The relationship between valence and chills in music: A corpus analysis[*]

Rémi de Fleurian[a]   Marcus T. Pearce[a,b]

[a]*Cognitive Science Research Group, Queen Mary University of London*
[b]*Center for Music in the Brain, Aarhus University & Royal Academy of Music*

Abstract

Chills experienced in response to music listening have been linked to both happiness and sadness expressed by music. To investigate these conflicting effects of valence on chills, we conducted a computational analysis on a corpus of 988 tracks previously reported to elicit chills, by comparing them with a control set of tracks matched by artist, duration, and popularity. We analysed track-level audio features obtained with the Spotify Web API across the two sets of tracks, resulting in confirmatory findings that tracks which cause chills were sadder than matched tracks, and exploratory findings that they were also slower, less intense, and more instrumental than matched tracks on average. We also found that the audio characteristics of chills tracks were related to the direction and magnitude of the difference in valence between the two sets of tracks. We discuss these results in light of the current literature on valence and chills in music, provide a new interpretation in terms of personality correlates of musical preference, and review the advantages and limitations of our computational approach.

Keywords: chills; music; valence

1   Introduction

Chills are a pleasurable psychophysiological response, often accompanied by piloerection, which can occur in response to music listening. They have been of great interest in scientific research on music for their involvement in emotional and aesthetic responses to music, as suggested by phenomenological reports of the experience and empirical findings about their physiological and neural correlates (for a review, see de Fleurian & Pearce, 2020). Particular attention has been given to their elicitors, resulting in the identification of a range of acoustic and musical features usually associated with the experience of chills, such as sudden dynamic changes, increased roughness, crescendi, or the entrance of new instruments (see de Fleurian & Pearce, 2020). These features represent

---

[*]Preprint submitted for peer review.

local musical and auditory events, and are therefore reflective of the fleeting nature of chills.

In contrast to continuous changes within musical stimuli, emotional characteristics of entire musical pieces have also been investigated. As a result, chills have been associated with perceived valence. However, there is disagreement about the direction of this relationship. While Grewe et al. (2011) reported an increase in frequency of chills for positively valenced music, Panksepp (1995) found that chills were more frequently associated with perceived sadness. In the latter study, however, both happy and sad music were found to elicit chills, reflecting subsequent findings that both emotions are linked with chills when they are expressed by music (Bannister, 2020; Mori & Iwanaga, 2017; Panksepp, 1995).

Conflicting effects of valence on chills have been discussed in the context of being moved, a mixed emotional state involving sadness and joy (Menninghaus et al., 2015). More specifically, being moved has been associated with chills when listening to music (Bannister, 2019, 2020; Bannister & Eerola, 2018; Benedek & Kaernbach, 2011), and has been found to mediate the relationship between liking and sadness in response to music (Vuoskoski & Eerola, 2017). Moving stimuli often feature narrative displays of social separation or reunion (Wassiliwizky et al., 2015), prosocial behaviour (Wassiliwizky et al., 2017), or self-sacrifice (Konečni et al., 2007), but it remains unclear how such narrative features translate to music, and how stimulus valence relates to the occurrence of chills. It could be that sad music provides an emotional context more conducive to the occurrence of chills (Panksepp, 1995). Another plausible explanation for the mixed effects of valence on chills comes from the possibility that chills encompass several phenomenologically distinct experiences, partly characterised by different degrees of felt emotions (Bannister, 2019; Maruskin et al., 2012).

In light of the current evidence, it remains difficult to establish the role of expressed stimulus valence on the incidence of chills in music. While behavioural approaches have contributed to identifying conflicting effects of happiness and sadness, they remain limited due to the number of stimuli which can be reasonably presented to participants, ranging here from 3 (Bannister & Eerola, 2018) to 23 (Grewe et al., 2011). Computational methods, however, can overcome such restrictions, and are well suited to the study of a large collection of naturalistic stimuli, at the cost of reduced control over experimental conditions. There has been, to our knowledge, no use of corpus-based analysis in research on chills, despite the success of similar approaches in research on music and emotion (e.g., Eerola, 2011). The analysis presented in this article is an attempt at addressing this gap in the literature, focusing on the effects of valence and other track-level audio features (i.e., features computed over entire musical pieces) on chills in music.

Specifically, we compare features between tracks known to elicit chills and a control set of tracks matched by artist, duration, and popularity. First, we conduct confirmatory analyses regarding the effect of the valence feature on chills, hypothesising a difference in expressed valence between the two sets of tracks. Then, we conduct two exploratory analyses, to investigate the influence

of other features on the occurrence of chills, and to assess whether these features influence the direction and magnitude of the difference in valence between both sets of tracks. We discuss these results and the advantages and limitations of our approach, and provide a new interpretation with reference to a theory of the personality correlates of musical preference (Rentfrow et al., 2011), identifying relationships between chills and several dimensions capturing musical preference (*sophisticated* music and *intense* music).

2 Materials

2.1 Dataset

For this study, we used *ChiM* (version 1.0.0, as included in de Fleurian & Pearce, 2020), a dataset prepared by compiling every mention of a piece of music reported to elicit chills in the literature reviewed in de Fleurian and Pearce (2020), whether these mentions consisted of author anecdotes, participant reports, empirical verifications, or discussion of prior results. This corresponds to 988 mentions of music confirmed to induce chills in at least one listener.

2.2 Features

Track-level audio features were collected using *spotifyr* (Thompson et al., 2020), an R package which enables pulling track information from Spotify's Web API (Spotify, n.d.). This allowed us to obtain, for most tracks, a range of features of interest, including *duration* (in milliseconds) and *popularity* (based on number and recency of plays), as well as nine track-level audio features: *acousticness* (confidence that the track is acoustic), *danceability* (based on tempo, rhythm stability, beat strength, and overall regularity), *energy* (based on dynamic range, perceived loudness, timbre, onset rate, and general entropy), *instrumentalness* (confidence that the track contains no vocal content), *liveness* (likelihood that the track was performed live), *loudness* (overall loudness in decibels), *speechiness* (presence of spoken words), *valence* (conveyed musical positiveness), and *tempo* (estimated in beats per minute). While Spotify does not share details about how these audio features are computed, they have successfully been used in previous research (e.g., Mas-Herrero et al., 2018; Melchiorre & Schedl, 2020).

3 Methods

3.1 Chills tracks

We removed 136 duplicated tracks from ChiM, before looking up track information by sending API queries for strings containing the artist (or arranger/interpreter, as indicated in ChiM) and title of each track. The top result for each query was retained, and used to pull the features described in Section 2.2. Throughout this process, an additional 103 tracks were removed due to unavailability on Spotify

3

or missing audio features, resulting in a dataset of 749 tracks which can cause chills.

3.2 Matched tracks

Our analysis aims to identify if specific track-level features are related to the occurrence of chills in ChiM. Therefore, a control set of tracks which do not cause chills was needed. Since it is impossible to assert that a specific track never causes chills for anyone, we approximated the construction of this control set. More specifically, we compared features across the chills tracks with features in another set of tracks, matched as fairly as possible with the chills tracks by artist, duration, and popularity. This strict matching procedure ensured there were as few differences as possible between both sets of tracks other than their potential to elicit chills. While it is possible that some matched tracks can elicit chills as well (see Section 5.2 for some limitations in our approach), it is unlikely that all of them do, meaning that any difference detected between the two sets of tracks can shed light on which factors affect the occurrence of chills.

First, we gathered potential matches by getting the first 50 tracks for each of the first 50 albums returned by an API query for each artist represented in the chills tracks. Then, we removed from these potential matches any track which was already present in the chills tracks, by comparing Spotify track IDs across the two sets. However, many duplicates remained across the two sets, since a piece of music on Spotify can have several distinct track IDs or slightly different titles. In order to mitigate this possibility, for each artist, we also removed from the potential matches any track with a title that had any number or any word of four letters or more (except the words *major* and *minor*) in common with the chills tracks. This process resulted in a pool of 205,717 potential matches.

Finally, we standardised track duration and popularity across the chills tracks and the potential matches. For each artist, the potential matches with the shortest Euclidean distance from each chills track for these two standardised features were selected as the best matches (see Figure 1A). Audio features were pulled for these 749 matched tracks, but were missing for 10 of them. A further 17 matched tracks were considered as outliers and removed for having Euclidean distances larger than three standard deviations from the mean, resulting in 722 pairs of chills and matched tracks (see Figure 1B).

3.3 Analysis

The confirmatory analysis consisted of assessing whether there is a difference in valence between chills tracks and matched tracks. We ran a logistic regression for the effect of the valence feature on track type (chills vs. matched). The presence of influential data points was checked with leave-one-out diagnostics, with the plan to run another logistic regression excluding data points that would, if left out, affect the slope by at least half of its original absolute value.

The exploratory analyses were twofold. First, we assessed whether there was a difference between chills tracks and matched tracks in the nine audio features
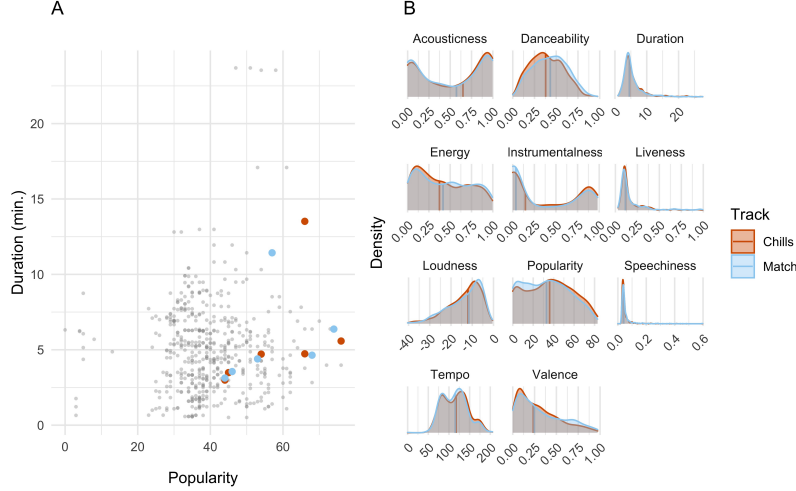
4

Figure 1: **A.** Example of the matching procedure, using Pink Floyd tracks. Tracks from ChiM are shown in orange, and potential matches gathered with the Spotify Web API are shown in grey. Potential matches with the shortest Euclidean distance from each chills track, in terms of duration and popularity, were selected as matched tracks, shown in blue. **B.** Densities and median values of audio features and metadata for the 722 resulting pairs of chills and matched tracks.

described in Section 2.2. Due to high collinearity between these features, we first ran a principal component analysis (PCA), with centering and scaling, resulting in two principal components with eigenvalues above one. We ran a logistic regression for the effects of these two principal components on track type (chills vs. matched), checking influential data points with leave-one-out diagnostics as described above.

Second, we assessed whether the audio features of the chills tracks had an effect on the *difference* in valence between chills tracks and matched tracks, which could mean that different types of chills arise in response to different auditory characteristics. As described above, we first ran a PCA due to collinearity, before running a linear regression for the effects of the two resulting principal components on the difference in valence between the chills tracks and the matched tracks, using leave-one-out diagnostics, and checking homoscedasticity and normality of residuals with residuals plots.

Finally, to check the robustness of the matching procedure, we conducted nonparametric paired t-tests to compare duration and popularity between chills tracks and matched tracks, expecting no differences in duration and popularity. Since there were some differences between both sets of tracks (see Section 4.2), we conducted mediation analyses, using the nonparametric bootstrap with 5000

Monte Carlo draws, as implemented in the *mediation* package (Tingley et al., 2014), to check if potential effects of the valence feature on track type were mediated by track duration and popularity. In addition, to mitigate this weakness of the matching procedure, all the analyses described above were replicated a total of 10 times, using a different set of matched tracks, each comprising one of the 10 tracks with the shortest Euclidean distance from each chills track (i.e., shortest Euclidean distance for iteration #1, second shortest for iteration #2, etc.)

4 Results

4.1 Effect of valence on track type

A logistic regression model yielded a significant fit ($\chi^2(1) = 6.33$, $p = .012$, Nagelkerke $R^2 = .006$), revealing a significant effect of the valence feature on track type ($b = 0.54$, $Z = 2.51$, $p = .012$), with the valence of chills tracks being lower than that of matched tracks by 0.033 on a 0–1 scale. This effect remained significant across all 10 iterations of the analysis (mean valence difference $= 0.042$, $SD = 0.009$). Results for the 10 iterations are shown in Table S1.

4.2 Mediating effects of duration and popularity

A Wilcoxon signed-rank test revealed no significant difference in duration ($V = 137199$, $p = .232$) and a significant difference in popularity ($V = 134593$, $p = .003$) between chills tracks and matched tracks (higher for chills tracks by 2.90 on a 1–100 scale), suggesting that the matching procedure did not result in an optimal set of matched tracks. The difference in popularity remained significant in all 10 iterations of the analysis, while the difference in duration became significant in the fourth as well as the last five iterations of the analysis, presumably due to the increasing Euclidean distance between chills tracks and each successive set of matched tracks (see Table S2).

To assess whether duration and popularity mediated the effect of the valence feature on track type as reported in Section 4.1, we conducted two separate causal mediation analyses. For duration, the average causal mediation effect was not significant ($ACME = -.006$, $p = .716$) and the average direct effect was significant ($ADE = -.128$, $p = .021$), suggesting that duration did not mediate the effect of valence on track type. For popularity, both average effects were significant ($ACME = .037$, $p = .002$; $ADE = -.171$, $p < .001$), suggesting that popularity partially, but not fully, mediated the effect of valence on track type. These results remained stable across the 10 iterations of the analysis, except for duration, which partially mediated the effect of valence on track type in the last iteration (see Table S3).

In this case, the mediation analyses each involved a linear regression (for the effect of valence on duration/popularity) and a logistic regression (for the effects of valence and duration/popularity on track type). It is worth noting

6

that for the linear models, some assumptions (homoscedasticity and normality of residuals) were violated. This is most likely due to the distribution of valence, duration, and popularity in our data (see Figure 1B). To confirm the results of the mediation analyses, we ran them again on a reduced (keeping only tracks with non-zero popularity due to this feature being zero-inflated) and transformed dataset (square root for popularity and log for valence and duration). These re-analyses did not fully eliminate the violations of assumptions for linear regression, but did replicate the findings presented above (see Table S4).

4.3 Effects of audio features on track type

We ran a PCA to reduce collinearity in the nine audio features. We retained two principal components with eigenvalues higher than one, accounting for 56.4% of cumulative proportion of variance explained. The first component featured high positive loadings (greater than .2) for energy, loudness, valence, danceability, and tempo, and high negative loadings (lower than -.2) for acousticness and instrumentalness. The second component featured high positive loadings for liveness and speechiness, and a high negative loading for danceability. The number of retained principal components and their associated loadings were consistent across all 10 iterations of the analysis (besides occasional but systematic sign differences, which are expected when conducting several PCAs—see Table S5).

A logistic regression model yielded a significant fit ($\chi^2(2) = 6.47$, $p = .039$, Nagelkerke $R^2 = .006$), revealing a significant effect of the first component on track type ($b = 0.06$, $Z = 2.34$, $p = .019$) and no significant effect of the second component ($b = 0.05$, $Z = 0.98$, $p = .328$), showing that chills tracks had lower scores than matched tracks on the first component (i.e., low energy, loudness, valence, danceability, and tempo, as well as high acousticness and instrumentalness—see Figure 2). The model fit remained significant in all but one iteration of the analysis ($\chi^2(2) = 5.33$, $p = .070$, Nagelkerke $R^2 = .005$), the effect of the first component remained significant in all iterations, and the effect of the second component became significant in four iterations, highlighting that in some cases, chills tracks had lower scores than matched tracks on the second component (i.e., low liveness and speechiness, as well as high danceability). It is worth noting that in the seventh iteration of the analysis, there were two influential data points (as described in Section 3.3). For this iteration, we ran the model both with and without the influential data points, leading to similar results in both cases (see Table S6).

4.4 Effects of audio features on difference in valence between track types

We ran a PCA on the nine audio features of the chills tracks only (as opposed to both track types in Section 4.3), to assess if properties of the chills track can predict the direction and magnitude of the difference in the valence feature between both track types. We retained two principal components with eigenvalues higher than one, accounting for 55.6% of cumulative proportion of variance
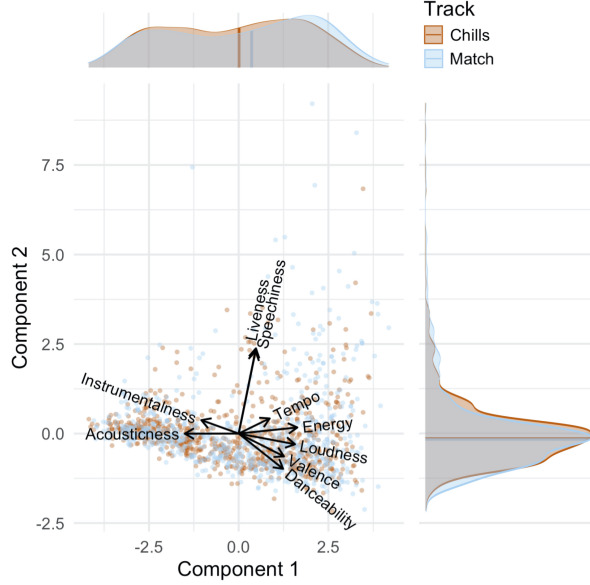
Figure 2: Biplot of chills and matched tracks for the first iteration of the analysis. Tracks are mapped onto the first two components obtained with PCA. Densities and median values for chills and matched tracks are shown in marginal plots, revealing a difference on Component 1. Audio feature loadings are shown as vectors, illustrating the high degree of collinearity between some features.

explained. Both components featured similar loadings as in the previous section. The number of retained principal components and their associated loadings were consistent across all 10 iterations of the analysis (see Table S7).

A multiple linear regression model yielded a significant fit ($F(2, 719) = 63.9$, $p < .001$, adjusted $R^2 = .149$), revealing a significant effect for both the first component ($\beta = .062$, $p < .001$) and the second component ($\beta = .039$, $p < .001$), suggesting that chills tracks were progressively happier than matched tracks as values on these two components increased, and vice versa. These effects held for all 10 iterations of the analysis (see Table S8).

5 Discussion

5.1 Results

In this experiment, we compared track-level audio features between tracks taken from ChiM, a dataset of pieces of music known to elicit chills, and several sets of tracks algorithmically matched by artist, duration, and popularity.

We compared the valence feature between chills tracks and matched tracks, and found that chills tracks were, on average, slightly lower in valence. This

echoes previous findings that chills are more frequently associated with perceived sadness (Panksepp, 1995), as opposed to perceived happiness (Grewe et al., 2011). The matching procedure resulted in a small difference in valence between track types, but it is worth noting that overall, the distribution of valence in ChiM is highly positively skewed, whereas it is relatively uniform across all tracks on Spotify (Spotify, n.d.). In other words, an effect of valence was identified despite the application of a strict matching procedure, which most likely resulted in high similarity between chills tracks and matched tracks. If control tracks had been selected randomly instead, most chills tracks would have had a much lower valence by comparison.

When taking all audio features into consideration, we found that chills tracks were characterised by smaller values on a component linked with high energy, loudness, valence, danceability, and tempo, as well as low acousticness and instrumentalness, meaning that overall, chills track were sadder, slower, less intense, and more instrumental than matched tracks. In a few occasions, chills tracks were also characterised by smaller values on a second component linked with high liveness and speechiness, as well as low danceability, therefore suggesting that chills-inducing music may be less likely to include spoken words and to feature a live audience, although these results are less robust than those for the first component. These findings can be interpreted with reference to an influential theory of the personality correlates of musical preference (Rentfrow et al., 2011). The musical characteristics we identified strongly match *sophisticated* music, which tends to be relaxing, quiet, non-danceable, slow, non-electric, and instrumental (Rentfrow et al., 2012), suggesting that chills tracks were more sophisticated than matched tracks in our analysis. Interestingly, preference for sophisticated music is associated with openness to experience (Schäfer & Mehlhorn, 2017), a personality characteristic strongly linked to the experience of chills (see de Fleurian & Pearce, 2020).

We also examined whether the audio features of chills tracks relate to the difference in valence between chills tracks and matched tracks. We found that tracks with higher energy, loudness, valence, liveness, and speechiness, as well as lower acousticness and instrumentalness, were more likely to be happier than their associated matched tracks, and vice versa. While these results are partly explained by valence loading on the first component obtained with PCA, the involvement of other audio features suggests a potential interpretation. Using the same classification as above (Rentfrow et al., 2012), it becomes apparent that, on average, *sophisticated* chills tracks were sadder than matched tracks, and *intense* chills tracks (i.e., non-relaxing, loud, electric, and featuring raspy or yelling voice) were happier than matched tracks. As is the case with sophisticated music, intense music is linked with openness to experience (Schäfer & Mehlhorn, 2017), a known personality correlate of chills. Interestingly, these results provide some support for the possibility that different types of chills are elicited by different types of feelings and affective states expressed or evoked by music (Bannister, 2019; Maruskin et al., 2012).

5.2 Limitations

We discussed the rationale and some limitations of the matching procedure in Section 3.2. Notably, despite our best efforts, matched tracks were slightly less popular than chills tracks. This could be due to a bias towards popular tracks when reporting music which causes chills. It should be possible to reduce this difference by picking matched tracks from a larger set of potential matches, but our current methods did not allow this due to limits on the rate of Spotify API requests and the number of records per request. As a result, we found that popularity partially, but not fully, mediated the difference in valence between chills tracks and matched tracks. This could be interpreted as sad songs being more popular, and in turn, popular songs being more likely to cause chills. However, the mediation analyses were not part of the planned primary analysis, but rather only intended as a procedural check, and we suspect the identified mediating effects are largely due to the idiosyncrasies of our data. Nonetheless, we attempted to address any limitations in the matching procedure by conducting 10 iterations of the full analysis with different sets of matched tracks, which led to consistent results across iterations.

There are other limitations to our approach. ChiM does not report the exact version of the pieces of music that elicit chills, which could have had some impact on the audio characteristics of the chills tracks. Then, some chills tracks might have been present in the matched set, despite our efforts to limit this possibility (see Section 3.2). In general, apart from a few sanity checks, we considered the lack of manual verification of our data as an acceptable trade-off for the large size of the dataset, which made a robust computational analysis possible. Another issue is the lack of transparency about how Spotify computes audio features. Again, we accepted this trade-off which allowed us to collect large amounts of audio data and metadata through API queries.

More importantly, effect sizes were small for most of our results. One possible explanation is that we tried to be as fair as possible with the matching procedure, which might have drastically reduced the differences in audio features between chills tracks and matched tracks—effect sizes would probably have been more pronounced if we randomly selected matched tracks, but this process would have introduced noise due to confounding differences between both track types. Also, as highlighted in Section 3.2, it is possible that some matched tracks also had the ability to elicit chills. Finally, chills are a localised phenomenon, and it is fully expected that track-level features would not capture local changes in acoustic and structural features, therefore limiting the explanatory power of our approach. However, we believe that the consistency of the results across several iterations of the analysis on different sets of matched tracks yielded robust and interpretable findings, despite their small effect size.

5.3 Conclusion

We conducted a corpus analysis of audio characteristics of music known to elicit chills and identified that such music is sadder than music matched by artist,

duration, and popularity. Exploratory analyses revealed that chills tracks are also slower, less intense, and more instrumental than matched tracks on average, and tend to possess the characteristics of sophisticated music. Moreover, within chills tracks, we identified a possible relationship between valence and type of music, with sophisticated music tending to be sadder than matched tracks, and intense music tending to be happier. These results show that, for research about chills in music, computational methods have a great and largely untapped potential to complement empirical studies.

Data availability

The data pulled from the Spotify Web API is proprietary, and can therefore not be published alongside this article. However, we have made the R script used to obtain and analyse this data available on GitHub (https://doi.org/10.5281/zenodo.4161610).

Acknowledgements

Funding

Declaration of interest

No potential conflict of interest was reported by the authors.

References

Bannister, S. (2019). Distinct varieties of aesthetic chills in response to multimedia. *PLOS ONE*, *14*(11), e0224974. https://doi.org/10.1371/journal.pone.0224974

Bannister, S. (2020). A survey into the experience of musically induced chills: Emotions, situations and music. *Psychology of Music*, *48*(2), 297–314. https://doi.org/10.1177/0305735618798024

Bannister, S., & Eerola, T. (2018). Suppressing the chills: Effects of musical manipulation on the chills response. *Frontiers in Psychology*, *9*, 2046. https://doi.org/10.3389/fpsyg.2018.02046

Benedek, M., & Kaernbach, C. (2011). Physiological correlates and emotional specificity of human piloerection. *Biological Psychology*, *86*(3), 320–329. https://doi.org/10.1016/j.biopsycho.2010.12.012

de Fleurian, R., & Pearce, M. T. (2020). Chills in music: An integrative review. *PsyArXiv*. https://doi.org/10.31234/osf.io/yc6d8

Eerola, T. (2011). Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres. *Journal of New Music Research*, *40*(4), 349–366. https://doi.org/10.1080/09298215.2011.602195

Grewe, O., Katzur, B., Kopiez, R., & Altenmüller, E. (2011). Chills in different sensory domains: Frisson elicited by acoustical, visual, tactile and gustatory stimuli. *Psychology of Music*, *39*(2), 220–239. https://doi.org/10.1177/0305735610362950

Konečni, V. J., Wanic, R. A., & Brown, A. (2007). Emotional and aesthetic antecedents and consequences of music-induced thrills. *American Journal of Psychology*, *120*(4), 619–643. https://doi.org/10.2307/20445428

Maruskin, L. A., Thrash, T. M., & Elliot, A. J. (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*(1), 135–157. https://doi.org/10.1037/a0028117

Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2018). Modulating musical reward sensitivity up and down with transcranial magnetic stimulation. *Nature Human Behaviour*, *2*, 27–32. https://doi.org/10.1038/s41562-017-0241-z

Melchiorre, A. B., & Schedl, M. (2020). Personality correlates of music audio preferences for modelling music listeners. In T. Kuflik & I. Torre (Eds.), *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '12* (pp. 313–317). Association for Computing Machinery. https://doi.org/10.1145/3340631.3394874

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Kuehnast, M., & Jacobsen, T. (2015). Towards a psychological construct of being moved. *PLOS ONE*, *10*(6), e0128451. https://doi.org/10.1371/journal.pone.0128451

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

Panksepp, J. (1995). The emotional sources of "chills" induced by music. *Music Perception*, *13*(2), 171–207. https://doi.org/10.2307/40285693

Rentfrow, P. J., Goldberg, L. R., & Levitin, D. J. (2011). The structure of musical preferences: A five-factor model. *Journal of Personality and Social Psychology*, *100*(6), 1139–1157. https://doi.org/10.1037/a0022406

Rentfrow, P. J., Goldberg, L. R., Stillwell, D. J., Kosinski, M., Gosling, S. D., & Levitin, D. J. (2012). The song remains the same: A replication and extension of the MUSIC model. *Music Perception*, *30*(2), 161–185. https://doi.org/10.1525/mp.2012.30.2.161

Schäfer, T., & Mehlhorn, C. (2017). Can personality traits predict musical style preferences? A meta-analysis. *Personality and Individual Differences*, *116*, 265–273. https://doi.org/10.1016/j.paid.2017.04.061

Spotify. (n.d.). *Web API: Spotify for developers.* https://developer.spotify.com/documentation/web-api/

Thompson, C., Parry, J., Phipps, D., & Wolff, T. (2020). *spotifyr: R wrapper for the 'Spotify' Web API (R package version 2.1.1.)* https://github.com/charlie86/spotifyr

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1–38. https://doi.org/10.18637/jss.v059.i05

Vuoskoski, J. K., & Eerola, T. (2017). The pleasure evoked by sad music is mediated by feelings of being moved. *Frontiers in Psychology*, *8*, 439. https://doi.org/10.3389/fpsyg.2017.00439

Wassiliwizky, E., Jacobsen, T., Heinrich, J., Schneiderbauer, M., & Menninghaus, W. (2017). Tears falling on goosebumps: Co-occurence of 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., Wagner, V., Jacobsen, T., & Menninghaus, W. (2015). Art-elicited chills indicate states of being moved. *Psychology of Aesthetics, Creativity, and the Arts*, *9*(4), 405–416. https://doi.org/10.1037/aca0000023

Supplementary materials

Table S1: Effect of valence on track type for all 10 iterations of the analysis.

Iteration	Model fit χ^2	p	Nagelkerke R^2	Valence b	Z	p
1	6.33	.012	.006	0.54	2.51	.012
2	12.04	< .001	.011	0.75	3.45	< .001
3	8.99	.003	.008	0.65	2.99	.003
4	5.13	.023	.005	0.51	2.26	.024
5	13.13	< .001	.012	0.80	3.60	< .001
6	6.42	.011	.006	0.56	2.53	.012
7	11.72	< .001	.011	0.75	3.40	< .001
8	15.46	< .001	.014	0.83	3.91	< .001
9	15.88	< .001	.015	0.86	3.96	< .001
10	14.98	< .001	.014	0.84	3.84	< .001

Table S2: Difference in duration and popularity between chills tracks and matched tracks.

Iteration	Duration		Popularity	
	V	p	V	p
1	137199	.232	134593	.003
2	142583	.051	145166	$< .001$
3	142185	.060	150597	$< .001$
4	146264	.005	152310	$< .001$
5	140323	.059	154629	$< .001$
6	141824	.026	159205	$< .001$
7	145974	.005	159076	$< .001$
8	144000	.013	162740	$< .001$
9	143062	.010	163894	$< .001$
10	144711	.004	165209	$< .001$

Table S3: Mediating effects of duration and popularity for the effect of valence on track type.

Iteration	Duration				Popularity			
	$ACME$	p	ADE	p	$ACME$	p	ADE	p
1	-.006	.716	-.128	.021	.037	.002	-.171	< .001
2	-.010	.503	-.173	.002	.049	< .001	-.232	< .001
3	-.012	.434	-.147	.004	.063	< .001	-.223	.001
4	-.028	.103	-.100	.095	.067	< .001	-.196	< .001
5	-.012	.419	-.184	< .001	.073	< .001	-.270	< .001
6	-.020	.232	-.118	.034	.072	< .001	-.209	< .001
7	-.019	.209	-.164	.002	.073	< .001	-.256	< .001
8	-.016	.267	-.189	< .001	.075	< .001	-.281	< .001
9	-.022	.145	-.188	.001	.073	< .001	-.285	< .001
10	-.032	.046	-.174	.002	.081	< .001	-.288	< .001

Table S4: Re-analyses of mediating effects of duration and popularity for the effect of valence on track type.

Iteration	Duration				Popularity			
	$ACME$	p	ADE	p	$ACME$	p	ADE	p
1	-.002	.656	-.023	.109	.010	.001	-.036	.010
2	-.005	.304	-.034	.024	.013	$< .001$	-.052	$< .001$
3	-.006	.136	-.019	.174	.017	$< .001$	-.043	.002
4	-.009	.086	-.024	.102	.020	$< .001$	-.052	$< .001$
5	-.006	.217	-.030	.032	.020	$< .001$	-.056	$< .001$
6	-.007	.094	-.017	.242	.021	$< .001$	-.045	$< .001$
7	-.008	.051	-.028	.058	.023	$< .001$	-.059	$< .001$
8	-.005	.198	-.028	.049	.023	$< .001$	-.056	$< .001$
9	-.008	.070	-.036	.014	.021	$< .001$	-.065	$< .001$
10	-.010	.026	-.036	.016	.026	$< .001$	-.072	$< .001$

Table S5: Principal component analysis on audio features for all tracks.

| PC | Iteration | **Audio feature loadings** | | | | | | | | |
		Tempo	Loudness	Valence	Dance.	Energy	Acoust.	Instru.	Speech.	Liveness
1	1	.242	.439	.353	.346	.457	-.419	-.291	.132	.134
	2	.222	.444	.356	.346	.459	-.422	-.298	.128	.118
	3	.233	.438	.350	.343	.458	-.421	-.289	.151	.142
	4	.220	.447	.351	.341	.461	-.428	-.300	.112	.117
	5	.221	.442	.356	.343	.461	-.422	-.297	.139	.117
	6	.243	.439	.354	.338	.456	-.420	-.292	.153	.124
	7	.231	.439	.346	.341	.459	-.425	-.301	.151	.115
	8	.241	.437	.352	.348	.453	-.419	-.302	.135	.125
	9	.220	.441	.356	.348	.456	-.421	-.293	.151	.124
	10	.241	.438	.351	.351	.456	-.422	-.292	.130	.122
2	1	.119	-.083	-.178	-.275	.048	-.001	.105	.645	.665
	2	.158	.078	.080	.148	.003	-.071	.046	-.660	-.706
	3	.138	-.029	-.293	-.365	.081	-.051	.098	.548	.665
	4	.075	-.052	-.240	-.300	.062	-.055	.020	.581	.706
	5	.050	.019	.219	.324	-.063	.029	.013	-.567	-.719
	6	.007	-.020	-.261	-.359	.076	-.048	.007	.498	.739
	7	.043	-.010	-.276	-.363	.048	-.038	-.045	.543	.699
	8	.037	.078	.150	.202	.017	-.046	.056	-.644	-.714
	9	.052	-.075	-.173	-.267	.020	.012	.035	.612	.717
	10	.088	.100	.101	.162	.007	-.051	.033	-.654	-.717

Note. PC = Principal component, Dance. = Danceability, Acoust. = Acousticness, Instru. = Instrumentalness, Speech. = Speechiness.

Table S6: Effect of first two principal components on track type.

Iteration	Model fit			Component 1			Component 2		
	χ^2	p	Nagelkerke R^2	b	Z	p	b	Z	p
1	6.47	.039	.006	0.06	2.34	.019	0.05	0.98	.328
2	10.39	.006	.010	0.07	2.42	.016	-0.11	-2.10	.036
3	6.57	.038	.006	0.07	2.51	.012	0.03	0.51	.612
4	13.04	.001	.012	0.07	2.63	.008	0.12	2.44	.015
5	9.00	.011	.008	0.07	2.76	.006	-0.06	-1.17	.243
6	5.33	.070	.005	0.06	2.15	.032	0.04	0.83	.404
7a	6.71	.035	.006	0.07	2.56	.010	0.02	0.37	.714
7b	6.96	.031	.006	0.07	2.61	.009	0.02	0.40	.687
8	13.28	.001	.012	0.07	2.74	.006	-0.12	-2.37	.018
9	7.20	.027	.008	0.07	2.61	.009	0.03	0.60	.551
10	12.53	.002	.012	0.07	2.69	.007	-0.12	-2.26	.024

Note. The analysis for iteration 7 was conducted with (7a) and without (7b) influential data points.

Table S7: Principal component analysis on audio features for chills tracks only.

PC	Iteration	Audio feature loadings								
		Tempo	Loudness	Valence	Dance.	Energy	Acoust.	Instru.	Speech.	Liveness
1	1	.251	.440	.345	.339	.463	-.419	-.283	.155	.126
	2	.249	.440	.346	.336	.463	-.420	-.283	.158	.127
	3	.249	.440	.347	.335	.463	-.421	-.283	.157	.125
	4	.247	.441	.346	.334	.464	-.421	-.282	.153	.129
	5	.249	.440	.346	.337	.463	-.420	-.281	.156	.129
	6	.253	.439	.344	.337	.463	-.420	-.282	.157	.125
	7	.250	.440	.346	.335	.463	-.421	-.282	.158	.126
	8	.250	.440	.346	.335	.463	-.421	-.282	.158	.126
	9	.249	.439	.346	.334	.463	-.421	-.285	.159	.126
	10	.250	.438	.345	.338	.462	-.420	-.286	.155	.127
2	1	.014	.088	.149	.265	-.020	-.007	-.037	-.650	-.690
	2	.008	.083	.164	.283	-.030	.004	-.049	-.646	-.682
	3	.009	.081	.164	.281	-.029	.004	-.047	-.646	-.684
	4	.005	.084	.162	.284	-.030	.005	-.049	-.654	-.674
	5	.025	.086	.148	.273	-.024	-.005	-.040	-.648	-.688
	6	.016	.086	.157	.269	-.024	-.004	-.040	-.647	-.689
	7	.015	.088	.152	.274	-.025	-.001	-.047	-.646	-.688
	8	.015	.088	.152	.274	-.025	-.001	-.047	-.646	-.688
	9	.009	.084	.162	.277	-.028	.000	-.052	-.646	-.685
	10	.011	.082	.160	.277	-.031	.000	-.048	-.650	-.682

Note. PC = Principal component, Dance. = Danceability, Acoust. = Acousticness, Instru. = Instrumentalness, Speech. = Speechiness.

Table S8: Effect of first two principal components on difference in valence between track types.

Iteration	Model fit			Component 1		Component 2	
	F	p	Adjusted R^2	β	p	β	p
1	63.9	$< .001$	.149	0.062	$< .001$	0.039	$< .001$
2	62.5	$< .001$	.145	0.059	$< .001$	0.040	$< .001$
3	65.2	$< .001$	.151	0.063	$< .001$	0.033	.002
4	64.8	$< .001$	.150	0.059	$< .001$	0.031	.002
5	55.7	$< .001$	.132	0.057	$< .001$	0.028	.007
6	47.5	$< .001$	.115	0.052	$< .001$	0.026	.012
7	51.8	$< .001$	.124	0.055	$< .001$	0.033	.002
8	51.4	$< .001$	.123	0.057	$< .001$	0.035	.002
9	51.7	$< .001$	.124	0.058	$< .001$	0.033	.003
10	55.2	$< .001$	.131	0.058	$< .001$	0.035	.001