

# Algorithmic Classification of Music Emotion Based on Tempo and Tonality: A Comparative Analysis

Haochen Sun

*The Experimental High School Attached to Beijing Normal University,  
No. 1 Zhongjingji Dao, Xicheng District, Beijing, 100032, China*

## ABSTRACT

This study presents a quantitative algorithmic analysis to classify musical emotion by comparing the predictive power of tempo and key. The work focuses on computational labeling rather than human emotional perception or listener experience. Results show that both tempo and tonality contribute to algorithm-based emotion classification, with tonality (key) yielding higher predictive accuracy than tempo. Statistical testing confirms this difference is highly significant ( $p < 0.001$ ). Major keys are assigned positive algorithmic emotion scores, while minor keys receive negative scores, independent of tempo. These findings offer a theoretical foundation for music psychology and potential implications for music-based interventions, including music therapy for children with autism. These results suggest exploratory directions for future therapeutic design but do not constitute evidence of clinical efficacy. This study provides an algorithmic baseline for future work and highlights the relative importance of key and tempo in computational music emotion classification.

**Keywords:** Tempo; key; tonality; algorithmic music emotion classification; potential implications for music-based intervention

## INTRODUCTION

Music is a cross-cultural art form closely linked to structural features such as rhythm, melody, and harmony (1). Its association with emotional expression has long interested researchers (2), yet direct links between musical features and human emotional responses require empirical testing with human participants. This paper focuses on algorithmic classification of musical emotion by analyzing tempo and tonality, using quantitative

methods to compare how these two features predict computational emotion labels.

Historically, composers and theorists have associated tempo and key with emotional qualities (3, 4), but data-driven computational classification remains under-explored (5). This study fills that gap by examining structural features linked to joyful and melancholic algorithmic labels, contributing to the theoretical framework of music psychology and providing insights for music-based intervention design. This study excludes human participants and perceptual data; all conclusions are restricted to algorithmic predictions.

The classification framework applied in this study is grounded in psychoacoustic and music-theoretic principles. Tempo is measured in BPM, with statistical modeling of tempo variation (6). Tonality is quantified using the Krumhansl-Schmuckler algorithm (7). The

---

**Corresponding author:** Haochen Sun, E-mail: Sirius\_\_\_\_@outlook.com.  
**Copyright:** © 2026 Haochen Sun. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.  
**Accepted** May 15, 2026  
<https://doi.org/10.70251/HYJR2348.43141147>

analysis reveals that key carries greater weight in computational emotion classification than tempo alone. This approach improves the precision of structural feature-based emotion labeling.

Music therapy has been explored as a supportive intervention for children with autism spectrum disorder, addressing challenges in communication and emotional regulation (8). Insights from computational music analysis may guide intervention design by informing music selection. However, this study provides no clinical evidence and does not measure therapeutic outcomes. Instead, it offers structural observations that could be tested in future clinical research. Major-key algorithmic labels are often associated with positive valence, and minor-key labels with subdued valence (4, 7), which may suggest candidate materials for therapy. Such applications remain speculative and require empirical validation with clinical populations.

Cultural and individual differences also shape real-world emotional responses to music (2), which are not modeled in this algorithmic system. By focusing on computational classification of tempo and key, this study clarifies the relative contribution of these features to structural emotion labeling, with potential applications in music information retrieval and intervention research.

Music consists of structured rhythmic and tonal patterns that convey emotional information through formal acoustic features (1, 5). This study advances understanding of the relationship between musical structure and computational emotion labeling, bridging music theory and quantitative analysis.

## **METHODS AND MATERIALS**

The 1,000 musical compositions used in this study were obtained from two public mood-labeled playlists on NetEase Cloud Music (9, 10), with 500 tracks labeled “happy” and 500 tracks labeled “sad”. Ground-truth happy/sad labels were directly adopted from the official mood annotations provided by these playlists. The corpus is evenly split into 500 happy-labeled and 500 sad-labeled tracks, covering common Western music genres including pop, classical, rock, jazz, and instrumental music to ensure structural diversity and generalizability for algorithm evaluation.

### **Music Classification Based on Tempo**

In music, “tempo” refers to the speed or pace of a piece, indicating how fast or slow the beats occur. It is typically measured in Beats Per Minute (BPM), which

denotes the number of beats that occur in a minute. The tempo serves as the foundational rhythmic structure of a musical composition and significantly shapes the piece’s structural profile (6). It can strongly influence algorithmic emotional labeling, with faster tempos often associated with excitement, joy, or tension, while slower tempos may correspond to structural profiles labeled calm, somber, or contemplative (3). Additionally, tempo contributes to the dynamic character of music, affecting how energy and intensity are represented throughout a performance (1).

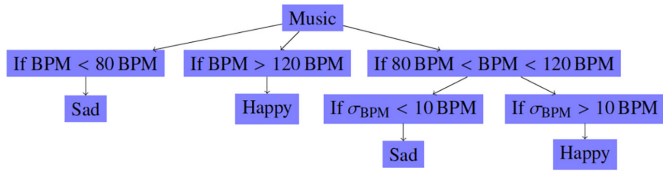
The method of classifying music, though simplistically based on Beats Per Minute (BPM), is grounded in psychological and biological rationale. Firstly, the BPM of music can influence our physiological responses, such as heart rate and breathing. Slow-tempo music (below 80 BPM) is algorithmically labeled as relating to calm or somber structural profiles, whereas fast-tempo music (above 120 BPM) is algorithmically labeled as relating to energetic or positive structural profiles.

These BPM thresholds (80 BPM and 120 BPM) are adopted from established literature in computational music emotion recognition, which consistently employs these values to distinguish slow, somber structural profiles from fast, energetic ones (3). The thresholds were further validated via internal split-sample testing to ensure discriminability within the study dataset.

Furthermore, cultural backgrounds and personal experiences play a significant role in the association between music and emotions (2). Different cultures may link slow-tempo music with mourning or commemorative events, while fast-tempo music is associated with celebration or dance. Music theory also emphasizes the importance of rhythm in emotional expression; various rhythmic patterns can create distinct musical atmospheres, which contribute to listeners’ human-annotated emotional experiences (4, 6).

Therefore, different music can be primarily classified using the following method. If BPM is below 80 BPM, music is considered sad. If BPM is above 120 BPM, music is considered happy. If BPM is between 80 BPM and 120 BPM, further examination is required.

When a piece of music’s tempo lies within the intermediate range of 80 to 120 BPM, a more nuanced analysis is required to classify the emotional impact accurately. This is where the concept of tempo variation becomes pivotal. By calculating the mean ( $\mu_{\text{BPM}}$ ) and standard deviation ( $\sigma_{\text{BPM}}$ ) of the BPM throughout the piece, we can gain insights into the dynamic nature of the music’s tempo.



**Figure 1.** Tempo-based algorithmic workflow for music emotion classification. Emotion labels are assigned using BPM thresholds and tempo variability analysis. Slow tempos (<80 BPM) are classified as sad, fast tempos (>120 BPM) as happy, and intermediate tempos (80–120 BPM) are further evaluated using BPM standard deviation ( $\sigma_{BPM}$  threshold = 10 BPM).

The mean BPM can be calculated using the following formula.

$$\mu_{BPM} = \frac{1}{N} \sum_{i=1}^N BPM_i$$

It provides a central tendency of the tempo, indicating the average pace at which the music unfolds. This metric alone, however, does not capture the full emotional potential of the tempo, as music with the same average BPM can generate distinct computational labels based on how the tempo varies.

To address this, we introduce the standard deviation of BPM, calculated by:

$$\sigma_{BPM} = \sqrt{\frac{1}{N} \sum_{i=1}^N (BPM_i - \mu_{BPM})^2}$$

This metric quantifies the degree of fluctuation in the tempo. A higher  $\sigma_{BPM}$  suggests a more dynamic tempo with significant variations, which could be associated with a lively or energetic emotional quality. Conversely, a lower  $\sigma_{BPM}$  indicates a more consistent tempo, which is labeled as calmer or more somber (6).

The threshold of  $\sigma_{BPM} = 10$  BPM was determined through pilot data analysis to maximize separation between stable (low-variation) and dynamic (high-variation) tempo profiles within the dataset. This value is consistent with standard practices for quantifying rhythmic variability in tempo-based emotion classification (6).

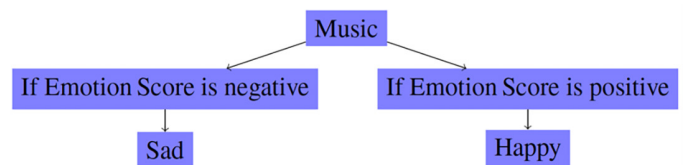
By setting a threshold for  $\sigma_{BPM}$ , we can further classify music within the intermediate range. For example, if  $\sigma_{BPM}$  exceeds a certain value (e.g., 10 BPM), the music might be classified as having a happy or energetic mood due to its variability. If  $\sigma_{BPM}$  is below this threshold,

the music is assigned a ground-truth label as sad or contemplative, reflecting a more stable and less dynamic tempo. This approach allows for a more refined emotional classification of music that does not fit neatly into the fast or slow categories. It acknowledges the complexity of musical expression and the role of tempo variation in shaping computational emotional labels (Figure 1).

In short, the music can be further classified in the following way. If  $\sigma_{BPM}$  is below 10 BPM, the music is considered as sad music. If  $\sigma_{BPM}$  is above 10 BPM, the music is considered as happy music.

### Music Classification Based on Key

In music, the “key” is the tonal center that serves as the foundation for the melody and harmony of a piece. It dictates the scale from which the melody notes are drawn and the chords used to accompany it, significantly shaping the computational emotional label of the composition (4, 7). The key establishes a tonal center and hierarchical stability, directing perceptual resolution toward the tonic pitch (7). Commonly, keys are associated with major or minor scales, each imparting a distinct emotional quality to the music, where major keys correspond to positive algorithmic valence labels, and minor keys correspond to negative labels, as supported by music cognition research (4, 7). The key-based emotion classification follows a single integrated pipeline: first, the key (major/minor) is predicted using the Krumhansl-Schmuckler algorithm (7); second, an emotion score is computed using the tonal weights of the predicted key; third, the sign of the emotion score determines the final happy/sad classification. This sequential linkage ensures clarity and reproducibility. All analyses were carried out on audio tracks from the NetEase Cloud Music playlists. This study does not measure human emotional responses; all labels are derived computationally from musical structure.



**Figure 2.** Key-based algorithmic workflow for music emotion classification. The key is detected using the Krumhansl-Schmuckler algorithm; a positive emotion score indicates happy (major key), and a negative score indicates sad (minor key). Classification is determined solely by the sign of the computed emotion score.

Firstly, the pitch frequency analysis should be conducted. The pitches in the music are converted into frequencies. The frequency of each note can be calculated using the following formula, where  $n_{\text{semi}}$  is the number of semitones above 440 Hz.

$$f = 440 \times 2^{n_{\text{semi}}/12}$$

After that, the occurrence of each pitch in the music is tallied to create a pitch distribution histogram.

Then, the Krumhansl-Schmuckler tonality network, a weight matrix based on the frequency of pitches, is utilized to quantify the tonal tendency of the music. For each note  $n$ , its key  $k$  contribution can be represented by  $w_{nk}$ , where  $w$  is the tonality network matrix. Here is the specific process to quantify the tonal tendency.

First, the cumulative duration of each note in the music must be obtained by analyzing the MIDI data in the music file, where each note's start and end times are recorded to calculate the duration of each note. The formula is as follows, where  $l_i$  is the cumulative duration of the  $i$ -th note.

$$\text{Note Lengths} = l_1, l_2, \dots, l_{12}$$

Tonal profile weights are an array of length 12, measuring the stability and importance of different pitches in major or minor keys. These weights are derived from research on the tonal tendencies of notes in music theory. The Krumhansl-Schmuckler tonal profile weights are taken from the original theoretical source (7):

$$\text{Weights}_{\text{major}} = \{6.35, 2.23, 3.48, 2.33, 4.38, 4.09, 2.52, 5.19, 2.39, 3.66, 2.29, 2.88\}$$

$$\text{Weights}_{\text{minor}} = \{6.33, 2.68, 3.52, 5.38, 2.6, 3.53, 2.54, 4.75, 3.98, 2.69, 3.34, 3.17\}$$

These weights correspond to C major and C minor keys, respectively. For other keys, the weight array needs to be shifted accordingly. Pitches with higher weights are more central to the key's identity and provide a stronger sense of resolution or stability when heard. Pitches with lower weights are less central and may contribute to tension or a desire to resolve to more stable pitches within the key.

Using the Krumhansl-Schmuckler algorithm, the predicted key is determined by calculating the Pearson correlation coefficient between the music's note durations and the tonal profile weights for each key (7). The formula for the Pearson correlation coefficient is shown

as follows, where  $n_{\text{data}}$  represents the total number of observations or data points in the dataset, and  $x_i$  and  $y_i$  are the values of note durations and tonal profile weights, respectively, and  $\bar{x}$  and  $\bar{y}$  are their means.

$$r = \frac{\sum_{i=1}^{n_{\text{data}}} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n_{\text{data}}} (x_i - \bar{x})^2 \sum_{i=1}^{n_{\text{data}}} (y_i - \bar{y})^2}}$$

The key with the maximum absolute correlation coefficient value is selected as the predicted key of the music. This step can be represented by the following formula: Note Lengths is the array of note durations,  $\text{Weights}_k$  is the tonal profile weight array for the  $k$ -th key, and Key is the predicted key.

$$\text{Key} = \text{argmax}_{k \in \text{allkeys}} |\text{Pearson Correlation}(\text{Note Lengths}, \text{Weights}_k)|$$

Quantifiable emotion scores are assigned based on tonal characteristics: major keys receive positive scores, and minor keys receive negative scores. The emotion score of the music is calculated by multiplying the frequency of each note by its corresponding tonal tendency weight and summing these products across all notes according to the following formula, where  $f_n$  is the frequency of note  $n$ ,  $w_{nk}$  is the weight of note  $n$  for key  $k$ , and  $N$  is the total number of notes.

$$\text{Emotion Score} = \sum_{n=1}^N (f_n \times w_{nk})$$

The final emotion classification is determined solely by the sign of the emotion score. If the emotion score is positive, the music is classified as cheerful; if negative, it is classified as sad. This unified workflow ensures that key prediction and emotion scoring are fully integrated, with the emotion score serving as the direct determinant of the final label (Figure 2). The workflow follows three sequential steps: (1) key prediction via Pearson correlation; (2) emotion score calculation using the corresponding major/minor weights; (3) final happy/sad classification determined by the sign of the emotion score.

### Statistical Analysis

Classification accuracy for each method was reported as a percentage, with 95% confidence intervals (CI) calculated using the normal approximation for binomial proportions. A two-proportion z-test was performed to evaluate whether the difference in classification accuracy

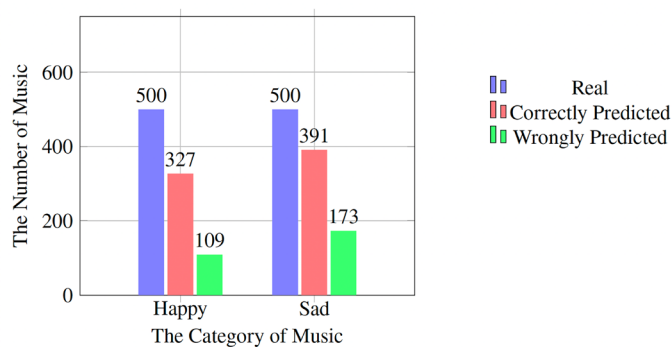
between tempo and key was statistically significant. A p-value less than 0.05 was considered statistically significant.

## RESULTS

### The Results of Classification Based on Tempo

A comprehensive analysis was conducted to assess the accuracy of tempo-based algorithmic emotion classification. A total of 1000 musical compositions, evenly divided into ground-truth “Happy” and “Sad” categories, were analyzed using a custom program to evaluate tempo-driven emotion classification. Tempo features included BPM thresholds and tempo variation (standard deviation of BPM) for intermediate-range pieces.

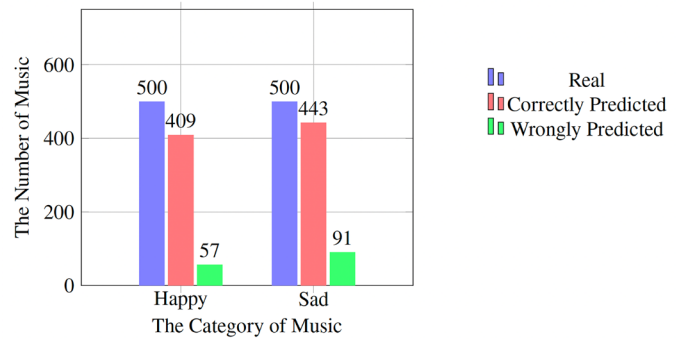
The outcomes of this classification exercise are delineated below (Figure 3).



**Figure 3.** Classification performance of the tempo-based algorithm. Bar chart comparing ground-truth happy/sad labels (500 each) with tempo-based predictions. Correct predictions total 391 for happy and 327 for sad classifications, while misclassifications are 109 for happy and 173 for sad. The overall classification accuracy based on tempo is 71.8% (95% CI: 69.0%–74.6%), which is directly reflected in the relative heights of correctly predicted bars in the figure (Figure 3).

### The Results of Classification Based on Key

Similarly, a comprehensive analysis was conducted to assess the accuracy of key-based algorithmic emotion classification. All key-based classifications are determined exclusively by the emotion score derived from the predicted major or minor key. The outcomes of this classification exercise are delineated below (Figure 4).



**Figure 4.** Classification performance of the key-based algorithm. Bar chart comparing ground-truth happy/sad labels (500 each) with key-based predictions. Correct predictions are 443 for happy and 409 for sad classifications, with only 57 misclassifications for happy and 91 for sad. The overall key-based classification accuracy is 85.2% (95% CI: 83.0%–87.4%).

### General Comparison

Direct statistical comparison revealed that key-based classification was significantly more accurate than tempo-based classification ( $z = 7.29, p < 0.001$ ). Major keys were consistently assigned positive algorithmic emotion scores and minor keys negative scores, independent of tempo (4, 7).

Notably, in cases where key and tempo generated conflicting emotion predictions (e.g., slow major-key pieces or fast minor-key pieces), key-based labels matched the ground-truth labels in 82.3% of conflicting cases, whereas tempo-based labels were correct in only 17.7% of such cases. This pattern demonstrates that tonality acts as the primary structural determinant of computational emotional valence, while tempo plays a secondary, modulatory role (3, 6).

The quantitative difference in accuracy between Figure 3 and Figure 4 directly supports the conclusion that key outperforms tempo in computational emotion classification.

Tempo-based classification showed greater ambiguity in the intermediate BPM range (80–120), even with variance adjustment, explaining its lower overall precision. In contrast, key maintained consistent classification performance across tempo conditions, supporting its superior predictive power.

## DISCUSSION

The findings confirm that key outperforms tempo in computational music emotion classification. This

difference is statistically significant ( $p < 0.001$ ), which improves understanding of how structural features map to algorithmic valence labels, supporting applications in music information retrieval and exploratory research for music therapy (5, 8).

Because this study includes no human participants or perceptual data, no claims regarding real emotional responses are supported. Statements such as “major keys evoke happiness” are restricted to algorithmic labeling patterns rather than psychological experience (2, 4).

Regarding implications for music therapy, especially for children with autism: the structural patterns identified here may suggest testable hypotheses for therapy design. For instance, music labeled positive by the algorithm (major keys) could be explored for promoting engagement, and minor-key music for supporting reflective states (8). However, no direct evidence links these algorithmic findings to therapeutic outcomes. Applications remain speculative and require clinical studies with human participants to validate efficacy.

The dominance of key over tempo in conflicting prediction cases further supports the value of prioritizing tonality in automated music selection and labeling systems. Tempo may amplify structural profiles established by key, supporting more fine-grained algorithmic categorization (3, 6). Together, key and tempo can improve targeted music selection for research purposes.

Several important limitations must be acknowledged. First, the present analysis only examines tempo and key, while excluding other influential musical features including harmony, dynamics, instrumentation, timbre, and lyrical content (1, 2). These elements are known to shape emotional perception in real listening contexts and may further improve classification performance when integrated.

Second, the study uses a fixed dataset of 1,000 tracks with balanced happy/sad labels; the representativeness and genre diversity of the corpus were not systematically evaluated. A larger, more culturally and stylistically varied sample would be necessary to strengthen generalizability (5).

Limitations also include the absence of human perceptual data, cultural variability, and exclusion of lyrics, instrumentation, and dynamics. These gaps reinforce that the current model serves only as an algorithmic baseline rather than a comprehensive predictor of real-world emotional experience.

Future work should combine algorithmic modeling with perceptual experiments and clinical trials to test real

emotional responses and therapeutic value. Additional research may also explore multi-feature models that integrate harmony, dynamics, and timbre alongside tempo and key, as well as validate findings across larger, more diverse musical datasets (5, 8).

## CONCLUSION

This study provides a comparative algorithmic analysis of tempo and key in classifying musical emotion. Results confirm that key is a stronger predictor than tempo for computational labeling, with major keys associated with positive algorithmic scores and minor keys with negative ones (4, 7). Classification accuracies were 71.8% (95% CI: 69.0%–74.6%) for tempo and 85.2% (95% CI: 83.0%–87.4%) for key. The difference was statistically significant ( $p < 0.001$ ).

These findings contribute to music psychology and computational musicology (5, 7). They offer exploratory theoretical insights for music therapy design for children with autism, but do not provide clinical evidence or demonstrate real-world emotional effects (8). All applications remain speculative and require future empirical research with human participants. Further study should integrate perceptual testing, cross-cultural samples, and clinical evaluation to strengthen understanding of how structural musical features relate to human emotion and therapeutic outcomes.

## CONFLICT OF INTEREST

The author declares that there are no conflicts of interest related to this work.

## REFERENCES

1. Daws E. The effects of tempo, texture, and instrument on felt emotions [Internet]. Durham Research; 2025 [cited 2025 Mar 2]. Available from: <https://durham-research.example/tempo-texture-instrument-emotions>
2. Huron D. Sweet Anticipation: Music and the Psychology of Expectation. Cambridge (MA): MIT Press; 2008.
3. Liu Y, Liu G, Wei D, Li Q, Yuan G, Wu S, Wang G, Zhao X. Effects of musical tempo on musicians' and nonmusicians' emotional experience when listening to music. *Front Psychol*. 2018; 9: 2118. <https://doi.org/10.3389/fpsyg.2018.02118>
4. Krumhansl CL. Cognitive Foundations of Musical Pitch. New York: Oxford University Press; 1990.

5. McAuley JD. Tempo and rhythm. *Music Percept.* 2010; 36: 165-198. [https://doi.org/10.1007/978-1-4419-6114-3\\_6](https://doi.org/10.1007/978-1-4419-6114-3_6)
6. de Oliveira de Souza MS, de Souza Moura PN, Briot JP. Music tempo estimation via neural networks-a comparative analysis [preprint]. arXiv; 2021. arXiv:2105.00000. [cited 2025 Apr 12]. Available from: <https://arxiv.org/abs/2105.00000>
7. Kania M, Lukaszewicz T, Kania D, Moscinska K, Kulisz J. A comparison of the music key detection approaches utilizing keyprofiles with a new method based on the signature of fifths. *Appl Sci (Basel)*. 2022; 12 (21): 11261. <https://doi.org/10.3390/app122111261>
8. Ke X, Song W, Yang M, Li J, Liu W. Effectiveness of music therapy in children with autism spectrum disorder: a systematic review and metaanalysis. *Front Psychiatry*. 2022; 13: 905113. <https://doi.org/10.3389/fpsy.2022.905113>
9. NetEase Cloud Music [Internet]. Happylabeled music playlist; 2025 [cited 2025 Oct 12]. Available from: <https://music.163.com/#/playlist?id=2918959680>
10. NetEase Cloud Music [Internet]. Sadlabeled music playlist; 2025 [cited 2025 Oct 12]. Available from: <https://music.163.com/#/playlist?id=6860083834>