Cognition 222 (2022) 105010

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Decoding peak emotional responses to music from computational acoustic and lyrical features

Kazuma Mori [a,b,c,*]

[a] Montreal Neurological Institute, McGill University, Canada
[b] Graduate School of Human Sciences, Osaka University, Japan
[c] Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology (NICT), Japan

A R T I C L E I N F O

Keywords:
Peak emotion
Audio
Lyric
Music information retrieval
Natural language processing
Machine learning

A B S T R A C T

Music can evoke strong emotions. Research has suggested that pleasurable chills (shivering) and tears (weeping) are peak emotional responses to music. The present study examines whether computational acoustic and lyrical features can decode chills and tears. The experiment comprises 186 pieces of self-selected music to evoke emotional responses from 54 Japanese participants. Machine learning analysis with L2-norm-regularization regression revealed the decoding accuracy and specified well-defined features. In Study 1, time-series acoustic features significantly decoded emotional chills, tears, and the absence of chills or tears by using information within a few seconds before and after the onset of the three responses. The classification results showed three significant periods, indicating that complex anticipation-resolution mechanisms lead to chills and tears. Evoking chills was particularly associated with rhythm uncertainty, while evoking tears was related to harmony. Violating rhythm expectancy may have been a trigger for chills, while the harmonious overlapping of acoustic spectra may have played a role in evoking tears. In Study 2, acoustic and lyrical features from the entire piece decoded tears but not chill frequency. Mixed emotions stemming from happiness were associated with major chords, while lyric content related to sad farewells can contribute to the prediction of emotional tears, indicating that distinctive emotions in music may evoke a tear response. When considered in tandem with theoretical studies, the violation of rhythm may biologically boost both the pleasure- and fight-related physiological response of chills, whereas tears may be evolutionarily embedded in the social bonding effect of musical harmony and play a unique role in emotional regulation.

1. Introduction

Music has existed in every known society, past and present (Mehr, Singh, Knox, et al., 2019). On average, people listen to music for more than 15% of their daily lives (Rentfrow, 2012). Its ubiquity demonstrates the importance of music in society. One of the main purposes of listening to music is the emotional experience; music is a universal pleasure for human beings (Dubé & le Bel, 2003). People can even experience intense pleasure while listening to music, which evokes physiological changes and activates the neural reward circuit (Blood & Zatorre, 2001; Salimpoor, Benovoy, Larcher, et al., 2011). Although previous peak emotional experience studies have confirmed listeners' psychophysiological responses when peak emotion occurs, the question of why music evokes such a rewarding experience remains to be answered (Goupil & Aucouturier, 2019).

Music is a sound sequence[1] that conveys acoustic features to the listener. Research on acoustic features should clarify why music evokes peak emotional experiences. Although past studies have repeatedly shown that one person's favorite music does not necessarily evoke peak emotion in another (Benedek & Kaernbach, 2011; Blood & Zatorre, 2001; Mori & Iwanaga, 2017; Mori & Iwanaga, 2021; Salimpoor et al., 2011), it seems that there are common acoustic features associated with peak emotional responses (Beier, Janata, Hulbert, & Ferreira, 2020; Grewe, Nagel, Kopiez, & Altenmüller, 2007; Nagel, Kopiez, Grewe, & Altenmüller, 2008; Sloboda, 1991). Peak emotional responses do not always occur due to certain acoustic features, because an emotional

* Corresponding author at: Montreal Neurological Institute, McGill University, Montreal, Quebec H3A 2B4, Canada
E-mail address: kazuma.mori@mcgill.ca.

[1] Although the current study primarily focuses on acoustic and lyrical features, music has multimodality. The visual information of the performer is a key factor in emotional responses. For example, the smells and atmosphere in concert halls or outdoor festivals may also play an effective role in eliciting emotional responses.

https://doi.org/10.1016/j.cognition.2021.105010
Received 30 May 2021; Received in revised form 16 December 2021; Accepted 26 December 2021
Available online 5 January 2022

stimulus may not always elicit the same emotional response in different individuals or even in the same individual (Sander & Nummenmaa, 2021). However, acoustic features can constitute the necessary conditions (minimum requirements) for evoking peak emotions.

Since several studies have examined the acoustic features of music that evoke emotional experiences (Gabrielsson & Lindström, 2010), researchers have investigated the links between musical features and emotional chills. An emotional chill is a well-known peak emotional response to music and is defined as "goose bumps" or "shivers down the spine." (Goldstein, 1980) An early study asked participants to recall emotional chill experiences and the respective pieces of music that evoked them, and to identify the moment of chills in the piece as accurately as possible (Sloboda, 1991). The results suggest that emotional chills are evoked by new or unanticipated harmonies and sudden changes in dynamics and texture. Later research analyzed the acoustic features in digital music files around chill onsets (Beier et al., 2020; Grewe et al., 2007; Nagel et al., 2008). Grewe et al. (2007) found that an increase in loudness and sharpness was synchronized with chill responses. Nagel et al. (2008) indicated that chills were linked to an increase in sound roughness and a decreased tone-to-noise ratio. Furthermore, a recent cross-cultural study showed that sudden peaks in loudness, brightness, and roughness were correlated with chill occurrence (Beier et al., 2020). Therefore, research on chill responses and the corresponding acoustic features has indicated that the response may often be elicited by dynamic changes.

There are, however, other types of peak emotional responses, such as emotional tears, which are defined as "weeping" or a "lump in the throat." Sloboda (1991) suggested that emotional tears occur when melodic appoggiaturas and harmonic sequences emerge in music; thus, according to these results, the acoustic features of chills are distinct from those of tears. However, to the best of my knowledge, no study has analyzed the link between acoustic features in a digital music file and emotional tears. An elaborate acoustic analysis may reveal the difference between chill- and tear-evoking acoustic features. To improve our understanding of why peak emotional responses occur through music, it is important to identify the specific acoustic features of music that evoke chills and tears. Although trait and state differences are crucial factors for experiencing peak emotions in music (Ferreri, Mas-Herrero, Zatorre, et al., 2019; Martínez-Molina, Mas-Herrero, Rodríguez-Fornells, et al., 2016; Mas-Herrero, Dagher, Farrés-Franch, & Zatorre, 2021; Mas-Herrero, Zatorre, Rodriguez-Fornells, & Marco-Pallarés, 2014), acoustic features can trigger peak emotions (Beier et al., 2020; Grewe et al., 2007; Nagel et al., 2008; Sloboda, 1991). Furthermore, several studies have extracted chills and tears as different psychological factors, although they are positively correlated (Menninghaus, Wagner, Hanich, et al., 2015; Silvia & Nusbaum, 2011; Zickfeld, Schubert, Seibt, et al., 2018). Tears evoke psychophysiological calming, which is a different response from the psychophysiological arousal of chills (Mori & Iwanaga, 2017; Mori & Iwanaga, 2021). When chills and tears co-occur, their physiological responses are mixed and cancel each other out (Mori & Iwanaga, 2021); therefore, capturing the characteristics of chills and tears when they are treated separately, rather than capturing the similarities between them, would yield useful results. The specific acoustic features for each chill or tear response should help explain the psychophysiological responses.

A multivariate decoding approach has the advantage of finding specific acoustic features that evoke chills and tears. Previous studies have performed univariate analysis (e.g., correlation analysis) to examine the associations between acoustic features and peak emotions (Bannister & Eerola, 2018; Beier et al., 2020; Grewe et al., 2007; Nagel et al., 2008). However, because many indices of acoustic features show a mid-to-strong correlation (e.g., loudness is strongly correlated with roughness (Lange & Frieler, 2018)), the results of univariate analysis between peak emotion and acoustic features may reflect spurious correlations. To resolve this problem and more precisely detect effective acoustic features, multivariate analysis, such as multiple regression, should be useful. In particular, L2-norm-regularized regression (i.e., ridge regression) could provide more robustness to collinearity for regression coefficients through the application of a penalty to coefficients (Machens, Wehr, & Zador, 2004). The recent music information retrieval (MIR) approach can compute several dozen acoustic features (Eerola, 2012; Lartillot, Lartillot, Toiviainen, & Toiviainen, 2007); therefore, the combination of regularized regression analysis and MIR feature extraction enables the identification of important acoustic features associated with chills and tears. In addition, the machine learning technique allows us to better specify the acoustic features that contribute to evoking chills and tears. As the cross-validation procedure and out-of-sample prediction can avoid overfitting problems (Yarkoni & Westfall, 2017), the machine learning approach is able to better evaluate the valid acoustic distinctions between music-evoked chills and tears than simple regression analysis.

Furthermore, lyric content may evoke peak emotional responses. Since songs have universality and diversity (Mehr et al., 2019), music often includes not only acoustic features, but also lyric content, which can influence emotional responses to music (Mori & Iwanaga, 2014a). In addition to the cultural universality of music and language (Brwon, 1991; Savage, Loui, Tarr, et al., 2020), cross-cultural studies of music have proven that vocal songs are a general feature of global music collections (Brown & Jordania, 2013; Savage, Brown, Sakai, & Currie, 2015). As previous studies have reported that films and novels can be powerful elicitors of chill and tear responses (Konečni, Wanic, & Brown, 2007; Schubert, Zickfeld, Seibt, & Fiske, 2018; Vingerhoets & Bylsma, 2016), we know that stories can evoke peak emotional responses. Taken together, the stories in song lyrics can effectively evoke peak emotional responses. To quantitatively investigate lyric content, the natural language processing (NLP) approach would be fruitful. NLP summarizes text information, computes word frequency, and identifies clusters of words with similar meanings. The potential contribution of NLP to lyric content analysis has recently been explored (Greenberg, Matz, Schwartz, & Fricke, 2020). Machine learning analysis with ridge regression may reveal the types of lyric content associated with chills and tears.

In the present research, I conducted two studies to investigate the relationship between acoustic and lyrical features and the two types of peak emotional responses. The objective was to determine whether acoustic and lyrical features decode which peak emotions occur and what features are important in evoking chills and tears. In Study 1, I examined the association between the multivariate time-series acoustic features from the MIR computational approach and emotional chills and tears. In Study 2, the relationship between both acoustic and lyrical features and peak emotional responses was examined using both the MIR and NLP approach derived from the entire piece. Using regularization regression and cross-validated machine learning techniques, the acoustic and lyrical features associated with peak emotions were tested.

2. Study 1: Decoding by time-series acoustic features

In Study 1, the relationship between time-series acoustic features and the onset of peak emotional responses was examined using MIR methodology. This approach can capture momentary acoustic changes that evoke peak emotions. Unlike Study 2, the lyric content was not investigated because it is difficult to conduct second-by-second quantitative assessments compared with acoustic features.

2.1. Methods

2.1.1. Participants

In the sampling phase, I recruited 158 participants from regularly scheduled university classes and asked them to complete a questionnaire to assess the frequency at which they experienced chills and tears while listening to music. As a result, 54 undergraduates (20 men and 34 women, mean age = 19.13) who often experienced chills and tears participated in the experiment. The Faculty Committee granted ethical approval and the participants provided written informed consent. Participants were compensated for their participation in the study. The experiment was the same as that reported by Mori and Iwanaga (2021).

2.1.2. Stimuli and procedure

The experiments were conducted in a sound-attenuated room. The stimuli were delivered through an amplifier and speakers. Before the experiment, participants were instructed to select three to six favorite pieces of chill- and tear-evoking music. A total of 222 pieces of music were collected. This procedure was performed because self-selected music is a useful stimulus to elicit emotional peaks in a laboratory setting (Mori & Iwanaga, 2014b; Salimpoor et al., 2011; Sumpf, Jentschke, & Koelsch, 2015). To ensure that the music remains uniform across participants and to analyze both acoustic and lyrical features, I instructed participants to select music stimuli from pop/rock songs, including lyrics. After all the song information was collected prior to the experiment, the experimenter prepared the digitized files (16 bit .wav files sampled at 44.1 kHz) from the original CD recordings. The mean duration of the pieces of music was 290.3 s.

The speakers were positioned in front of the participants and connected to a computer outside the booth. Once in the sound-attenuated room, the participant was seated in front of a computer screen. First, participants engaged in a training session to familiarize themselves with the experimental procedure. In the session, the musical stimuli were adjusted to each participant's comfortable listening level to promote the evoking of strong emotional responses. In the main trials, participants' self-selected music was presented. Participants were instructed to complete button presses while listening to the music. This involved pressing the left button of a computer mouse whenever they felt chills (referred to as “goosebumps” or “shivers down the spine”) and the right button of the mouse whenever they felt tears (instructed as “weeping” or “a lump in the throat”). “Goosebumps” and “shivers down the spine” as well as “weeping” and “a lump in the throat” have a strong positive correlation (Bannister, 2019; Mori & Iwanaga, 2015; Sachs, Ellis, Schlaug, & Loui, 2016). Mouse button signals were recorded at a frequency of 10 Hz. After each song was played, the participants relaxed for one minute.

2.1.3. Peak emotion target

The onset time was recorded for each chill and tear response. The extracted chills and tears were used as targets for machine learning analysis. The acoustic features were analyzed for the onset of emotional peaks, with the period from 10 s before to 10 s after the button press. Research has indicated that chill-related acoustic features emerge during this period (Grewe et al., 2007; Koumura, Nakatani, Liao, & Kondo, 2020; Nagel et al., 2008). For each left and right mouse button press (chill and tear responses, respectively), button presses that occurred less than 10 s after the previous presses were not analyzed to avoid the duplication of acoustic features. Since participants often reported chills and tears simultaneously (Mori & Iwanaga, 2021; Wassiliwizky, Jacobsen, Heinrich, et al., 2017), chills and tears were separated to extract mutually exclusive acoustic features. Specifically, chill and tear responses were extracted only when a chill (tear) was reported; there was no occurrence of a tear (chill) within 10 s before or after the onset of a chill (tear). Furthermore, since the present examination requires acoustic features around the peak onset, chills and tears that occurred within the first or last 10 s of the piece were excluded from analysis.

Neutral onsets were also extracted as a control condition to investigate the acoustic specificity of chills and tears. Neutral onsets were extracted using random sampling from the period that did not include the onset of chills and tears. Random sampling for neutral onsets was performed after excluding the 10 s before and after the onset of chills and tears. The number of neutral onsets was set to half the total number of chills and tears within the same pieces of music. Random sampling was repeated ten times to correct stable time-series acoustic features, and the average score was used for machine learning analysis.

2.1.4. Music selection

Analysis objects were selected from all 222 pieces of music according to the following criteria. First, six pieces of music were discarded from the analysis because the participants did not report any chill or tear responses. Next, any trial containing a number of emotional peaks exceeding three standard deviations from the mean was treated as an outlier. I calculated the outlier index for each of the numbers of chills, tears, and the sum of both responses; as a result, five pieces of music were removed from analysis. Finally, several pieces of music were selected by multiple participants. If the analysis is conducted using the same music, the results would show an unreasonably high accuracy rate; 25 pieces of music were discarded to avoid the repetitive use of acoustic features. When several participants selected the same piece of music, one participant who reported more chills and tears in the experiment was selected to collect as many peak emotional responses as possible. Consequently, 186 J-Pop/J-Rock pieces of self-selected music from 54 Japanese undergraduate participants that contained at least one chill or tear response were analyzed (Supplemental Table 1).

2.1.5. Acoustic feature extraction

Sound analysis was conducted using the MIR toolbox (Lartillot et al., 2007) in MATLAB R2018b (MathWorks, USA). A total of 32 acoustic features were computed for each excerpt (Table 1). These features, which are described in the toolbox manual and Eerola (2012), can be grouped into six domains: dynamics, timbre, harmony, rhythm, articulation, and structure. Several studies have introduced the perceptual interpretations of acoustic features (Alluri, Toiviainen, Jääskeläinen, et al., 2012; Eerola, 2012; Lange & Frieler, 2018). This interpretation is briefly described in Table 1. The dynamic features include root-mean-square (RMS) and low energy. Timbre features include spectral centroid, spectral spread, spectral flux, spectral irregularity, roughness, brightness, and 13 mel-frequency cepstral coefficients (MFCCs). Harmony features consist of mode, key clarity, and harmonic change detection function (HCDF). Rhythm features include event density, pulse clarity, tempo, fluctuation centroid, and fluctuation entropy. Articulation features include the attack time and attack slope, while structural features include novelty. Although several acoustic features

Table 1
Acoustic features extracted by the MIR toolbox.

Musical domain	Acoustic feature	Perceptual meaning	Time window
Dynamics	RMS	Loudness	50 ms (50%)
	Low energy	Loudness contrast	2 s
Timbre	Spectral centroid	Sharpness	50 ms (50%)
	Spectral spread	Fullness	50 ms (50%)
	Spectral flux	Percussive onsets	50 ms (50%)
	Spectral irregularity	Polyphony	50 ms (50%)
	Roughness	Dissonance	50 ms (50%)
	Brightness	Brightness	50 ms (50%)
	MFCC 1–13	No direct interpretation	50 ms (50%)
Harmony	Mode	Major/minor (happy/sad)	3 s (33%)
	Key clarity	Chord certainty	3 s (33%)
	HCDF	Chord change	3 s (33%)
Rhythm	Event density	Beat numerosity	3 s (33%)
	Pulse clarity	Beat strength	3 s (33%)
	Tempo	Tempo	3 s (33%)
	Fluctuation centroid	Rhythmicity	3 s (33%)
	Fluctuation entropy	Rhythmic complexity	3 s (33%)
Articulation	Attack time	Staccato or legato	2 s
	Attack slope	Staccato or legato	2 s
Structure	Novelty	Musical contrast	100 ms

Note. The numbers in parentheses indicate the overlap rates. Only novelty was computed for the entire piece of music. RMS, root-mean-square; MFCC, mel-frequency cepstral coefficients; HCDF, harmonic change detection function.

showed a high correlation coefficient (see Supplemental Fig. 1), each feature could have different perceptual meanings. A brief description of these parameters follows (more details can be found in the MIR toolbox manual).

The RMS energy corresponds to the root average of the square of the amplitude of the waveform. Low energy is defined as the percentage of analysis frames that have less RMS energy than the average RMS energy (Tzanetakis & Cook, 2002).

Of the timbre features, the spectral centroid represents the weighted mean of the amplitude of the frequencies present in the audio signal, while the spectral spread indicates the weighted standard deviation of these frequencies. The spectral flux corresponds to a time-varying descriptor, calculated as the Euclidean distance between the spectra of successive audio frames normalized for amplitude (Scheirer & Slaney, 1997). Spectral irregularity is defined as the degree of variation in successive peaks of the spectrum. Roughness is a measure of the amount of sensory dissonance in the signal, which is computed by considering all possible pairs of peaks in the spectrum (Sethares, 2005). Brightness, which measures the amount of energy above the threshold frequency of 1500 Hz, reflects the proportion of high frequencies in the spectrum (Juslin, 2000). Furthermore, I used the first 13 MFCCs (excluding the 0th orthogonal component that indicates the average energy of the signal). For each frame, MFCCs provide a low-dimensional parameterization of the overall shape of the signal's Mel-spectrum; that is, a spectral representation that considers the human near-logarithmic perception of sound in magnitude (log-magnitude) and frequency (Mel scale).

For the harmonic features, mode is estimated as the difference between the best scores for a major and minor key obtained by correlation with the Krumhansl-Kessler key profiles (Krumhansl, 2004). Key clarity represents the score corresponding to the best-fitting key from 12 chromas (C, C#/Db, D, D#/Eb, E, F, F#/Gb, G, G#/Ab, A, A#/Bb, B) for major and minor keys. The HCDF is an important tonal space representation that distorts a 12-chroma vector and accumulates the energy of each of the chromatic notes. Using the distance between tonal spaces, the HCDF detects harmonic boundaries in the chord of successive audio frames (Cohn, 1998; Harte, Sandler, & Gasser, 2006).

Regarding rhythmic features, tempo is computed from the spectral decomposition of the onset detection curve (Ellis, 2007). Event density is an estimation of the average number of note onsets (detected from peaks in the amplitude envelope) per second, whereas pulse clarity is the degree of a clear and stable beat in music and is defined as the maximum correlation value in the autocorrelation function computed from the amplitude envelope (Eerola, 2012). The fluctuation centroid is the geometric mean of the fluctuation spectrum, representing the global repartition of rhythm periodicities within the range of 0–10 Hz (Pampalk, Rauber, & Merkl, 2002). This feature indicates the average frequency of these periodicities. The fluctuation entropy corresponds to the relative entropy (Shannon's entropy (Shannon, 1948)) of the fluctuation spectrum, representing the global repartition of rhythm periodicities. High fluctuation entropy indicates several co-existing rhythms of different periodicities (Alluri et al., 2012).

The attack time is the time interval during which the amplitude envelope increases to its peak intensity (Bello, Daudet, Abdallah, et al., 2005). The attack slope is defined as the average temporal slope of the energy during the attack segment. In addition, novelty represents the degree of temporal repetition of the spectrum or chromagram features across time based on the detection of edges within the diagonal of the self-similarity matrix (Foote & Cooper, 2003). The MIR toolbox computes novelty using both spectrum and chromagram features.

As shown in Table 1, short-term features, including dynamic RMS and all the timbre domain features, were evaluated on 50 ms frames, with a 50% overlap between frames (default settings). Long-term features, including harmony and rhythm features, were evaluated on 3 s frames with a 33% overlap (Alluri et al., 2012). This window length corresponds to typical estimates of the auditory sensory memory length (Fraisse, 1982). However, the dynamics of low-energy and articulation features were computed on 2 s frames. A structural feature (novelty) was computed in units of 100 ms for the entire piece of music because the overall music information is essential for the measurement.

To obtain peak emotion-related acoustic features, all the acoustic features before and after 10 s of each chill, tear, and neutral onset were extracted and synchronized with a 10 Hz button press measurement. All the short-term features were downsampled by retaining one sample for every 80 samples to match the sampling rate of 0.5 Hz of long-term and other features. The structural features were downsampled by retaining one out of every twenty samples. As a result, the length of each acoustic feature vector was ten samples corresponding to 20 s.

2.1.6. Machine learning analysis

For the machine learning analysis and statistical tests, the accuracy of decoding peak emotional responses was examined using time-series acoustic features. The examined time moved from 10 s before to 10 s after the peak onset. In correspondence with the sampling rate, the time was segmented every 2 s.

Machine learning analysis was used to determine whether the multivariate pattern of acoustic features around the peak onset could be classified as chill, tear, and neutral responses. The analysis was conducted using scikit-learn 1.0.1 in Python 3.8.12. After robust feature scaling (removing the median and scales with data quantiles), a linear ridge classifier was trained, wherein an L2 regularization parameter λ was used for shrinkage and tuned to control overfitting. After converting the target values into {chill (0), neutral, (1), tear (2)}, a one-vs-rest classification task was performed. The estimator is the regression of the response variable y onto X, where the estimation of the w weights is subject to an L2 penalty term λ in the objective function:

$$\widehat{w} = arg\ min_w \{ \parallel y - Xw \parallel^2 + \lambda \parallel w \parallel^2 \} \tag{1}$$

Projection of fitted $\widehat{w}$ weights used to generate predictions, $\widehat{y}$:

$$\widehat{y} = X\widehat{w} \tag{2}$$

Nested ten-fold cross-validation (10F-CV) was applied, with outer 10F-CV estimating the generalizability of the model and the inner 10F-CV to determine the optimal parameter λ for the ridge classifier. In the outer 10F-CV, the data were divided into ten participant subsets. Specifically, the acoustic features were sorted according to the peak responses, and participant subsets were randomly assigned while retaining nearly the same number for each subset. Participant subsets 1 to 9 were initially used as the training set, with participant subset 10 used as the testing set. Based on the optimal λ (see next paragraph), a model was trained using all subjects in the training set. The model was then used to predict the outcome of all subjects in the testing set. Similarly, participant subsets 2–10 were used as the training set and participant subset 1 was used as the testing set; the above procedure was repeated.

Within each loop of the outer 10F-CV, the inner 10F-CVs were applied to determine the optimal λ. The training set for each loop of the outer 10F-CV was further partitioned into ten subsets of participants, similar to the outer loop. Nine participant subsets were selected to train the model under a given λ in the range [0.001, 0.01, 0.1, 1, 10, 100, 1000], and the remaining participant subset was used to test the model. This procedure was repeated ten times to ensure that each participant subset was used once as the testing dataset, resulting in a total of ten inner 10F-CV loops. For each λ value, the accuracy between the actual and predicted outcomes was calculated for each inner 10F-CV loop and averaged across the ten inner loops. The mean accuracy was defined as the inner prediction accuracy, and λ with the highest inner prediction accuracy was selected as the optimal λ for the outer 10F-CV.

2.1.7. Statistical tests

Non-parametric statistical tests were used to evaluate the machine learning results every 2 s. To evaluate if the prediction performance was significantly better than expected by chance, I performed a permutation test. The prediction procedure was repeated 5000 times (Marozzi, 2004; Winkler, Ridgway, Douaud, et al., 2016). For repeated outer 10F-CV, the labels of chill, neutral, and tear (0,1,2) responses across the training data were scrambled 5000 times. The permutated classifier on each testing dataset was tested to produce a null distribution for the empirical *p*-value. The *p*-value was calculated to obtain the accuracy values in the null distributions that were equal to or higher than the true accuracy from the analyses. Bonferroni correction was applied to correct for multiple comparisons by multiplying each *p*-value by the total number of tests performed ($p < .05/10$).

Furthermore, I confirmed that regression coefficients are highly predictive of peak responses to understand the important acoustic features for classification. Significance tests on parameter estimates from the ridge classifier were performed via bootstrap resampling (Kohoutová, Heo, Cha, et al., 2020). This was repeated 5000 times, leading to a set of resampled acoustic features and peak emotional responses. The bootstrap distribution was then characterized by mean and 95% confidence intervals (CIs). A 95% CI that does not contain 0 indicates a significant effect on the classification of chills and tears.

2.2. Results and discussion

A total of 186 pieces of self-selected music from 54 Japanese undergraduate participants were analyzed. From 186 pieces of music, chill responses were collected 379 times ($M = 2.65$, $SD = 1.45$), tear responses were collected 381 times ($M = 2.63$, $SD = 1.40$), and neutral responses were collected 380 times ($M = 2.09$, $SD = 0.94$). Fig. 1 shows the mean time courses of 31 acoustic features. Machine learning analysis was performed using all 1140 responses.

Although 102 pieces of music evoked both chills and tears, these responses are time-independent; that is, the period of chills and tears has a time distance. When the co-occurrence of chills and tears was defined by reporting these responses within 2 s, the number of chills with tears amounted to only 67 (see preliminary results of the quadratic classification task, Supplemental Fig. 2). Therefore, chills and tears may be evoked primarily by different acoustic features.

2.2.1. Classification accuracy

To examine the efficiency of the machine learning algorithm in the current peak emotion classification task (chill vs. tear vs. neutral responses), its performance was compared in terms of the prediction accuracy score. Fig. 2-A shows the accuracy of the ridge classification. The permutation test showed that six accuracy scores in ten periods were significantly higher than chance. Therefore, the machine learning analysis successfully decoded peak emotional responses in specific time periods. Although the classification accuracy for time 2 s to 4 s was biased for chills, the accuracy for other times was well balanced (Fig. 2-B).

Specifically, the best accuracy score was 43.3% between −2 s to onset ($p = .002$, Bonferroni correction). Before onset, accuracy scores for the other two time points were statistically significant (−10 s to −8 s, 38.6%, $p = .012$, Bonferroni correction; −8 s to −6 s, 38.9%, $p = .006$, Bonferroni correction; −6 s to −4 s, 38.6%; $p = .008$, Bonferroni correction). Accuracy scores for the two time points after onset were also significant (2 s to 4 s, 39.3%, $p = .002$, Bonferroni correction; 4 s to 6 s, 38.0%, $p = .026$, Bonferroni correction). These results indicate that changes in acoustic features started before onset, were remarkable at onset, and continued after onset. The three periods were intermittent,

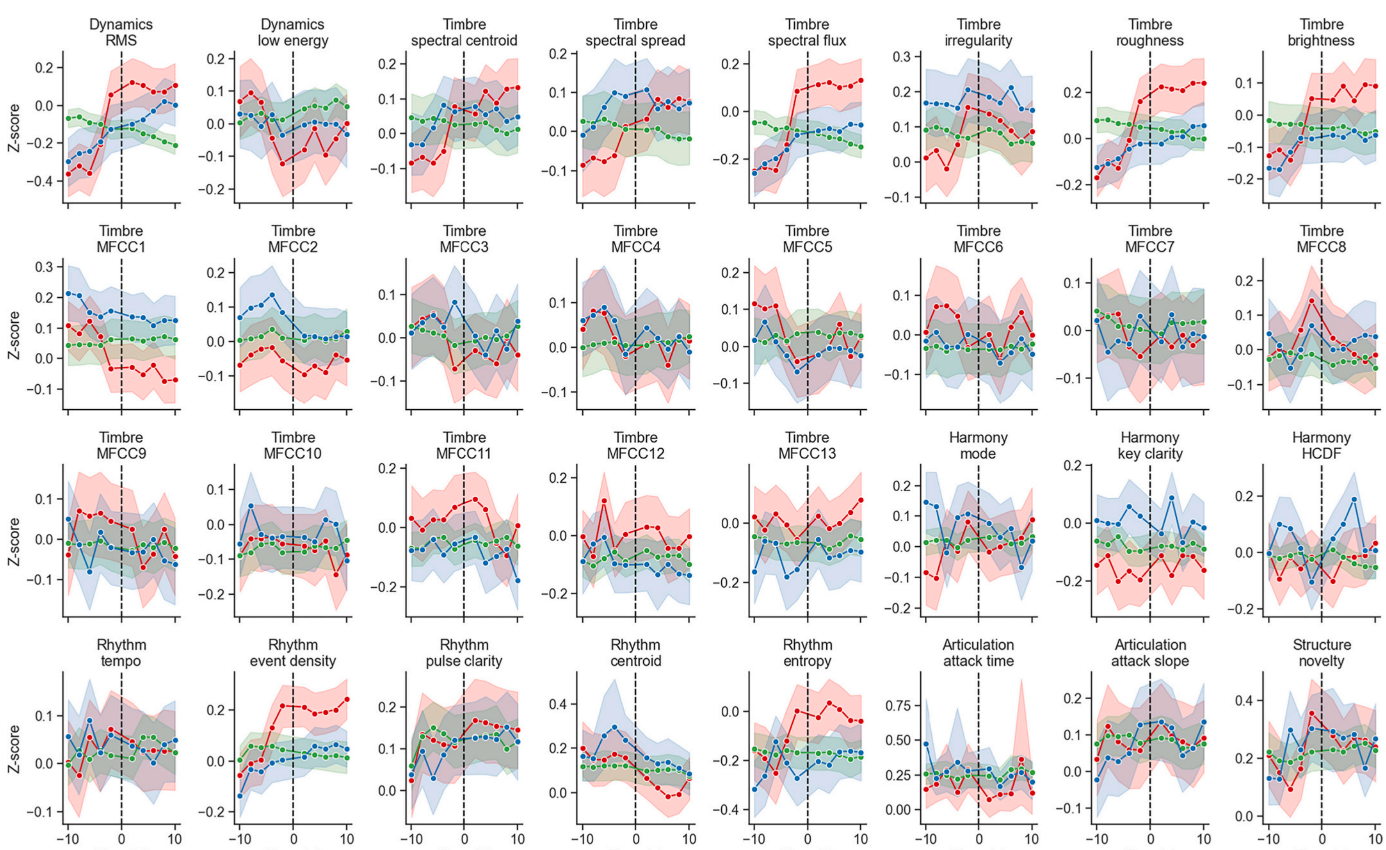

Fig. 1. Time course of 32 acoustic features averaged across all moments of chill (red), tear (blue), and neutral (green) responses. The shaded ribbons indicate the 95% CIs for the estimates. Time represents 20 s around the onset of each response. Each dot represents 2 s. RMS, root-mean-square; MFCC, mel-frequency cepstral coefficients; HCDF, harmonic change detection function. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

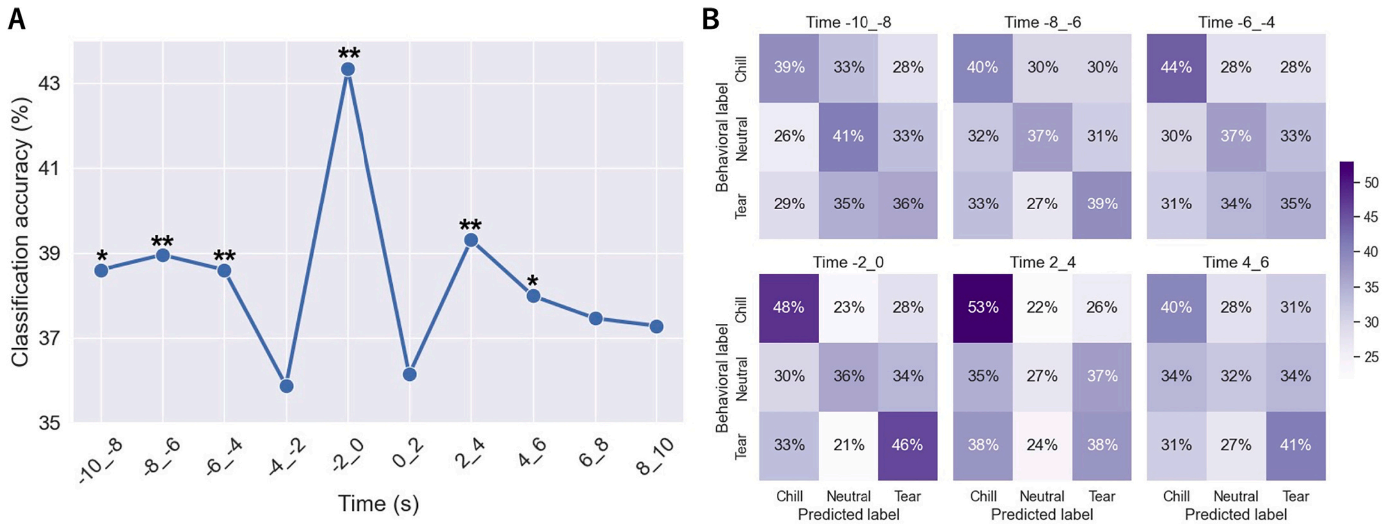

Fig. 2. Accuracy scores for acoustic decoding for peak emotion classification. (A) Classifier performance is plotted across time relative to peak onset. Time represents the 20 s around the onset of chill, tear, and neutral responses. *p < .05, ** p < .01, Bonferroni correction. (B) Confusion matrices to depict the performance of classifiers trained to identify peak emotions based on the acoustic features for each statistically significant time.

suggesting that the acoustic changes had multiple steps.

2.2.2. *Acoustic feature importance*

The next question is what acoustic features are important in classifying emotional chills and tears. First, the period of −10 s to −8 s was removed from consideration because the confusion matrix indicated that the acoustic features in the period best decoded the neutral response (Fig. 2-B) which may have made interpretation difficult (see the bootstrap results in the period, Supplemental Fig. 3). As shown in Fig. 3, when the accuracy was significantly higher than the chance level, the bootstrap test revealed that 20 acoustic feature weights significantly contributed to classifying peak emotions (p < .05, all p-values in Supplemental Table 2). If positive weight is significant, the higher acoustic features score may promote emotional chills and tears, whereas if negative weight is significant, the lower acoustic features score may promote emotional chills and tears (the higher acoustic features score may suppress emotional chills and tears). Since classification accuracy suggested multi-step acoustic changes, different combinations of acoustic features contributed to the classification of each of the three periods (Fig. 3-A [before onset], B [onset], C [after onset]). Note that the significant coefficient of neutral responses is difficult to interpret because the responses were collected using random sampling.

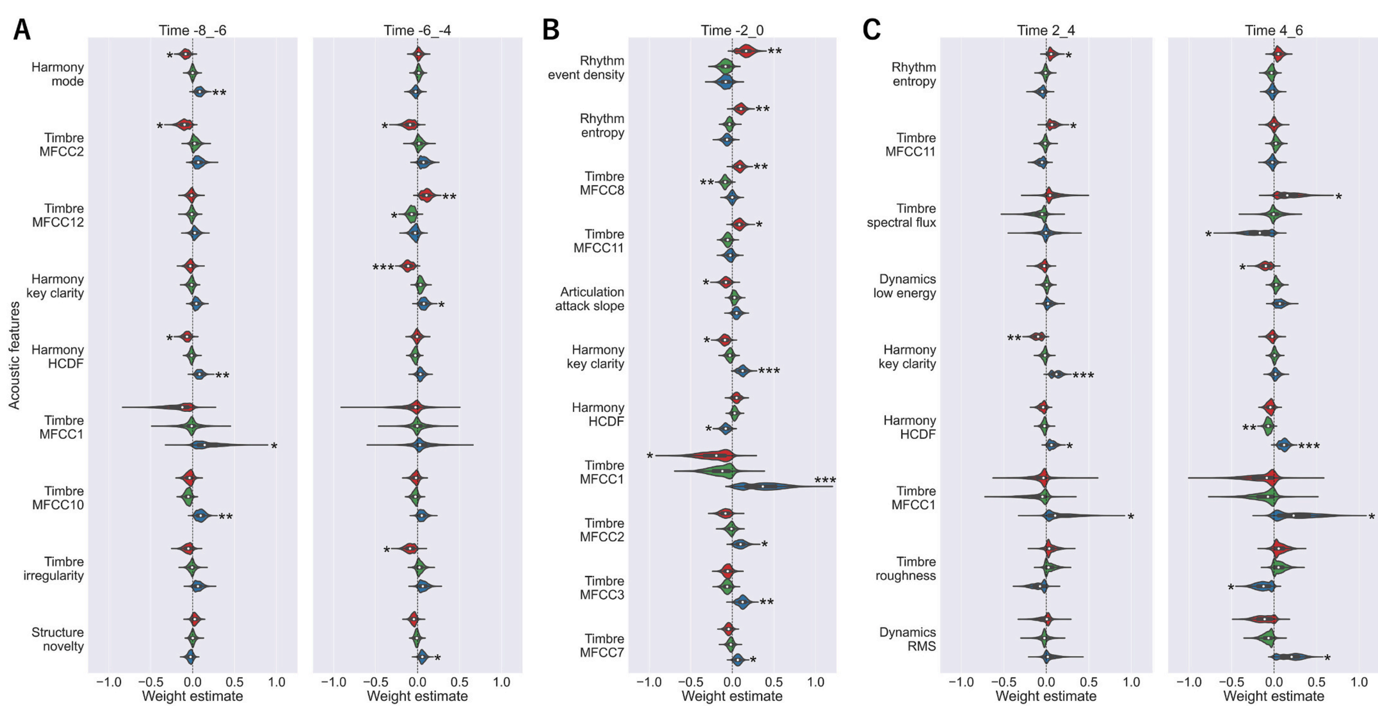

Fig. 3. Violin plot of the contributions of acoustic features to time-series classifications. Red, blue, and green indicate the contribution to chill, tear, and neutral responses of a given acoustic feature, respectively. Violin plots represent kernel probability density of bootstrap resampling. White dots represent the average. Black error bars indicate SD. *p < .05, **p < .01, ***p < .001. RMS, root-mean-square; MFCC, mel-frequency cepstral coefficients; HCDF, harmonic change detection function. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

For chill responses, minor mode (negative coefficient for mode) with lower MFCC2, followed by higher MFCC12 and lower irregularity were the effective features for the pre-onset period. Higher rhythm event density, entropy, MFCC8/11, and lower MFCC1 were effective features in the onset period alone. In this period, the effect of a lower attack slope may reflect many sound events. Higher spectral flux, rhythm entropy, and MFCC11 with lower energy were effective features for the after-onset period. These results suggest that a minor mode with a timbral component may lead to the anticipation of what will happen next. Then, many events and uncertain rhythm trigger a chill response. After a few seconds, percussive and high-energy sounds provide a further impression. MFCCs may be a timbre change associated with effective acoustic features for each period. Additionally, lower key clarity promoted chill responses for all periods; chroma may be unclear across the three periods.

For tear responses, higher key clarity and MFCC1 were consistently included in the effective acoustic features. Major mode (positive coefficient for mode), higher HCDF, and MFCC1/10, followed by higher novelty, were the most effective features for the pre-onset period. Higher MFCC1, 2, 3, and 7 were the effective features for the onset period alone. Further higher HCDF and RMS with lower spectral flux and roughness were effective features for the after-onset period. Since MFCC1 showed a strong negative correlation with brightness ($r = -0.86$) and spectral centroid ($r = -0.61$) (Supplemental Fig. 1), these results suggest that the tear sections always include clear chords and dark timbre (Eerola, 2012; Lange & Frieler, 2018). The contribution of HCDFs for two periods indicates that the chord may change before and after onset. After a major mode with a timbral component and novelty section emerged, another timbre represented by four MFCCs may trigger a tear response. Chord changes with loud but not percussive and dissonant sounds may lead to more excitement for the music that follows.

In summation, the results of Study 1 indicate that multivariate acoustic features can classify peak emotional responses to music. Furthermore, the examination of a time-series machine learning analysis indicates when and what acoustic features are effective for classification.

3. Study 2: Decoding by acoustic and lyrical features from the entire piece

In Study 2, the relationship between the acoustic and lyrical features of the whole piece and the number of peak emotional responses for a piece of music were examined using both MIR and NLP methodology. In contrast to Study 1, the approaches reflect listeners' impressions of the overall musical content for peak emotions.

3.1. Methods

3.1.1. Peak emotion target

Chill and tear responses were obtained from the same experiment as in Study 1. In Study 2, the number of chill and tear responses for 186 pieces of music was set as the target for machine learning analysis.

3.1.2. Acoustic feature extraction

The acoustic features were the same as in Study 1 (see Table 1 and Fig. 1), with the exception of novelty, because novelty is a measurement of time-series acoustic features. Using the MIR toolbox, 31 acoustic feature scores for each of the 186 songs were computed. These features represent summary information of the entire piece of music.

3.1.3. Lyrical feature extraction

I investigated both Japanese and English texts because both Japanese and English sentences are often included in Japanese song lyrics. However, all songs in the current experiment primarily included Japanese lyrics and few English lyrics (see the number of tokens). I applied the NLP technique to analyze the lyric texts. I used regular expressions to use Japanese and English words separately. A regular expression is a rule for describing strings with common properties that provide a method of searching for specific strings from a set of text documents. The Japanese texts were segmented into words and lemmatized using a Japanese morphological analyzer, MeCab (Kudo, Yamamoto, & Matsumoto, 2004), with MeCab-ipadic-NEologd used as a Japanese word dictionary (Satou, 2015). For English lyric extraction, the texts were tokenized and lemmatized using nltk in Python (Bird, Klein, & Loper, 2009). I removed stop words in Japanese using an open-source Ginza library (https://github.com/megagonlabs/ginza) and English using a customized nltk stop words library. First-person pronouns ('I', 'my', 'me', 'myself') and second-person pronouns ('you', 'your') were not set as stop words because they are important expressions in lyrics (Hu & Stephen Downie, 2010).

I used two prominent NLP methods: bag of words (BoW) and latent Dirichlet allocation (LDA). First, an effective approach of representing documents is the BoW model (Schwartz, Eichstaedt, Kern, et al., 2013). In this model, word histograms are constructed, in which the frequencies of words in a dictionary within a text document are counted. I determined the relative frequency with which users used words (unigrams) and two-word phrases (bigrams). To examine the effective features, I removed words and phrases that were used in less than 3% and in more than 95% of the lyrics, yielding a 186 (pieces of music) $\times$ 454 (number of tokens) BoW matrix from the total lyric lines. The tokens included 438 Japanese and 16 English words. Finally, I weighted the BoW matrix to term frequency-inverse document frequency (tf-idf) to emphasize the important lyric words and phrases. Tf-idf reflects how quantitively important a word is to a document in all documents and is one of the most widely used term weight algorithms (Aizawa, 2003).

Second, topics are clusters of semantically related words created using the LDA algorithm (Blei, Carin, & Dunson, 2010). LDA assumes that a document (in this case, individual lyrics) is a mixture of a fixed number of latent topics, where each topic is a cluster of related words (this fixed number is specified in advance by the analyst). Through an iterative procedure, LDA identifies and refines the specified number of word clusters in the lyric samples. Words in the same topic tend to co-occur in lyrics and are automatically identified by the LDA hierarchical Bayesian algorithm. I fit an LDA model for the BoW matrix weighted by tf-idf using an implementation provided in the scikit-learn package, setting the number of topics to 20. This produced 20 naturally occurring topics, each consisting of many words with relative weights. I adjusted only one hyperparameter ($\alpha = 0.1$) to favor fewer topics per document. The α-value was smaller than the guidelines (Griffiths & Steyvers, 2004) because individual lyrics tend to contain fewer topics than typical documents (e.g., newspaper and encyclopedia articles) to which LDA is applied. I then calculated all 20 topic scores for each lyric, defined as the probability of using a topic.

3.1.4. Machine learning and statistical tests

A machine learning regression task with 186 pieces of music from 54 subjects was used to determine how the multivariate pattern of acoustic and lyrical features predicts the number of chill and tear responses. The total number of features was 505 (31 acoustic features and 474 lyrical features). Using the scikit-learn library, I trained a linear ridge regressor with feature scaling to predict the number of chill and tear responses. As in Study 1, a nested 10F-CV was applied for model generalizability for unseen participants. Machine learning accuracy was evaluated using the correlation coefficient between the self-reported score and the machine learning predicted score. For the correlation coefficient, a permutation test was applied to obtain a p-value with Bonferroni correction ($p < .05/2$). In addition, the ridge regression coefficient was tested using bootstrap resampling (see the Method section of Study 1).

3.2. Results and discussion

3.2.1. Prediction accuracy

As shown in Fig. 4, the prediction score from ridge regression was positively correlated with the number of chill (r = 0.02) and tear (r = 0.35) responses. The permutation test with Bonferroni correction indicated that the correlation of chill responses was not statistically significant (p = .78, Bonferroni correction), but the correlation of tear responses was highly significant ($p < .001$, Bonferroni correction). The results indicate that audio and lyric content predict the frequency of tear responses but not chill responses. The audio and lyric regression models overwhelmed the prediction accuracy of tear responses for each audio (r = 0.28) and lyric (r = 0.30) regression model (see Supplemental Fig. 4).

3.2.2. Acoustic and lyrical feature importance for tear responses

The important regression coefficients for tear response were confirmed by bootstrap estimation of the CI, resulting in a total of 64 features (10 acoustics, 47 words, and 7 topics; the detailed ridge weights can be found in Supplemental Table 3) that were statistically significant ($p < .05$, all p-values can be found in Supplemental Table 3).

As shown in Fig. 5 (circular bar plot), in the acoustic features, "Harmony_key clarity" and "Harmony_mode" showed a strong positive coefficient (blue, A+), whereas "Rhythm_centroid" and "Rhythm_entropy" showed a strong negative coefficient (black, A-). These results indicate that major chords that can express positive emotions (Gabrielsson & Lindström, 2010) with distinct tonalities promote tear responses; however, rhythmicity and its complexity suppress tear responses. These effective features partly corresponded with the results of Study 1. In the BoW features, positive coefficient words (blue, L+) include "someday," "farewell," "meet," and "can go," whereas negative coefficient words (black, L-) include "destiny," "want," and "now" (the Japanese words were semi-automatically translated into English using Deep L; https://www.deepl.com/ja/home). These words were effective in increasing and decreasing tear responses, respectively.

Furthermore, in the topic features in the circular bar plot, topics with positive coefficients included topics 6 and 14 (Fig. 5, top). Specifically, topic 6 represents "you" (strongest loading of "you") and topic 14 represent "painful separation" (includes "goodbye," "I can," and "sadness"). However, topics with negative coefficients included topics 1, 3, 8, and 18 (Fig. 5, bottom). Topic 1 reflects "mental conflict" (e.g., "memories," "I believe," "not"), topic 3 reflects "enthusiasm" (e.g., "a lot," "gather," "fun"), topic 8 reflects "calmness" (e.g., "now," "quiet," "close"), and topic 18 represents "positive" (e.g., "stand up," "born again," "follow me").

In summation, tear responses can be associated with the acoustics of positive impression and less rhythmicity, and lyrics drawing on a significant person and sad farewell, but not lyrics drawing on positivity, enthusiasm, calm feelings, and mental conflict.

4. General discussion

In the present study, I conducted machine learning classification and regression tasks for peak emotional responses to music by applying the MIR and NLP methods. Multivariate acoustic features successfully decoded chill and tear responses during specific time periods. Acoustic and lyrical features derived from the entire piece decoded the frequency of tear responses, but not chill responses. The computational acoustic and lyrical features associated with chill and tear responses may indicate why these responses occurred and what they were.

4.1. Decoding emotional chills and tears by temporary acoustic features

The multivariate analytic approach revealed that acoustic features classify emotional chills and tears. The current time-series examination found multistep contributions of acoustic features for classification. Acoustic differences between chills and tears have three key periods: the first period occurs several seconds before, the second period occurs at onset, and the third period occurs after onset. The classification accuracy

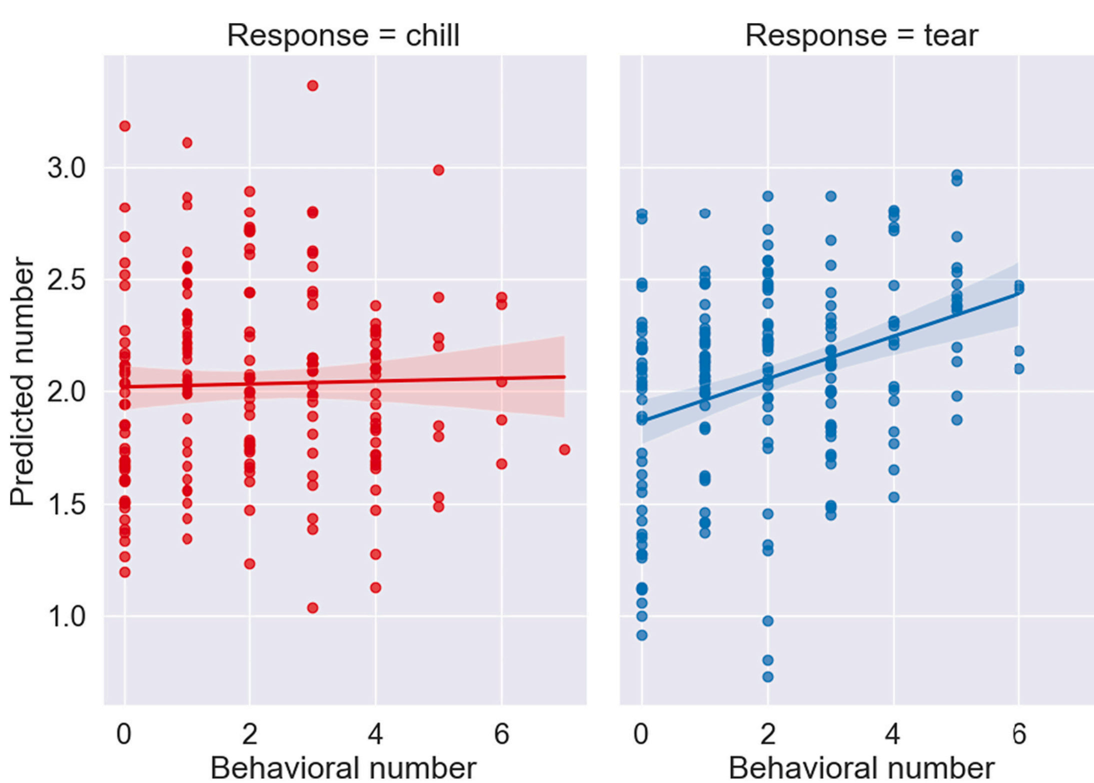

Fig. 4. Scatter plots of correlations between the behavioral number of peak emotions and the predicted number of peak emotions. The predicted number is computed by cross-validated prediction. Shaded ribbons indicate the 95% CIs for the estimates.

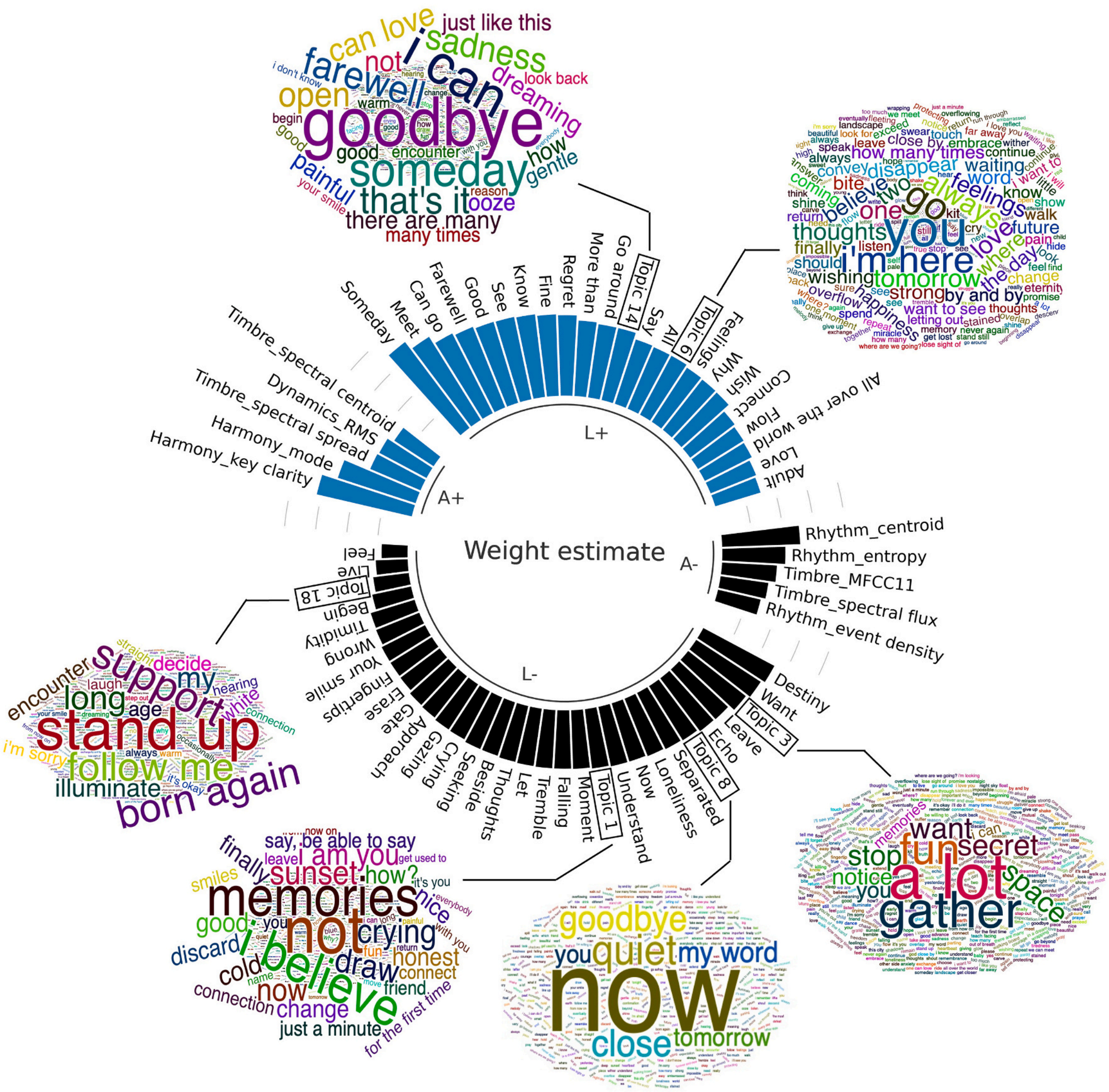

Fig. 5. Acoustic and lyrical features showing a significant predictor for tear responses. The circular bar plot contains acoustic and lyrical features with a significant regression weight with the number of tear responses. Bar color indicates positive (blue) or negative (black) direction. A+ and L+ refers to the positive coefficient of the audio and lyrics, respectively; A- and L- refers to the negative coefficient of the audio and lyrics. The surrounding word clouds are the six topics of semantically related words. Within topics, word size indicates word prevalence; the color is merely a visualization aid. RMS, root-mean-square; MFCC, mel-frequency cepstral coefficients. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of chill responses was better than that of tear responses. Time-series acoustic features explain chill responses, as opposed to acoustic features from the entire piece that failed to predict chill responses in Study 2.

The former two periods suggest that two-step changes in acoustic features can lead to chills and tears. Previous studies have examined acoustic changes, such as loudness just around chill onset (Beier et al., 2020; Grewe et al., 2007; Guhn, Hamm, & Zentner, 2007; Nagel et al., 2008). Classification accuracy was best for onset, and the period was the most important and a trigger for evoking chill and tear responses. However, the current findings of two-step acoustic changes suggest that evoking peak emotional responses requires multiple stages of the cognitive state. It has long been believed that musical emotions are derived from expectations and their violations (Meyer, 1956). It is possible that the expectation for acoustic change starts when the first acoustic changes emerge, after which the second acoustic cues follow or violate the expectation. Theoretical research has suggested that musical expectancy is an involuntary reaction to potentially important events (Juslin, 2013). Such an expectation evokes tension, and the anticipated resolution evokes emotional responses to music even after the listener is familiar with the music (Huron, 2006; Koelsch, 2012). When considered in tandem with the current results, within a few seconds, the interplay between listeners' anticipation-resolution and successive musical events may lead to peak emotion.

Additionally, acoustic contribution after onset may reflect listeners' prediction of the next musical event when they experience peak emotions. Since the experimental stimuli used in this research were the favorite music of each participant, the participant could predict the progression of the music. When participants reported chills or tears, they were able to imagine how the piece would progress. As listeners of music, we constantly generate plausible hypotheses about what may occur next (Koelsch, Vuust, & Friston, 2018); thus, it is possible that when anticipation-resolution ends, the next anticipation-resolution has already started. This leads to the speculation that convoluted anticipation-resolution progress evokes strong emotions, such as chills and tears. This complexity may explain why familiar music frequently evokes peak emotional responses rather than non-familiar music (Koelsch, 2012; Lehne & Koelsch, 2015). Future studies should confirm this hypothesis by conducting experiments using controlled stimuli.

4.2. Important temporary acoustic features that decode emotional chills and tears

The current results of multivariate acoustic analysis shed light on the acoustic features that are effective in evoking emotional chills and tears. Although much insight has been gained from previous studies that have examined the univariate correlation between acoustic information and chill responses (Bannister & Eerola, 2018; Beier et al., 2020; Grewe et al., 2007; Guhn et al., 2007; Nagel et al., 2008), it is difficult to determine which acoustic features are more important than others because the acoustic features of music are mutually correlated (Lange & Frieler, 2018). Here, the L2 norm ridge classification and out-of-sample prediction resolved the problem and provided specifically important features.

For chills, ambiguous minor mode (negative coefficient of mode and key clarity) and timbral components were significant several seconds before peak emotion onset, whereas rhythm event density and entropy with other timbral features were effective from the onset. This can be interpreted as meaning that temporally ambiguous minor mode and a timbre increase tension or uncertainty, while high event density and entropy of rhythm, accompanied by a new timbre, provide online resolution of tension or violate listeners' expectations; this has been identified as a key mechanism for the evocation of emotions in music (Juslin, 2013; Koelsch, 2012), thereby resulting in chills being evoked. Moreover, in the resolution time, the musical key is still ambiguous, and participants may anticipate further excitement due to percussive and loud sounds after a few seconds. Chill-related physiological arousal (Beier et al., 2020; Mori & Iwanaga, 2021), as well as rewards (Blood & Zatorre, 2001; Salimpoor et al., 2011), may be evoked by the complex manner of anticipation-resolution, which may provide special excitement. The multistep acoustic change should be further examined to reveal the chill-evoking mechanisms. Note that the dynamics increased chill onset, as shown in Fig. 1, which is similar to the findings of previous studies (Beier et al., 2020; Grewe et al., 2007; Guhn et al., 2007; Nagel et al., 2008); however, the dynamic effect did not survive in the multivariate analysis. The dynamic effect for chills may be a spurious correlation by the mid-to-high correlation with event density ($r = 0.44$) and spectral flux ($r = 0.73$) (see Supplemental Fig. 1).

The current analysis indicates that rhythmic features may become a trigger and are especially important for evoking chills, which aligns with previous findings that sudden dynamic or textural changes are associated with chills (Sloboda, 1991). Musical systems are decomposable in a series of components related to pitch and rhythm (Brown & Jordania, 2013), and rhythm is commonly cited as a putative universal in music evolution research (Savage et al., 2015). Most music has an isochronous beat and metric structure (Savage et al., 2015; Savage et al., 2020). Higher rhythm event density and entropy may violate these predictable and repetitive rhythmic components. The universality of rhythm may support that the violation of rhythm is a biological/evolutionary chills signal. While being surprised by violated expectations may be a biological failure to anticipate the future, the surprise would play the role of an emotional amplifier (Huron, 2006). The violation of rhythm may boost both the pleasure of musical chills and “fight” physiological response (Huron, 2006). The physiological response may be acquired as an evolutionary defense reflex (Shwartz, Gonzalez-celeiro, Chen, et al., 2020); however, the safe situation of music listening can be accompanied by a pleasurable feeling. People may be attracted to these specific pleasures and the reward (Blood & Zatorre, 2001; Salimpoor et al., 2011; Zatorre & Salimpoor, 2013) of chills. If chill-evoking music in various cultures includes high rhythm event density and entropy, the violation of rhythm may be a biologically acquired chills cue.

For tears, since key clarity and MFCC1 indicate the degree of key clearness (Eerola, 2012; Lange & Frieler, 2018) and a strong negative correlation with brightness, respectively, the tears section in music constantly has a clear chord and dark timbre. Furthermore, HCDF, which represents a sign of harmony or chord change (Lange & Frieler, 2018), contributes to the prediction of tears both before and after peak onset. Within the harmonic movement and dark timbre, the major mode (positive coefficient of mode), followed by high-novelty acoustics, were significant several seconds before peak emotion onset. This section may provide the anticipation or tension of the next musical event. After the anticipation period, a timbre component with MFCC1, 2, 3, and 7 emerged at tear onset. Although MFCC is difficult to interpret directly (Lange & Frieler, 2018), timbre could be a trigger that evokes tears. A pioneer survey suggested that melodic appoggiaturas are likely to evoke emotional tears (Sloboda, 1991). The timbral features for tear onset can be speculated to reflect melodic appoggiaturas. Additionally, in the resolution time, participants may predict the events in a few seconds, with chord changes including loud, non-dissonant sounds, and the pleasure that would accompany it. These results indicate that the anticipation-resolution rule for tears may differ from that of chills. The physiological calming effect of tears (Mori & Iwanaga, 2017; Mori & Iwanaga, 2021) may be evoked by the acoustic mechanisms of sound resolution from tension.

Although a timbral component may be a trigger for tears, harmony and its movement with a dark timbre showed positive effects for tears before and after onset. Harmonic features were also important in decoding emotional tears from the entire piece in Study 2. This effect contrasts with the chills section, where harmony was unclear. Pitch is an essential component of music and is cited, along with rhythm, as a putative universal in music evolution research (Brown & Jordania, 2013; Savage et al., 2015). Discrete pitch in harmony is widespread throughout global musical systems (Savage et al., 2015). The music and social bonding hypothesis indicate the evolutionary social bonding effects of harmony and melody; thus, tears may be evoked by social bonding feelings, as supported by general tear perspectives (Gračanin, Bylsma, & Vingerhoets, 2018; Vingerhoets, 2013). The hypothesis also suggests the social bonding effects of rhythm and other musical characteristics; therefore, I speculate that the bonding promoted by the harmonious overlapping of acoustic spectra (Savage et al., 2020), including dark sounds, is specifically important for tear responses.

4.3. Decoding emotional chills and tears by acoustic and lyrical features from an entire piece

Machine learning using the information extracted from the whole piece revealed that emotional tears, but not chills, could be predicted by audio and lyric content. The results supported a previous survey that found that tears were mainly evoked by music consisting of vocal and instrumental components (Hanser, Mark, & Vingerhoets, 2021). Based on evidence from Studies 1 and 2, chills are primarily evoked by relatively lower cognitive functions (i.e., perceptual processing for momentary acoustic features), whereas tears are evoked by not only lower- but also higher-order cognitive functions (i.e., a fusion of acoustic features and lyric content processing). The combination of acoustic and lyrical features is speculated to be effective in evoking a tear response, rather than each feature, because the interaction effect of audio and lyrics, such as musical harmony, enhances the meaning of language; words deepen the musical image. Moreover, the experience of tears may be why vocal songs, including lyrics, are a global feature of music across many cultures (Mehr et al., 2019; Savage et al., 2015). It is speculated that people seek tear responses through vocal songs. Tear evoking with higher-order processing may also be related to the fact that humans are the only animals who can weep (Bellieni, 2017).

4.4. Important acoustic and lyrical features that decode emotional tears

The machine learning model in Study 2 also revealed a combination of acoustic and lyrical features that can evoke emotional tears. Acoustic coefficients suggest that a major chord with less rhythmicity promotes the prevalence of tear responses. The major chord is biologically important for social sound signals (Bowling & Purves, 2015); thus, the social bonding effect for tears corresponds with Study 1. Therefore, tear responses may be evoked by the same local and global acoustic functional role. In contrast, words and topic coefficients indicated that a story of a sad farewell and a significant person ("you") but not positive and enthusiastic feelings increased the frequency of tears. Emotional tears from music are accompanied by strong pleasure (Mori & Iwanaga, 2017; Mori & Iwanaga, 2021); therefore, the fusion analysis of audio and lyrics prove that happiness from major chords with sad lyric content can evoke pleasurable tears. These findings support the connection between mixed feelings of positive and negative affect and strong emotional experiences in artwork (Berrios, Totterdell, & Kellett, 2015; Wassiliwizky & Menninghaus, 2021). Mixed feelings are integrated as a complex emotional trajectory yielding a distinctive feeling, such as sadness for the loss but also joy at the memory of what has been lost (Cullhed, 2019; Hanich, Wagner, Shah, et al., 2014). The listener can experience complex emotions in music that would evoke sad but pleasurable tears through psychological mechanisms such as contagion, visual imagery, and episodic memory (Juslin, 2013; Juslin, 2021). Complex emotions may have the emotional regulation and mood stabilizing effect of tears (Bellieni, 2017; Miceli & Castelfranchi, 2003) which is unique to the social bonding process, such as easing the pain of separation from loved ones by listening to music that evokes shared memories (Kornhaber, 2020; Savage et al., 2020). Since studies have measured mixed emotions using subjective ratings, further research on the combination of acoustic and lyric content and emotional tears may open new horizons for mixed emotions in art.

Furthermore, it may be culture-specific to the link between tear responses and the combination of major chords and sad lyric content. Recent cross-cultural studies between the US and China have revealed that bittersweet, mixed emotions do not show a high correlation (Cowen, Fang, Sauter, & Keltner, 2019). Mixed emotions perceived from the combination of audio and lyrics can differ from one culture to another. It is interesting to note that different emotional combinations are related to tear experiences in different kinds of cultural music, especially traditional music. Tear responses for vocal songs may also not emerge in cultures where major and minor chords are not perceived as having different emotional qualities (Mcdermott, Schultz, Undurraga, & Godoy, 2016). Similar to the emotional response to the chord, the tear response to a song may be shaped by culture.

4.5. Limitations

The current study has several limitations. First, specific acoustic features may be necessary but not sufficient to evoke peak emotions because an emotional stimulus may not always elicit the same emotional response in different or even the same individual (Sander & Nummenmaa, 2021). To understand the conditions that evoke peak emotions, future studies should examine both stimulus features and evaluative processes for emotions (Scherer & Moors, 2019). Second, the analyzed musical stimuli were self-selected. Although peak emotions are easily evoked by self-selected music (Mori & Iwanaga, 2017; Salimpoor et al., 2011; Savage, Jansen, Stringer, et al., 2018), self-selected music has strong biases, such as high familiarity, listening memory, and knowledge of the artist's history. Since several studies have shown that specific music can evoke peak emotions even when selected by the experimenter (Beier et al., 2020; Salimpoor, van den Bosch, Kovacevic, et al., 2013), future studies should investigate the relationship between acoustic and lyrical features and peak emotions evoked by novel music for participants. Third, the analyzed musical stimuli were limited to the pop or rock genres. Previous studies have indicated that people listen to and recognize several genres of music in everyday life (Rentfrow & Gosling, 2003; Rentfrow & Levitin, 2012), and acoustic and lyrical features differ from genre to genre (Mauch, MacCallum, Levy, & Leroi, 2015; Ying, Doraisamy, & Abdullah, 2012). The examination of multiple genres of music may reveal common and different features to decode emotional chills and tears and generalize the current results. Fourth, chills and tears are sometimes co-activated (Mori & Iwanaga, 2021; Wassiliwizky et al., 2017), and each response may be of different types (Bannister, 2020; Maruskin, Thrash, & Elliot, 2012; Vingerhoets, 2013). Although I examined the quad classification task between chill, tear, and neutral responses, as well as chills with tears as a preliminary test (see Supplemental Fig. 2), future studies should investigate the decoding possibility of not only the co-activation of chills and tears, but also of each subtype. Finally, studies have indicated that trait and state differences are key factors in evoking peak emotions in music (Ferreri et al., 2019; Martínez-Molina et al., 2016; Mas-Herrero et al., 2014; Mas-Herrero et al., 2021). Acoustic and lyrical features may have interactive effects on the traits and states. For example, individuals with musical anhedonia (Martínez-Molina et al., 2016; Mas-Herrero et al., 2014) may find it difficult to respond to chill-related acoustic cues; furthermore, listeners may easily experience tears by lyrical features when fronto-striatal pathways in the brain are activated (Mas-Herrero et al., 2021). An integrated examination of music features and these trait and state differences would provide a deeper understanding of why peak emotional responses occur through music.

5. Conclusion

In the present study, I examined whether computational acoustic and lyrical features decode peak emotional responses to music. Machine learning analysis by L2-norm-regularized regression revealed that time-series acoustic features decode emotional chills and tears using the information a few seconds before and after peak onset. Both chills and tears may be evoked by the complex anticipation-resolution mechanism; however, chills are primarily associated with the violation of rhythm expectation, whereas tears are primarily related to harmonics. Acoustic and lyrical features from the entire piece decode tear frequency, but not chill frequency. Mixed emotions from happiness stemming from major chords and sad farewell lyrical content can help predict emotional tears. When considered in tandem with theoretical studies, the violation of rhythm may biologically boost both the pleasure- and the fight-related physiological responses of chills. However, tears may be evolutionarily embedded in the social bonding effect of musical harmony and play a unique role in emotion regulation. The current study establishes a new perspective regarding why peak emotional responses are evoked by music and what these responses are.

Acknowledgments

This research was supported by the Japan Society for the Promotion of Science (JSPS), Grant-in-Aid for JSPS Fellows (#2306670), and Fostering Joint International Research (A) (#19KK0316).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2021.105010.

References

Aizawa, A. (2003). An information-theoretic perspective of tf-idf measures. *Information Processing and Management, 39*, 45–65. https://doi.org/10.1016/S0306-4573(02)00021-3

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., et al. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage, 59*, 3677–3689. https://doi.org/10.1016/j.neuroimage.2011.11.019

Bannister, S. (2019). Distinct varieties of aesthetic chills in response to multimedia. *PLoS One, 14*. https://doi.org/10.1371/journal.pone.0224974

Bannister, S. (2020). A vigilance explanation of musical chills? Effects of loudness and brightness manipulations. *Music & Science, 3*. https://doi.org/10.1177/2059204320915654, 205920432091565.

Bannister, S., & Eerola, T. (2018). Suppressing the chills: Effects of musical manipulation on the chills response. *Frontiers in Psychology, 9*, 2046.

Beier, E. J., Janata, P., Hulbert, J. C., & Ferreira, F. (2020). *Do you chill when I chill? A cross-cultural study of strong emotional responses to music. Psychology of aesthetics, creativity, and the arts*. https://doi.org/10.1037/aca0000310

Bellieni, C. V. (2017). Meaning and importance of weeping. *New Ideas in Psychology, 47*, 72–76. https://doi.org/10.1016/j.newideapsych.2017.06.003

Bello, J. P., Daudet, L., Abdallah, S., et al. (2005). A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing, 13*, 1035–1046. https://doi.org/10.1109/TSA.2005.851998

Benedek, M., & Kaernbach, C. (2011). Physiological correlates and emotional specificity of human piloerection. *Biological Psychology, 86*, 320–329. https://doi.org/10.1016/j.biopsycho.2010.12.012

Berrios, R., Totterdell, P., & Kellett, S. (2015). Eliciting mixed emotions: A meta-analysis comparing models, types, and measures. *Frontiers in Psychology, 6*, 1–15. https://doi.org/10.3389/fpsyg.2015.00428

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*.

Blei, D., Carin, L., & Dunson, D. (2010). Probabilistic topic models. *IEEE Signal Processing Magazine, 27*, 55–65. https://doi.org/10.1109/MSP.2010.938079

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 11818–11823. https://doi.org/10.1073/pnas.191355898

Bowling, D. L., & Purves, D. (2015). A biological rationale for musical consonance. *Proceedings of the National Academy of Sciences of the United States of America, 112*, 11155–11160. https://doi.org/10.1073/pnas.1505768112

Brown, S., & Jordania, J. (2013). Universals in the world's musics. *Psychology of Music, 41*, 229–248. https://doi.org/10.1177/0305735611425896

Brwon, D. E. (1991). *Human universals*. New York: McGraw Hill.

Cohn, R. (1998). Introduction to neo-Riemannian theory: A survey and a historical perspective. *Journal of Music Theory, 42*, 167–180.

Cowen, A. S., Fang, X., Sauter, D., & Keltner, D. (2019). What music makes us feel : At least 13 dimensions organize subjective experiences associated with music across different cultures. *Proceedings of the National Academy of Sciences*, 1–11. https://doi.org/10.1073/pnas.1910704117

Cullhed, E. (2019). What evokes being moved ? *Emotion Review*. https://doi.org/10.1177/1754073919875216

Dubé, L., & le Bel, J. L. (2003). The content and structure of laypeople's concept of pleasure. *Cognition and Emotion, 17*, 263–295. https://doi.org/10.1080/02699930302295

Eerola, T. (2012). Modeling listeners' emotional response to music. *Topics in Cognitive Science, 4*, 607–624. https://doi.org/10.1111/j.1756-8765.2012.01188.x

Ellis, D. P. W. (2007). Beat tracking by dynamic programming. *Journal of New Music Research, 36*, 51–60. https://doi.org/10.1080/09298210701653344

Ferreri, L., Mas-Herrero, E., Zatorre, R. J., et al. (2019). Dopamine modulates the reward experiences elicited by music. *Proceedings of the National Academy of Sciences of the United States of America, 116*, 3793–3798. https://doi.org/10.1073/pnas.1811878116

Foote, J. T., & Cooper, M. L. (2003). Media segmentation using self-similarity decomposition. *Storage and Retrieval for Media Databases, 5021*, 167. https://doi.org/10.1117/12.476302, 2003.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *Psychology of music*. New York: Academic Press.

Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. N. Juslin, & J. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 367–400). Oxford University Press.

Goldstein, A. (1980). Thrills in response to music and other stimuli. *Physiological Psychology, 8*, 126–129.

Goupil, L., & Aucouturier, J.-J. (2019). Musical pleasure and musical emotions. *Proceedings of the National Academy of Sciences, 201900369*. https://doi.org/10.1073/pnas.1900369116

Gračanin, A., Bylsma, L. M., & Vingerhoets, A. J. J. M. (2018). Why only humans shed emotional tears. *Human Nature, 29*, 104–133.

Greenberg, D. M., Matz, S. C., Schwartz, H. A., & Fricke, K. R. (2020). The self-congruity effect of music. *Journal of Personality and Social Psychology, 2*.

Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2007). Listening to music as a re-creative process: Physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Perception: An Interdisciplinary Journal, 24*, 297–315.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America, 101*(Suppl), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Guhn, M., Hamm, A., & Zentner, M. (2007). Physiological and musico-acoustic correlates of the chill response. *Music Perception: An Interdisciplinary Journal, 24*, 473–484.

Hanich, J., Wagner, V., Shah, M., et al. (2014). Why we like to watch sad films. The pleasure of being moved in aesthetic experiences. *Psychology of Aesthetics, Creativity, and the Arts, 8*, 130–143. https://doi.org/10.1037/a0035690

Hanser, W. E., Mark, R. E., & Vingerhoets, A. J. J. M. (2021). Everyday crying over music: A survey. *Musicae Scientiae*. https://doi.org/10.1177/1029864920981110

Harte, C., Sandler, M., & Gasser, M. (2006). Detecting harmonic change in musical audio. In *Proceedings of the ACM international multimedia conference and exhibition* (pp. 21–26). https://doi.org/10.1145/1178723.1178727

Hu, X., & Stephen Downie, J. (2010). When lyrics outperform audio for music mood classification: A feature analysis. In *, 2010. Proceedings of the 11th international society for music information retrieval conference* (pp. 619–624). ISMIR.

Huron, D. (2006). *Sweet anticipation*. https://doi.org/10.7551/mitpress/6575.001.0001

Juslin, P. (2021). Mind the gap: The mediating role of emotion mechanisms in social bonding through musical activities. *Behavioral and Brain Sciences, 44*, Article e80. https://doi.org/10.1017/s0140525x2000120x

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1797–1813.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews, 10*, 235–266. https://doi.org/10.1016/j.plrev.2013.05.008

Koelsch, S. (2012). *Brain and music*. West Sussex: JohnWiley & Sons, Ltd.

Koelsch, S., Vuust, P., & Friston, K. (2018). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences, xx*, 1–15. https://doi.org/10.1016/j.tics.2018.10.006

Kohoutová, L., Heo, J., Cha, S., et al. (2020). Toward a unified framework for interpreting machine-learning models in neuroimaging. *Nature Protocols, 15*. https://doi.org/10.1038/s41596-019-0289-5

Konečni, V. J., Wanic, R. A., & Brown, A. (2007). Emotional and aesthetic antecedents and consequences of music-induced thrills. *American Journal of Psychology, 120*, 619–643. https://doi.org/10.2307/20445428

Kornhaber, S. (2020). The coronavirus has forced a repurposing of music. In *The Atlantic*.

Koumura, T., Nakatani, M., Liao, H.-I., & Kondo, H. M. (2020). Dark, loud, and compact sounds induce frisson. *Quarterly Journal of Experimental Psychology, 174702182097717*. https://doi.org/10.1177/1747021820977174

Krumhansl, C. L. (2004). The cognition of tonality–as we know it today. *Journal of New Music Research, 33*, 253–268. https://doi.org/10.1080/0929821042000317831

Kudo, T., Yamamoto, K., & Matsumoto, Y. (2004). Applying conditional random fields to Japanese morphological analysis. In *Proceedings of the conference on empirical methods in natural language processing - EMNLP'04. Barcelona, Spain* (pp. 230–237).

Lange, E. B., & Frieler, K. (2018). Challenges and opportunities of predicting musical emotions with perceptual and automatized features. *Music Perception: An Interdisciplinary Journal, 36*, 217–242. https://doi.org/10.1525/mp.2018.36.2.217

Lartillot, O., Lartillot, O., Toiviainen, P., & Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio. In *International conference on digital audio …* (pp. 1–8). https://doi.org/10.1007/978-3-540-78246-9_31

Lehne, M., & Koelsch, S. (2015). Toward a general psychological model of tension and suspense. *Frontiers in Psychology, 6*, 1–11. https://doi.org/10.3389/fpsyg.2015.00079

Machens, C. K., Wehr, M. S., & Zador, A. M. (2004). Linearity of cortical receptive fields measured with natural sounds. *Journal of Neuroscience, 24*, 1089–1100. https://doi.org/10.1523/JNEUROSCI.4445-03.2004

Marozzi, M. (2004). Some remarks about the number of permutations one should consider to perform a permutation test. *Statistica, 64*, 193–201. https://doi.org/10.6092/issn.1973-2201/32

Martínez-Molina, N., Mas-Herrero, E., Rodríguez-Fornells, A., et al. (2016). Neural correlates of specific musical anhedonia. *Proceedings of the National Academy of Sciences, 113*, E7337–E7345.

Maruskin, L., Thrash, T., & Elliot, A. (2012). The chills as a psychological construct: Content universe, factor structure, affective composition, elicitors, trait antecedents, and consequences. *Journal of Personality and Social Psychology, 103*, 135–157. https://doi.org/10.1037/a0028117

Mas-Herrero, E., Dagher, A., Farrés-Franch, M., & Zatorre, R. J. (2021). Unraveling the temporal dynamics of reward signals in music-induced pleasure with TMS. *Journal of Neuroscience, 41*. https://doi.org/10.1523/jneurosci.0727-20.2020. JN-RM-0727-20.

Mas-Herrero, E., Zatorre, R. J., Rodriguez-Fornells, A., & Marco-Pallarés, J. (2014). Dissociation between musical and monetary reward responses in specific musical anhedonia. *Current Biology: CB, 24*, 699–704. https://doi.org/10.1016/j.cub.2014.01.068

Mauch, M., MacCallum, R. M., Levy, M., & Leroi, A. M. (2015). The evolution of popular music: USA 1960–2010. *Royal Society Open Science, 2*, Article 150081. https://doi.org/10.1098/rsos.150081

Mcdermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature, 25*, 21–25. https://doi.org/10.1038/nature18635

Mehr, S. A., Singh, M., Knox, D., et al. (2019). Universality and diversity in human song. *Science, 366*, 1–17. https://doi.org/10.1126/science.aax0868

Menninghaus, W., Wagner, V., Hanich, J., et al. (2015). Towards a psychological construct of being moved. *PLoS One, 10*, Article e0128451. https://doi.org/10.1371/journal.pone.0128451

Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago, IL: Chicago University Press.

Miceli, M., & Castelfranchi, C. (2003). Crying: Discussing its basic reasons and uses. *New Ideas in Psychology, 21*, 247–273. https://doi.org/10.1016/j.newideapsych.2003.09.001

Mori, K., & Iwanaga, M. (2014a). Pleasure generated by sadness: Effect of sad lyrics on the emotions induced by happy music. *Psychology of Music, 42*, 643–652. https://doi.org/10.1177/0305735613483667

Mori, K., & Iwanaga, M. (2014b). Resting physiological arousal is associated with the experience of music-induced chills. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology, 93*, 220–226. https://doi.org/10.1016/j.ijpsycho.2014.05.001

Mori, K., & Iwanaga, M. (2015). General reward sensitivity predicts intensity of chills. *Music Perception: An Interdisciplinary Journal*, 484–492.

Mori, K., & Iwanaga, M. (2017). Two types of peak emotional responses to music: The psychophysiology of chills and tears. *Scientific Reports, 7*, 46063. https://doi.org/10.1038/srep46063

Mori, K., & Iwanaga, M. (2021). Being emotionally moved is associated with phasic physiological calming during tonic physiological arousal from pleasant tears. *International Journal of Psychophysiology, 159*, 47–59. https://doi.org/10.1016/j.ijpsycho.2020.11.006

Nagel, F., Kopiez, R., Grewe, O., & Altenmüller, E. (2008). Psychoacoustical correlates of musically induced chills. *Musicae Scientiae, 12*, 101–113.

Pampalk, E., Rauber, A., & Merkl, D. (2002). Content-based organization and visualization of music archives. *Proceedings of the ACM International Multimedia Conference and Exhibition*, 570–579. https://doi.org/10.1145/641007.641121

Rentfrow, P. J. (2012). The role of music in everyday life: Current directions in the social psychology of music. *Social and Personality Psychology Compass, 6*, 402–416. https://doi.org/10.1111/j.1751-9004.2012.00434.x

Rentfrow, P. J., & Gosling, S. D. (2003). The do re Mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84*, 1236–1256. https://doi.org/10.1037/0022-3514.84.6.1236

Rentfrow, P. J., & Levitin, D. J. (2012). The structure of musical preferences: A five-factor model. *Journal of Personality and Social Psychology, 100*, 1139–1157. https://doi.org/10.1037/a0022406.The

Sachs, M. E., Ellis, R. J., Schlaug, G., & Loui, P. (2016). Brain connectivity reflects human aesthetic responses to music. *Social Cognitive and Affective Neuroscience, 11*, 884–891.

Salimpoor, V. N., Benovoy, M., Larcher, K., et al. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience, 14*, 257–262. https://doi.org/10.1038/nn.2726

Salimpoor, V. N., van den Bosch, I., Kovacevic, N., et al. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science (New York, N.Y.), 340*, 216–219. https://doi.org/10.1126/science.1231059

Sander, D., & Nummenmaa, L. (2021). Reward and emotion: An affective neuroscience approach. *Current Opinion in Behavioral Sciences, 39*, 161–167. https://doi.org/10.1016/j.cobeha.2021.03.016

Satou, T. (2015). *Mecab-ipadic-neologd*.

Savage, J. E., Jansen, P. R., Stringer, S., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics, 50*, 912–919. https://doi.org/10.1038/s41588-018-0152-6

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences USA.*. https://doi.org/10.1073/pnas.1414495112

Savage, P. E., Loui, P., Tarr, B., et al. (2020). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X20000333

Scheirer, E., & Slaney, M. (1997). Construction and evaluation of a robust multifeature speech/music discriminator. In *, 2. ICASSP, IEEE international conference on acoustics, speech and signal processing - Proceedings* (pp. 1331–1334). https://doi.org/10.1109/icassp.1997.596192

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology, 70*, 719–745. https://doi.org/10.1146/annurev-psych-122216-011854

Schubert, T. W., Zickfeld, J. H., Seibt, B., & Fiske, A. P. (2018). Moment-to-moment changes in feeling moved match changes in closeness, tears, goosebumps, and warmth: Time series analyses. *Cognition and Emotion, 32*, 174–184. https://doi.org/10.1080/02699931.2016.1268998

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., et al. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS One, 8*. https://doi.org/10.1371/journal.pone.0073791

Sethares, W. A. (2005). *Tuning, timbre, spectrum, scale*. London, UK: Springer.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Shwartz, Y., Gonzalez-celeiro, M., Chen, C., et al. (2020). Cell types promoting goosebumps form a niche to regulate hair follicle stem cells. *Cell*, 1–16. https://doi.org/10.1016/j.cell.2020.06.031

Silvia, P. J., & Nusbaum, E. C. (2011). On personality and piloerection: Individual differences in aesthetic chills and other unusual aesthetic experiences. *Psychology of Aesthetics, Creativity, and the Arts, 5*, 208–214. https://doi.org/10.1037/a0021914

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music, 19*, 110–120. https://doi.org/10.1177/0305735691192002

Sumpf, M., Jentschke, S., & Koelsch, S. (2015). Effects of aesthetic chills on a cardiac signature of emotionality. *PLoS One, 10*, Article e0130117. https://doi.org/10.1371/journal.pone.0130117

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals using geometric methods. *IEEE Transactions on Speech and Audio Processing, 10*, 293–302.

Vingerhoets. (2013). *Why only humans weep: Unraveling the mysteries of tears*. Oxford: Oxford university press.

Vingerhoets, A. J. J. M., & Bylsma, L. M. (2016). The riddle of human emotional crying: A challenge for emotion researchers. *Emotion Review, 8*, 207–217. https://doi.org/10.1177/1754073915586226

Wassiliwizky, E., Jacobsen, T., Heinrich, J., et al. (2017). Tears falling on goosebumps: Co-occurrence of emotional lacrimation and emotional piloerection indicates a psychophysiological climax in emotional arousal. *Frontiers in Psychology, 8*, 41. https://doi.org/10.3389/fpsyg.2017.00041

Wassiliwizky, E., & Menninghaus, W. (2021). Why and how should cognitive science care about aesthetics ? *Trends in Cognitive Sciences, xx*, 1–13. https://doi.org/10.1016/j.tics.2021.03.008

Winkler, A. M., Ridgway, G. R., Douaud, G., et al. (2016). Faster permutation inference in brain imaging. *NeuroImage, 141*, 502–516. https://doi.org/10.1016/j.neuroimage.2016.05.068

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*, 1100–1122. https://doi.org/10.1177/1745691617693393

Ying, T. C., Doraisamy, S., & Abdullah, L. N. (2012). Genre and mood classification using lyric features. In *Proceedings - 2012 international conference on information retrieval and knowledge management, CAMP'12* (pp. 260–263). https://doi.org/10.1109/InfRKM.2012.6204985

Zatorre, R. J., & Salimpoor, V. N. (2013). From perception to pleasure: Music and its neural substrates. *Proceedings of the National Academy of Sciences of the United States of America, 110*(Suppl), 10430–10437. https://doi.org/10.1073/pnas.1301228110

Zickfeld, J. H., Schubert, T. W., Seibt, B., et al. (2018). Kama Muta: Conceptualizing and measuring the experience often labelled being moved across 19 nations and 15 languages. *Emotion.*. https://doi.org/10.1037/emo0000450