

# Assessing aesthetic music-evoked emotions in a minute or less: A comparison of the GEMS-45 and the GEMS-9

Musicae Scientiae

1–9

© The Author(s) 2024



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/10298649241256252

[journals.sagepub.com/home/msx](https://journals.sagepub.com/home/msx)**Peer-Ole Jacobsen\*** 

Universität Innsbruck, Austria

**Hannah Strauss\***

Universität Innsbruck, Austria

**Julia Vigl**

Universität Innsbruck, Austria

**Eva Zangerle**

Universität Innsbruck, Austria

**Marcel Zentner** 

Universität Innsbruck, Austria

## Abstract

The Geneva Emotional Music Scale (GEMS) is a widely used instrument to measure emotions evoked by music. Its original version includes 45 emotion-related adjectives that can be grouped into nine dimensions and three second-order factors. Because time is often critical, the same authors introduced a checklist that assesses each dimension with one item only (GEMS-9). The checklist is being increasingly used, but it remains at present unclear whether the two instruments produce comparable scores. To redress this gap, we had 192 participants rate 18 music excerpts from various music genres with both instruments. We found that although scores on the nine GEMS emotions did converge in terms of profile similarity, the GEMS-9 tended to produce somewhat higher absolute scores. Yet, when dimensions of the GEMS-45 were represented by their highest-scoring scale item, the absolute scores were consistent as well. We conclude that if researchers have time constraints but still wish to capture some of the distinct features of music-evoked emotion, the GEMS-9 provides an interesting alternative to the GEMS-45.

## Keywords

music, emotion, measures, GEMS, profile similarity

---

\*These two authors contributed equally to the work.

## Corresponding author:

Marcel Zentner, Department of Psychology, University of Innsbruck, Universitätsstrasse 15, 6020 Innsbruck, Austria.

Email: [Marcel.Zentner@uibk.ac.at](mailto:Marcel.Zentner@uibk.ac.at)

Over the past two decades, music-evoked emotions have become an important area of study across disciplines, from anthropology, neuroscience, and psychology to domains that use signal-processing and machine-learning methods to characterize or predict musical emotions (e.g., Gómez-Cañón et al., 2021; Han et al., 2022; Juslin, 2019; Koelsch, 2018). The availability of tools to accurately assess music-evoked emotions is therefore of importance to a growing number of research communities. Initially, research on music and emotions relied on domain-unspecific emotion models, such as the affective circumplex (e.g., Russell, 1980), or basic emotions model (e.g., Izard, 2007). Later, it was found that music-evoked emotions have distinctive features that are best captured by a domain-specific approach (Zentner et al., 2008).

Specifically, Zentner et al. (2008) started with an initial set of 515 affect terms and successively eliminated those terms that were rarely used to describe music-evoked emotions, retaining a core set of 45 musically significant emotion terms. The authors also found that a hierarchical structure underlies this set of items, comprising three second-order and nine first-order factors: (1) Sublimity (wonder, transcendence, tenderness, nostalgia, and peacefulness); (2) Vitality (joyful activation and power); and (3) Unease (tension and sadness). This structure is sometimes referred to as GEMS model, in accordance with the name of the scale used to measure emotions predicated by the model (see below). In recent years, researchers have independently discerned a factorial structure of music-evoked emotion similar to the GEMS model (Chelkowska-Zacharewicz & Janowski, 2021) and recognized the importance of several of the components of the model for characterizing musically evoked emotions, such as wonder-awe, tenderness, and nostalgia (e.g., Barrett et al., 2010; Juslin, 2013).

To assess music-evoked emotions along these lines, the authors devised the Geneva Emotional Music Scale (GEMS). It allows the nine components of the model to be assessed using 45 items. Although the scale is well-suited for an in-depth assessment of music-evoked emotion, tasking participants with rating 45 items is not always feasible and is particularly challenging when time is tight, such as in neuroimaging studies. Recent years have also seen the advent of music databases that incorporate information about the emotional effects of songs (e.g., Chen et al., 2015; Soleymani et al., 2013; Zhang et al., 2018). While a few hundred music excerpts can be rated using the GEMS when time and resources are adequate (see Strauss et al., 2024), obtaining similarly detailed ratings of thousands of music excerpts is likely to overtax the resources available to most researchers.

To counteract this practical limitation, a brief checklist measuring each of the nine dimensions of the GEMS with a single item can be used (see Zentner et al., 2008, Study 4). Due to its brevity and ease of administration, this checklist (henceforth GEMS-9) is being increasingly used in work characterizing or classifying music-evoked emotion (e.g., Hashim et al., 2020; Kaelen et al., 2015; Pearce & Halpern, 2015). A key difference between the instruments is that, whereas with the GEMS-9, each dimension is assessed directly, scale scores for the full-length version of the GEMS (henceforth GEMS-45) are derived from aggregating several items belonging to a given dimension. Thus, although both instruments are derived from the same model, it is not a foregone conclusion that the GEMS-9 provides a truthful approximation to the scale scores provided by the GEMS-45.

To elucidate this question, the objective of this study was to examine whether the GEMS-45 and the GEMS-9 produce similar average profiles for the same music excerpts. We also examined levels of interrater agreement for the GEMS-45 and the GEMS-9 since eventual differences in interrater reliability between the two instruments could bias the findings. To this end, we had participants rate 18 music excerpts from various music genres and different expressive characteristics using both instruments. The analyses focus on a comparison of the emotion profiles

obtained for the music excerpts via the GEMS-45 and the GEMS-9 using standard and alternative scoring procedures.

## Methods

### Sample

A total of 192 participants aged a mean of 27.5 years ( $SD = 10.3$ ) took part in the study (62.5 female, 37% male, 0.5% non-binary). The majority of participants were university/college students or graduates ( $n = 170$ , 88.5%). Of the 192 participants, 105 (54.7%) were German-speaking, and 87 (45.3%) English-speaking. The former were recruited via a mailing list from the University of Innsbruck; the latter were recruited through the crowd-sourcing tool Prolific (prolific.com). Psychology students at the University of Innsbruck ( $n = 91$ , 47.4%) received course credits, whereas participants recruited via Prolific ( $n = 87$ , 45.3%) were paid £11 for their participation.

### Stimuli

Music excerpts were drawn from the Emotion-to-Music Mapping Atlas (EMMA; [musemap-tools.uibk.ac.at/emma](http://musemap-tools.uibk.ac.at/emma)). EMMA is a new online database that comprises GEMS ratings for 817 music excerpts from 7 music genres, of which 364 excerpts from the genres hip-hop/rap, pop, and classical were rated with the full GEMS-45 (Strauss et al., 2024). For this study, we selected excerpts that were rated similarly by different raters, exhibited good levels of interrater agreement ( $ICC > .75$ ) and that varied in terms of subgenre, instrumentation, and tempo. The final set of excerpts comprised 6 music excerpts from each of the 3 genres, resulting in a total of 18 excerpts, with an average duration of 46 s ( $SD = 10$ ; range: 22–60 s). While all the pop excerpts comprised vocals, there were two fully instrumental excerpts for hip-hop/rap and five for classical music. Detailed information about the excerpts is reported in Supplemental Table S1.

### Measures

**Socio-demographic Information.** Participants were asked to indicate their gender, age, educational attainment, and fluency in the study language.

**GEMS.** Music excerpts were rated using the GEMS-45, and the GEMS-9 (Zentner et al., 2008). The GEMS-45 comprises 45 items that assess 9 musically relevant emotion dimensions, each scale containing between 3 and 6 items.<sup>1</sup> In contrast, the GEMS-9 measures each of the nine dimensions with one single item, named after the label of the dimension (see Supplemental Table S2). Although the original answer format consists of a 5-point Likert-type scale, we used visual analog scales ranging from 0 to 100 in this study. We did so to allow participants to characterize their emotional experiences with greater nuance, a practice that is frequently recommended for affect ratings (Zentner & Eerola, 2010).

### Procedure and scoring

**Procedure.** Participants completed the ratings remotely online via LimeSurvey (v. 2.64.1, LimeSurvey GmbH, n.d.). On the landing page, participants were provided with information regarding the content and duration of the study as well as relevant compensation. Participants

rated the music excerpts using the GEMS-45 and the GEMS-9 in two separate sessions. They were randomly assigned to either the GEMS-45 or the GEMS-9 condition and were instructed to complete the other version of the GEMS 1–7 days later. The average time interval between the sessions was 3.07 ( $SD = 3.12$ ,  $Mdn = 2$ , range: 0–21) days.

In the GEMS-45 condition, participants were first introduced to the GEMS. This introduction included the presentation of all 45 emotion terms, each term being illustrated by an information icon that, if activated, displayed synonyms to facilitate the understanding of the terms. The order of terms was randomly allocated to participants but remained unchanged throughout the study for any individual participant.

In terms of scoring procedures, there are two procedures for administering and scoring the GEMS-45. The first is *Scan-select*, whereby participants are instructed to scan all terms and rate only those that match the experienced emotions (see Zentner et al., 2008, Study 3). The second is *All-item*, whereby participants are instructed to rate all items (see Zentner et al., 2008, Study 4). Because of the relatively large number of excerpts to be rated, we used the scan-select procedure in this study. Thus, the instructions specified that participants should listen carefully to a music excerpt and then select the terms matching their emotional experience. The selected terms reappeared in a different location of the page when participants were played the same excerpt for a second time. This time participants were instructed to rate the intensity of the selected terms using a slider, on a visual analog scale ranging from 1 (i.e., indicating a very weak emotional response) to 100 (i.e., indicating a very intense emotional response). Because participants were instructed to deliberately discard emotions they had not experienced from their ratings, unselected emotions were accorded a value of 0 (see Zentner et al., 2008, Study 3).

For the GEMS-9, the same rating instruction was used but the rating procedure differed somewhat. Specifically, all nine items representing the nine GEMS dimensions were displayed next to the same visual analog slider as in the GEMS-45 rating. The default value was set to 0 and participants instructed to change the slider position only for those items that reflected an experienced emotion.

**Scoring.** Scale scores for the GEMS-45 are derived from aggregating three to six items belonging to a given dimension (Zentner et al., 2008). As noted above, in the scan-select procedure used here, discarded items are awarded a value of 0. This could lower scale scores relative to those obtained with the GEMS-9 if a simple scale average is used, since it is far less likely for items on the GEMS-9 to be discarded because there are so few of them. To account for this possibility, the scores for the nine dimensions of the GEMS-45 were computed in two different ways:

1. Using *maximum-value scoring*, whereby the item with the highest value of a given scale is the scale score, effectively acting as an ambassador, so to speak, for the entire scale. The reasoning underlying this approach is that, when faced with several related terms, listeners will tend to choose the term that best matches their emotional experience rather than similarly relevant, but less-suitable terms. To the extent that this assumption is correct, unselected scale items do not necessarily reflect an absence of the respective states, but rather a commitment to the most suitable term for the experienced emotion.
2. Using *weighted-mean scoring*, which accounts for both the number of chosen emotion terms and their intensity using the equation displayed in Figure 1. This formula was introduced to overcome the limitations of calculating a simple average across all scale

$$\left( \bar{x} + \frac{\bar{x} \times \sum s}{N} \right) / 2$$

**Figure 1.** Equation used to compute weighted-mean scores for the GEMS-45.

Note.  $\bar{x}$  = mean intensity of the emotion dimension or in total,  $\sum s$  = number of selected emotions, and  $N$  = number of emotions available for the emotion dimension or in total.

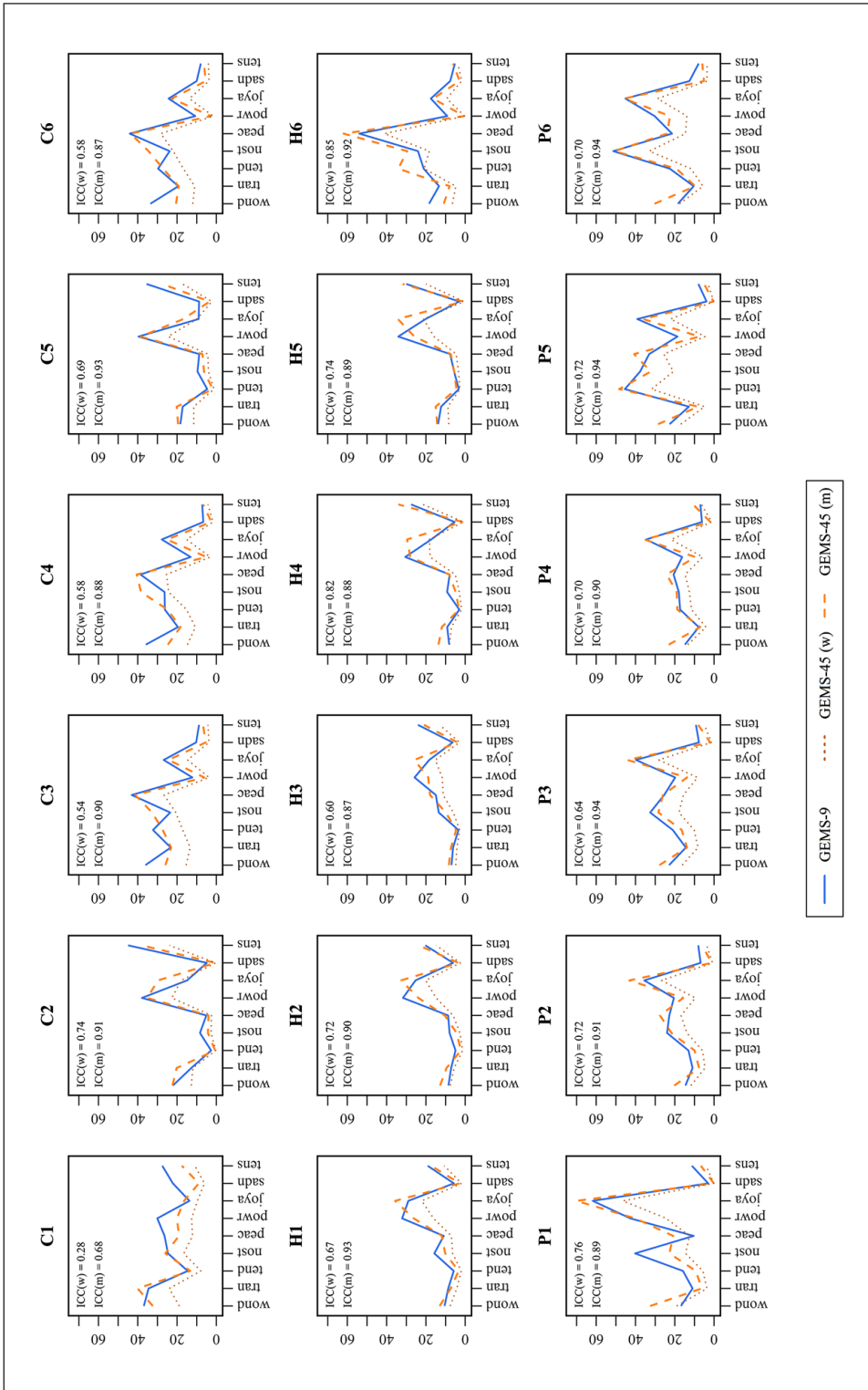
items, which are best illustrated by an example: Respondent 1 selects five out of six items of the Wonder scale but gives all of them a low-intensity rating (e.g., 10), while Respondent 2 selects only one Wonder item but gives it a higher intensity rating (e.g., 50). In this scenario, a simple average across all scale items would yield a score of 8.3 for both respondents. This calculation would give too much weight to the number of selected emotion terms at the expense of intensity. If, by contrast, the average were only computed across the selected items, Respondent 1 would be awarded a score of 10, while Respondent 2 would be awarded a score of 50, thus failing to take the number of selected emotion terms sufficiently into account. Using the weighted-mean formula would produce a score of 9.17 for Respondent 1 and a score of 29.17 for Respondent 2, taking both the number of selected emotions and their respective intensities into proportionate account (see Gerstgrasser et al., 2023, for more details).

## Results

Descriptive statistics for both GEMS-45 and GEMS-9, including single-item statistics for the GEMS-45, are shown in Supplemental Tables S3, S4, and Figure S1. To ensure that the results for concordances between GEMS-45 and GEMS-9 were not biased by eventual differences in the reliability of the instruments, we first examined levels of interrater agreement for both instruments. Specifically, we examined agreement in GEMS-9 and GEMS-45 profiles across participants for each music excerpt separately. To this end, we used the two-way mixed-effect consistency model, or ICC(C,k), to account for consistency (i.e., profile similarity) between multiple raters when generalizing to other raters is not intended (McGraw & Wong, 1996). The coefficients ranged from  $ICC = .95-.99$ , regardless of whether they were obtained from the GEMS-9 or the GEMS-45, and regardless of the scoring method,<sup>2</sup> indicating that the instruments produced similar levels of interrater agreement.

To examine the concordance between GEMS-9 and GEMS-45 scores, we compared the respective GEMS profiles for each of the 18 music excerpts. To this end, GEMS-45 and GEMS-9 ratings were averaged across participants for each of the nine GEMS emotions. We derived two types of values for the GEMS-45, in accordance with the maximum-value and weighted-mean scoring procedures described above. We treated the aggregate scores of the GEMS-45 and the GEMS-9 as *raters*, and the emotions as *subjects*. To obtain a comprehensive metric of concordance, we not only tested for profile similarity but also similarity in elevation (i.e., absolute values). We did so by computing ICCs for absolute agreement rather than consistency. Specifically, we used the ICC(A,1) or two-way mixed-effects model, single measure (see McGraw & Wong, 1996, Case 3).

The patterns of concordance between the GEMS-45 and the GEMS-9 are shown in Figure 2, with the three GEMS profiles for the classical music pieces displayed in the top row (C1-C6), the hip-hop pieces in the middle row (H1-H6), and the pop music pieces in the bottom row (P1-P6). Using weighted-mean scoring for the GEMS-45, the average ICC was  $.67$  ( $SD = .13$ ,



**Figure 2.** Patterns of concordance between the GEMS-45 and the GEMS-9. Note. ICC values in the upper left corner of the figures refer to congruence between the GEMS-45 scores (orange lines) with those obtained with the GEMS-9 (blue lines); w = weighted mean scoring; m = maximum value scoring. Labels on X-axis are abbreviated GEMS dimensions (Wonder, Transcendence, Tenderness, Nostalgia, Peacefulness, Power, Joyful Activation, Sadness, Tensio). Data relating to these analyses are available under: <https://osf.io/fsnk7/>



range: .28–.85), whereas using maximum-value scoring the average ICC was .89 ( $SD = .06$ , range: .68–.94). More detailed statistical values are reported in Supplemental Table S3.

## Discussion

The main question addressed in this study was whether the full-length version of the GEMS-45, and its popular 9-item checklist derivative (GEMS-9), produce similar levels of interrater agreement and similar average profiles for the same music excerpts. In terms of agreement between individual raters, results obtained for the GEMS-45 and GEMS-9 were satisfactory and similar in magnitude. With regard to emotion scale scores for the music excerpts, results obtained with the GEMS-45 and the GEMS-9 were similar in terms of their shape, whereas results for agreement in absolute values depended on the way the GEMS-45 scale scores were computed.

Using the maximum-scoring method, the scores obtained with the GEMS-45 and the GEMS-9 were largely similar. If scale scores were computed using weighted-mean scoring, the picture was less clear, with acceptable overlap for some music excerpts but non-trivial differences for others. In general, the differences resulted from GEMS-45 scale scores being lower relative to the scores obtained with the GEMS-9. Because several items make up each of the GEMS-45 scales and unselected scale-items were accorded a value of 0, this outcome is unsurprising. An adjustment for this issue was provided by the maximum-scoring method, as it only takes the highest-scoring item per scale into account.

We should note that differences in absolute values are not necessarily a concern, and they will matter primarily if scores of the GEMS-45 need to be directly compared or combined with those obtained with the GEMS-9. Furthermore, discrepancies in absolute values between GEMS-45 and GEMS-9 were limited to only a few dimensions. The largest discrepancies were found for Wonder, which is not entirely surprising for two reasons. First, in the GEMS-45, *happy* is part of the Wonder dimension (Zentner et al., 2008), whereas in the GEMS-9 Wonder has no happiness connotation because it is only represented by the term *wonder*, presented along with the example terms *filled with wonder*, *moved*, and *dazzled*. It is also worth noting that the Wonder scale has been found to exhibit somewhat lower levels of internal consistency compared to the other GEMS scales, which has been explained by interpretational issues with some of the scale items (Vuoskoski & Eerola, 2011).

A question left unanswered by this study is whether results would have differed if participants had been given the all-item rather than the scan-select rating instruction. Using the exhaustive all-item rating procedure might have reduced the number of items receiving a score of 0, thereby elevating the absolute scale values to a level more consistent with that of the GEMS-9. However, compelling participants to rate all 45 items for each music excerpt can be tiring, and excessively so when the number of excerpts to be rated is large. Fatigue can cause participants to miss terms or rate their feelings inattentively, thus potentially offsetting any benefits of an exhaustive rating.

The study has some limitations. First, we strove to select music excerpts that reflect at least some degree of diversity across and within musical genres. Even so, the extent to which the current results generalize to other types of music remains a matter for future research. Second, we should emphasize that the main objective of this research was not to introduce a new scoring method for the GEMS-45. Rather, we proposed scoring alternatives to resolve some discrepancies between the GEMS-45 and the GEMS-9. Furthermore, procedures other than the suggested ones are conceivable, such as standardizing the scores obtained with the GEMS-45 and GEM-9 if the main goal is to consolidate data obtained with both instruments.

Third, the focus of the study was on convergence between the GEMS-45 and GEMS-9 in characterizing emotional effects of music excerpts, rather than on similarity in associations with criterion variables obtained with both versions of the GEMS. Fourth, we compared ratings of induced rather than expressed emotion. This may seem an obvious point since the GEMS is primarily an instrument for measuring induced emotion, but it is one worth recalling in light of the field's emphasis on perceived emotion (Warrenburg, 2021). Finally, for all its practical advantages, the GEMS-9 cannot replace the in-depth assessment of emotions induced by music offered by the GEMS-45. Despite these limitations, this study shows that if researchers wish to capture some of the distinct features of music-evoked emotions but do not have the time or resources for an in-depth assessment, the GEMS-9 can be a viable alternative to the GEMS-45.



### Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded in part by the Austrian Science Fund (FWF): <https://doi.org/10.55776/P33526>

### ORCID iDs

Peer-Ole Jacobsen  <https://orcid.org/0009-0003-3416-3326>  
Marcel Zentner  <https://orcid.org/0000-0001-8580-8030>

### Ethics approval

The questionnaire and methodology for this study was approved by the Ethics Board of the Department of Psychology, University of Innsbruck.

### Availability of data and materials

The datasets generated and analyzed during the current study are available in the a OSF repository with the following link: <https://osf.io/fsnk7/>

### Supplemental material

Supplemental material for this article is available online.

### Notes

1. The GEMS-45 represents the consolidated terms from Study 3 and Study 4 by Zentner et al. (2008).
2. The average single measure ICC(C,1) was .21 for the GEMS-45 for both scoring types using the weighted mean ( $SD = .09$ , range: .08–.42), and .26 for the GEMS-9 ( $SD = 0.09$ , range: .10–.46).

### References

- Barrett, F. S., Grimm, K. J., Robins, R. W., Wildschut, T., Sedikides, C., & Janata, P. (2010). Music-evoked nostalgia: Affect, memory, and personality. *Emotion, 10*(3), 390–403. <https://doi.org/10.1037/a0019006>
- Chelkowska-Zacharewicz, M., & Janowski, M. (2021). Polish adaptation of the Geneva Emotional Music Scale: Factor structure and reliability. *Psychology of Music, 49*(5), 1117–1131. <https://doi.org/10.1177/0305735620927474>



- Chen, Y.-A., Yang, Y.-H., Wang, J.-C., & Chen, H. (2015, April 19–24). The AMG1608 dataset for music emotion recognition. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 693–697). IEEE. <https://doi.org/10.1109/ICASSP.2015.7178058>
- Gerstgrasser, S., Vigl, J., & Zentner, M. (2023). The role of listener features in musical emotion induction: The contributions of musical expertise, personality dispositions, and mood state. *Psychology of Aesthetics, Creativity, and the Arts*, *17*(2), 211–224. <https://doi.org/10.1037/aca0000468>
- Gómez-Cañón, J. S., Cano, E., Eerola, T., Herrera, P., Hu, X., Yang, Y.-H., & Gómez, E. (2021). Music emotion recognition: Toward new, robust standards in personalized and context-sensitive applications. *IEEE Signal Processing Magazine*, *38*(6), 106–114. <https://doi.org/10.1109/MSP.2021.3106232>
- Han, D., Kong, Y., Han, J., & Wang, G. (2022). A survey of music emotion recognition. *Frontiers of Computer Science*, *166*(6), Article 166335. <https://doi.org/10.1007/s11704-021-0569-4>
- Hashim, S., Stewart, L., & Küssner, M. B. (2020). Saccadic eye-movements suppress visual mental imagery and partly reduce emotional response during music listening. *Music & Science*, *3*, 205920432095958. <https://doi.org/10.1177/2059204320959580>
- Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*(3), 260–280. <https://doi.org/10.1111/j.1745-6916.2007.00044.x>
- Juslin, P. N. (2013). What does music express? Basic emotions and beyond. *Frontiers in Psychology*, *4*, Article 596. <https://doi.org/10.3389/fpsyg.2013.00596>
- Juslin, P. N. (2019). *Musical emotions explained: Unlocking the secrets of musical affect* (1st ed.). Oxford University Press.
- Kaelen, M., Barrett, F. S., Roseman, L., Lorenz, R., Family, N., Bolstridge, M., Curran, H. V., Feilding, A., Nutt, D. J., & Carhart-Harris, R. L. (2015). LSD enhances the emotional response to music. *Psychopharmacology*, *232*(19), 3607–3614. <https://doi.org/10.1007/s00213-015-4014-y>
- Koelsch, S. (2018). Investigating the neural encoding of emotion with music. *Neuron*, *98*(6), 1075–1079. <https://doi.org/10.1016/j.neuron.2018.04.029>
- LimeSurvey GmbH. (n.d.). *LimeSurvey* [Computer software].
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30–46. <https://doi.org/10.1037/1082-989X.1.1.30>
- Pearce, M. T., & Halpern, A. R. (2015). Age-related patterns in emotions evoked by music. *Psychology of Aesthetics, Creativity, and the Arts*, *9*(3), 248–253. <https://doi.org/10.1037/a0039279>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. <https://doi.org/10.1037/h0077714>
- Soleymani, M., Caro, M. N., Schmidt, E. M., Sha, C.-Y., & Yang, Y.-H. (2013, October 22). 1000 songs for emotional analysis of music. In *Proceedings of the 2nd ACM International Workshop on Crowdsourcing for Multimedia* (pp. 1–6). Association for Computing Machinery. <https://doi.org/10.1145/2506364.2506365>
- Strauss, H., Vigl, J., Jacobsen, P.-O. J., Bayer, M., Talamini, F., Zangerle, E., & Zentner, M. (2024). The Emotion-to-Music Mapping Atlas (EMMA): A systematically organized online database of emotionally evocative music excerpts. *Behavior Research Methods*. Advance online publication. <https://doi.org/10.3758/s13428-024-02336-0>
- Vuoskoski, J. K., & Eerola, T. (2011). Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Scientiae*, *15*(2), 159–173. <https://doi.org/10.1177/1029864911403367>
- Warrenburg, L. A. (2021). The PUMS database: A corpus of previously-used musical stimuli in 306 studies of music and emotion. *Empirical Musicology Review*, *16*(1), 145–150. <https://doi.org/10.18061/emr.v16i1.7208>
- Zentner, M., & Eerola, T. (2010). Self-report measure and models. In P. N. Juslin & J. A. Sloboda. (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 187–221). Oxford University Press.
- Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*(4), 494–521. <https://doi.org/10.1037/1528-3542.8.4.494>
- Zhang, K., Zhang, H., Li, S., Yang, C., & Sun, L. (2018, June 11–14). The PMemo dataset for music emotion recognition. *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval* (pp. 135–142). Association for Computing Machinery. <https://doi.org/10.1145/3206025.3206037>