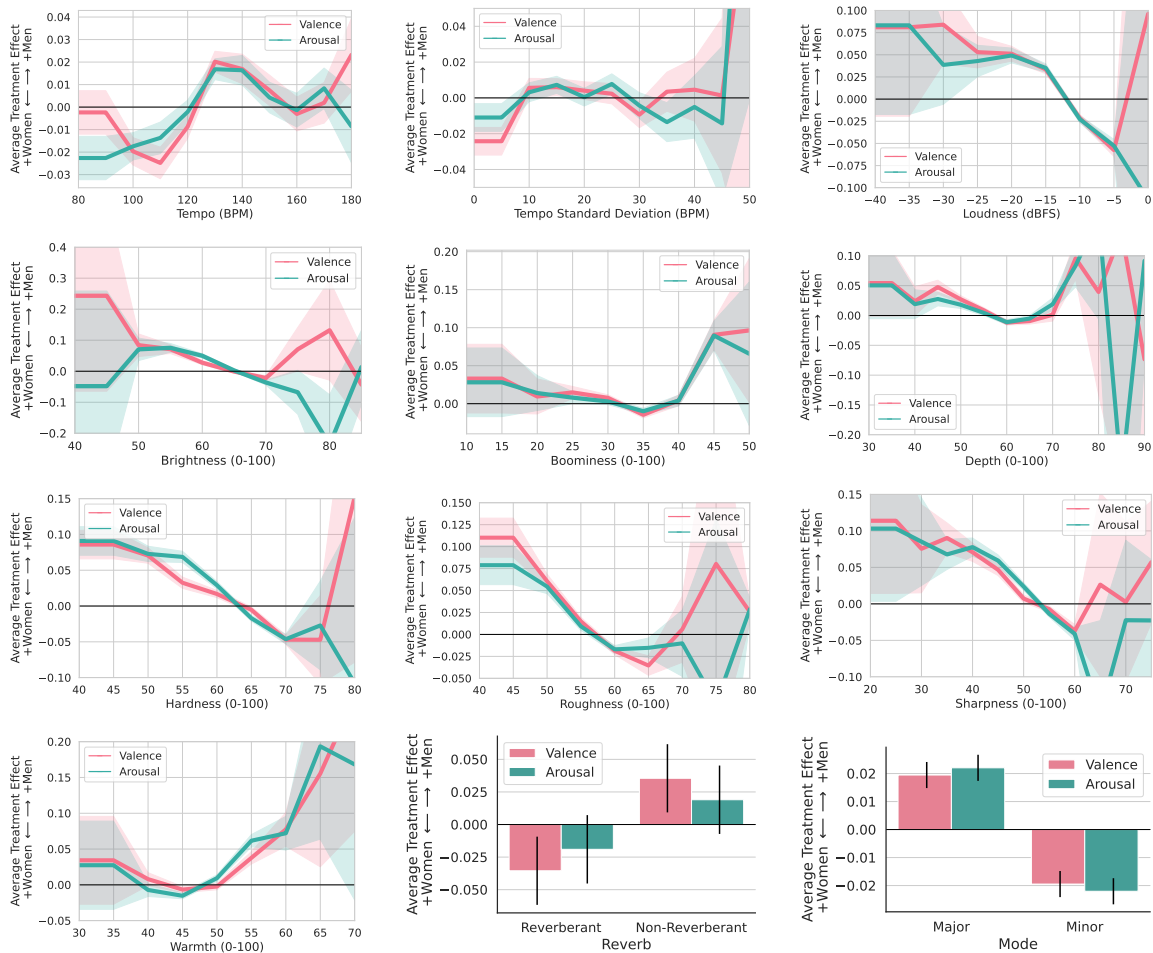


Figure 24: Average treatment effects of listener **gender** on response valence and arousal relative to **musical** features. A *positive* ATE here indicates a larger percent increase in valence or arousal for *men*, and a *negative* ATE here indicates a larger percent increase in valence or arousal for *women*. Standard errors are shown; **valence** in red, **arousal** in blue.

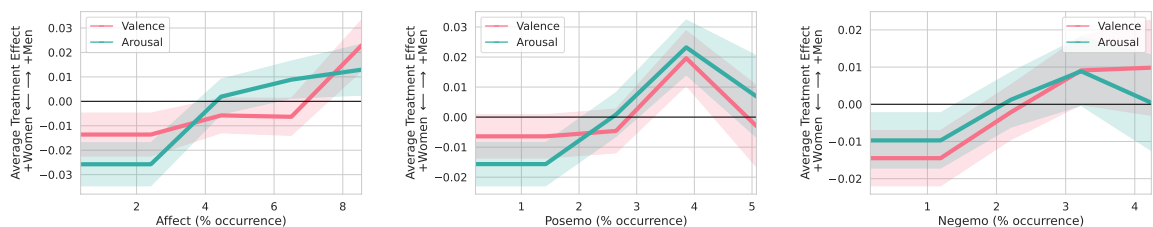


Figure 25: Average treatment effects of listener **gender** on response valence and arousal relative to **lyrical** features on LIWC affective processes. Observations show that men are more positively affected by greater *posemo* use, while women are more negatively affected by greater *negemo* use. With the intent to reduce noise at the extremities, x-axis limits are capped at their 95% quantile values. Standard errors are shown; **valence** in red, **arousal** in blue.

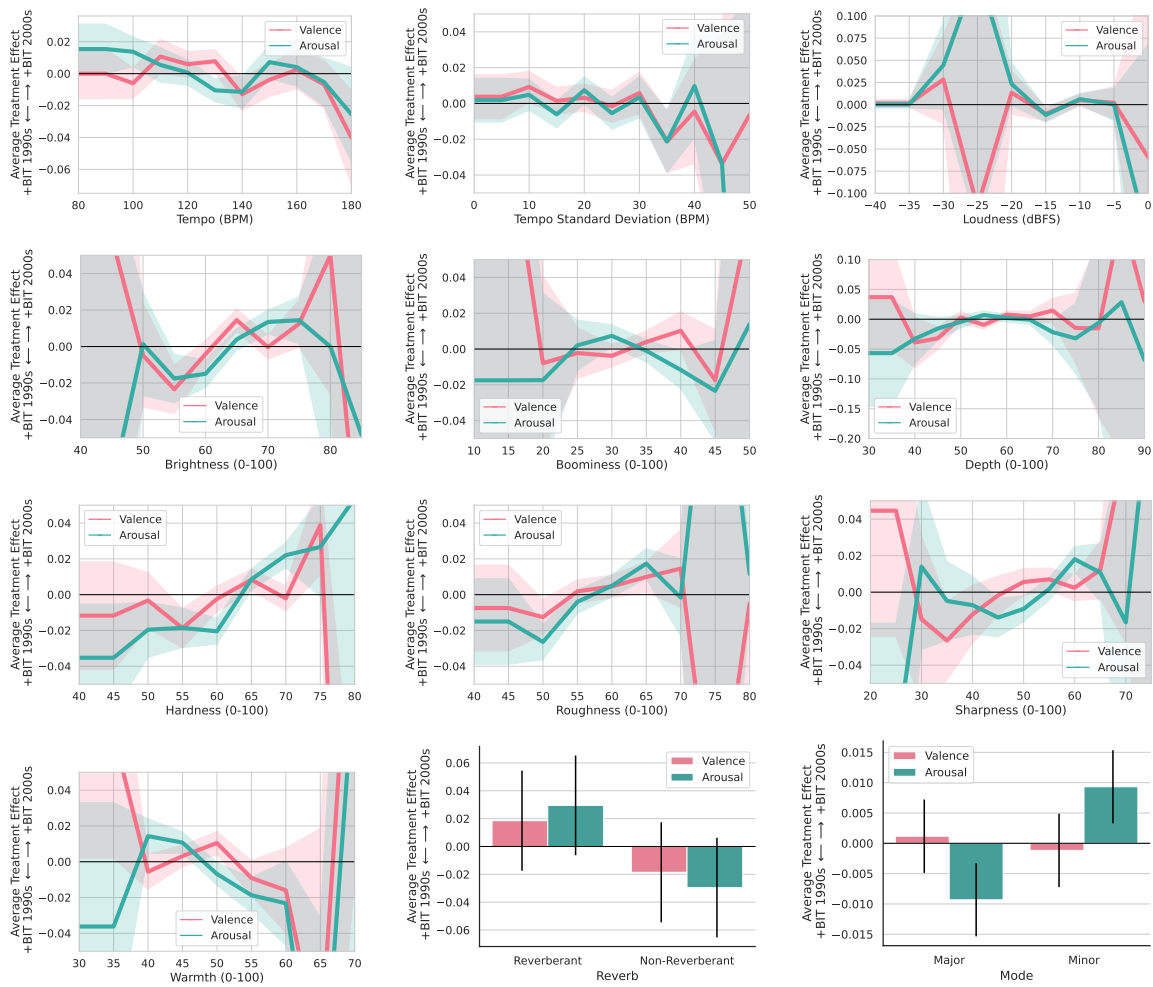


Figure 26: Average treatment effects of listener **age** on response valence and arousal relative to **musical** features. A *positive* ATE here indicates a larger percent increase in valence or arousal for those born in the (b.i.t.) *2000s*, and a *negative* ATE here indicates a larger percent increase in valence or arousal for those b.i.t. *1990s*. Standard errors are shown; **valence** in red, **arousal** in blue.