

Parametric Affect: An Empirical Study on Pitch and Tempo in Generative Music

Özlem Salman*
RPTU Kaiserslautern-Landau
Kaiserslautern, Germany
nyp64cex@rptu.de

Ko Watanabe*
German Research Center for Artificial
Intelligence
Kaiserslautern, Germany
ko.watanabe@dfki.de

Javier Melo*
RPTU Kaiserslautern-Landau &
German Research Center for Artificial
Intelligence
Kaiserslautern, Germany
javier.melo@dfki.de

Andreas Dengel
RPTU Kaiserslautern-Landau &
German Research Center for Artificial
Intelligence
Kaiserslautern, Germany
andreas.dengel@dfki.de

Shoya Ishimaru
Osaka Metropolitan University
Osaka, Japan
ishimaru@omu.ac.jp

Abstract

Generative AI is increasingly embedded in everyday audio experiences, from personalized playlists to text-to-music tools that synthesize compositions on demand. Yet little is known about how listeners emotionally perceive these AI-generated pieces, or which acoustic features lead listeners to judge a piece as human- or machine-made. We report an exploratory listener study ($N=21$) using 24 five-second clips generated with *MusicGen*, twelve prompted as “happy” and twelve as “sad”. Participants rated each clip on emotional valence, the perceived influence of pitch and tempo, and whether it sounded human-made or AI-generated. Participants recovered the intended emotion in approximately 79% of trials, and *tempo* ratings were systematically associated with: high-tempo clips were more often labeled *happy* and more often attributed to a *human* composer. Pitch effects were statistically significant but small. We discuss implications for designing emotionally and socially appropriate generative audio in ubiquitous computing settings.

CCS Concepts

• **Human-centered computing** → **User studies**; *Sound-based input / output*; • **Computing methodologies** → *Sound and music computing*; Neural networks.

Keywords

Generative AI, music perception, emotion, ubiquitous computing

ACM Reference Format:

Özlem Salman, Ko Watanabe, Javier Melo, Andreas Dengel, and Shoya Ishimaru. 2026. Parametric Affect: An Empirical Study on Pitch and Tempo in Generative Music. In *Companion of the 2026 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp Companion*

'26), October 11–15, 2026, Shanghai, China. ACM, New York, NY, USA, 5 pages.
<https://doi.org/10.1145/3798063.3837171>

1 Introduction

Artificial intelligence has rapidly extended into creative practice, with specialized generative tools now producing visual art, poetry, and music [11]. Adoption of these tools is rising globally [2], and in some perceptual studies AI-generated pieces are reported to match or surpass human work on aesthetic measures [10]. This shift raises questions not only about how AI systems create, but also about how humans *perceive* and emotionally respond to what they produce [16]. Understanding that perceptual bridge is central to designing AI experiences that align with human expectations, a long-running concern of ubiquitous computing, where sensor-rich environments and mobile or wearable devices increasingly mediate how everyday behavior, mood, and affective experience are sensed, interpreted, and acted upon [7, 18].

Among the arts, music has a particularly tight coupling with emotion. People often report stronger emotional responses to music than to other art forms [17], and peak musical moments have been linked to endogenous dopamine release in the striatum [12]. Because of this, music has long been used as a probe for affect in pervasive and wearable computing. However, most prior work has framed AI as a *responder* to human emotion, for instance recommending songs that fit a detected mood [17]. Text-to-music generation reverses that direction: the listener *describes* an intended emotion in natural language, and the system synthesizes audio in return [1, 20]. Whether the resulting piece evokes the intended affect, and whether listeners can tell its output apart from a human composition, remains an open empirical question.

Listeners’ answer to this second question matters because attribution shapes response: people often rate music as less emotionally engaging when they believe it was produced by AI [5]. As AI-generated audio becomes routine in streaming apps, in-car systems, and wearable devices, the cues by which listeners decide that a piece sounds “human” become design-relevant.

In this study, we ask which acoustic features influence two distinct judgments about short AI-generated clips: (i) the perceived

*These authors contributed equally to this research.



emotional valence of the clip, and (ii) the perceived authorship (human vs. AI). We focus on *tempo* and *pitch*, both well-established affective cues in music. Twenty-one participants listened to 24 MusicGen-generated five-second clips (twelve clips for each “happy”, and “sad”), and rated each on liking, valence, the influence of pitch and tempo, and perceived authorship. Our findings are twofold. First, participants recovered the intended emotion in roughly 79% of trials, but tempo confounded valence judgments: low-tempo happy clips were sometimes heard as sad and vice versa. Second, and more novel for the ubiComp community, higher-tempo clips were systematically more often attributed to a human composer. Pitch had statistically significant but substantially smaller effects on both judgments.

These findings contribute an early, exploratory account of how listeners perceive AI-generated music and suggest that tempo can act as a perceptual cue for human agency. We see this as a step toward an empirically grounded design vocabulary for pervasive AI audio.

2 Related Work

2.1 Generative AI music

Recent text-to-music systems generate audio from natural-language descriptions of mood, instrumentation, and style [1, 20]. Evaluations in this line have largely focused on generation quality and on emotional alignment from a computational perspective [4]. Adjacent work in ubiComp has explored context-driven generation: Suzawa et al. [14, 15] use real-time AI music to support online meetings, foregrounding system utility rather than the listener’s subjective response. Our focus is the listener side.

2.2 Acoustic features and emotion

Acoustic features such as tempo, mode, and pitch are well-known to shape perceived affect. Schellenberg et al. [13] manipulated pitch and rhythm in short melodies across happy, sad, and scary categories, and reported that pitch differences influenced affect ratings more strongly than rhythm, with rhythm mattering chiefly when it interacted with pitch. More recent neural evidence from Yang et al. [19] indicates that tempo can in fact dominate musical mode in shaping emotional perception. These mixed results motivate examining pitch and tempo together rather than in isolation.

2.3 Perceiving AI authorship

How listeners *frame* a piece, as human or machine, also shapes how they respond. Listeners typically report less intense emotional responses and lower aesthetic ratings when they attribute a work to AI, especially when they view creativity as a distinctly human capacity [5, 16]. Maintaining a balance between predictability and stylistic novelty has been suggested as one way to preserve engagement [9]. What remains less clear is which acoustic cues drive that attribution in the first place. Our study addresses that gap by examining how pitch and tempo influence both emotional perception and authorship attribution of short AI-generated clips.

Table 1: Pre-survey: demographic and background questions.

Item	Question	Response Type
Q1	Your gender	Multiple-choice
Q2	Your age	Open-ended
Q3	Nationality	Open-ended
Q4	On a scale from 1 (no knowledge) to 5 (expert), how would you rate your understanding of music theory?	5-point scale
Q5	How many years of formal music training have you completed?	Multiple-choice
Q6	Can you play at least one instrument currently?	Yes/No

Table 2: Post-survey: questions answered after each clip.

Item	Question	Response Type
Q7	Do you like the song?	5-point scale
Q8	Is this a happy or sad song?	Multiple-choice
Q9	Why did you rate it as happy or sad?	Open-ended
Q10	How much did the pitch of the song influence your rating?	5-point scale
Q11	How much did the tempo of the song influence your rating?	5-point scale
Q12	Is this song human-made or AI-generated?	4-point ordinal

3 Methodology and Data Collection

3.1 Materials

We generated 24 clips using HuggingFace MusicGen¹. Twelve were prompted as “happy” and twelve as “sad” using emotion words drawn from the music-mood dataset of Cano and Morisio [3]. Each prompt specified mood, instrumentation, and style (e.g., “a cheerful, jolly wedding song with piano and violin”). Each clip was exactly five seconds long. Clips were presented in their original generated form, with no subsequent pitch or tempo manipulation, embedded in AidaForm alongside the survey items. All participants evaluated the same set of clips.

3.2 Participants

A total of 21 participants (11 female, 9 male, 1 non-binary) took part in the study. All participants were based in Germany and were either university students or employees. Each participant self-assessed their level of musical expertise on a five-point scale (1 = lowest, 5 = highest). For the analyses, participants who rated their expertise as 3 or below were categorized as having low musical knowledge, whereas those who rated 4 or 5 were categorized as having high musical knowledge. An informed consent form was provided to all participants before the experiment, and allowed to opt out at any time without penalty. Participants received one participation hour as compensation for their contribution when applicable.

3.3 Procedure

Participants first completed a pre-survey covering demographics and musical background (Table 1). They then listened to each of the 24 clips and, for each clip, completed the post-survey items

¹https://huggingface.co/docs/transformers/model_doc/musicgen

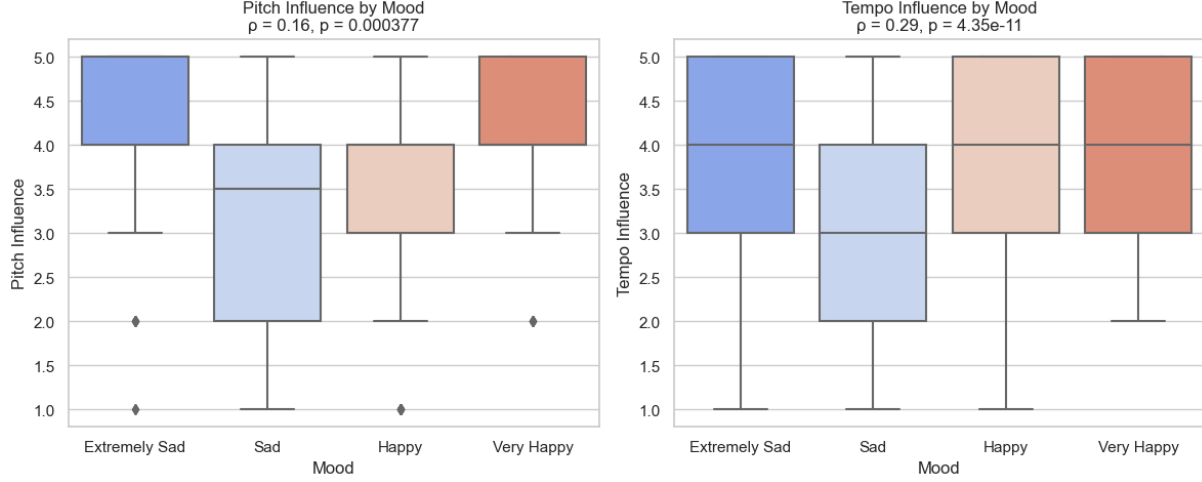


Figure 1: Effect of tempo and pitch on emotional perception of the songs. It is seen that higher tempo has a direct relation with songs being labeled as happy.

in Table 2. Items used a 5-point Likert scale except for emotional valence (happy/sad, binary) and authorship attribution, which used a 4-point ordinal scale: *definitely human-made, maybe human-made, maybe AI-generated, definitely AI-generated*. Participants were also encouraged to briefly justify their valence judgment in open text (“Why did you rate it as happy or sad?”). All participants listened to the same clips in a fixed order, with the twelve sad-prompted clips first, followed by the twelve happy-prompted clips. The session took approximately 25–30 minutes.

3.4 Data Analysis

We analyzed the data in Python (pandas, scipy). Because each participant rated all 24 clips, the resulting design is repeated-measures at the participant level but pools trials when comparing across stimulus categories. For the between-category contrasts on perceived influence of pitch/tempo and on authorship rating, we used two-tailed Welch’s t-tests (unequal variances) with $\alpha = .05$; this is reflected in the non-integer degrees of freedom reported below. Effect sizes are reported as Cohen’s d , with thresholds following Lovakov and Agadullina [8]. To characterize the strength of evidence for the strongest effects, we report Bayes factors [6]. Open-ended responses were not included in the statistical analysis.

4 Result and Discussion

All 21 participants completed the study by providing ratings for emotional perception (happy vs. sad), authorship attribution (AI vs. human), and the influence of tempo and pitch perception for each song.

4.1 Emotional Perception

Subjective emotional classifications of the songs from participants generally aligned with the intended emotional labels. Out of 24 songs, the overall correct emotion identification was approximately 79% and for the happy-prompted clips, it was 70% accuracy. Sad-prompted clips were mostly recognized, but several mid-tempo

sad clips were misclassified as happy. Interestingly, some of the lower-tempo happy songs were rated as sad ones, suggesting that the tempo or energy level of the song may have a greater influence on emotional perception than other musical cues such as harmonic structure. These results suggest that participants were mostly able to recognize the intended emotional tone of the songs, although differences in tempo sometimes affected their judgments and caused confusion between happy and sad pieces.

4.2 Acoustic Features and Emotional Perception

Our two-tailed paired samples t-test revealed a small, but significant difference in perceived emotional influence for pitch ($t(280.61) = 2.15, p < .05, d = 0.24$). On the other hand, the effect of tempo was of medium size [8] and highly significant ($t(287.75) = 5.12, p < .001, d = 0.57$). A Bayes factor of $BF_{10} = 2.45 \times 10^4$ provides decisive evidence for a tempo-related difference [6]. These results indicate that perceived tempo influence was strongly associated with the participants. While pitch and tempo had significant effects, the statistics show that higher tempo strongly correlates with songs being labeled as happy by the participants. The effect is illustrated in Figure 1.

4.3 Acoustic Features and Authorship Perception

For the analysis of authorship perception, a two-tailed independent samples t-test was applied, and this revealed that both pitch and tempo had a significant effect on authorship perception. The influence of pitch was small but significant ($t(455.33) = -2.53, p = .012, d = 0.22$), whereas the effect of tempo was stronger and highly significant ($t(486.27) = -5.62, p < .001, d = 0.48$), with a Bayes factor ($BF_{10} = 3.1 \times 10^5$) providing a strong evidence for a tempo-related difference. These results showed that songs with higher tempo were more frequently perceived as human-made, indicating that tempo may be used as a cue for perceived human agency in music generation, as shown in Figure 2.

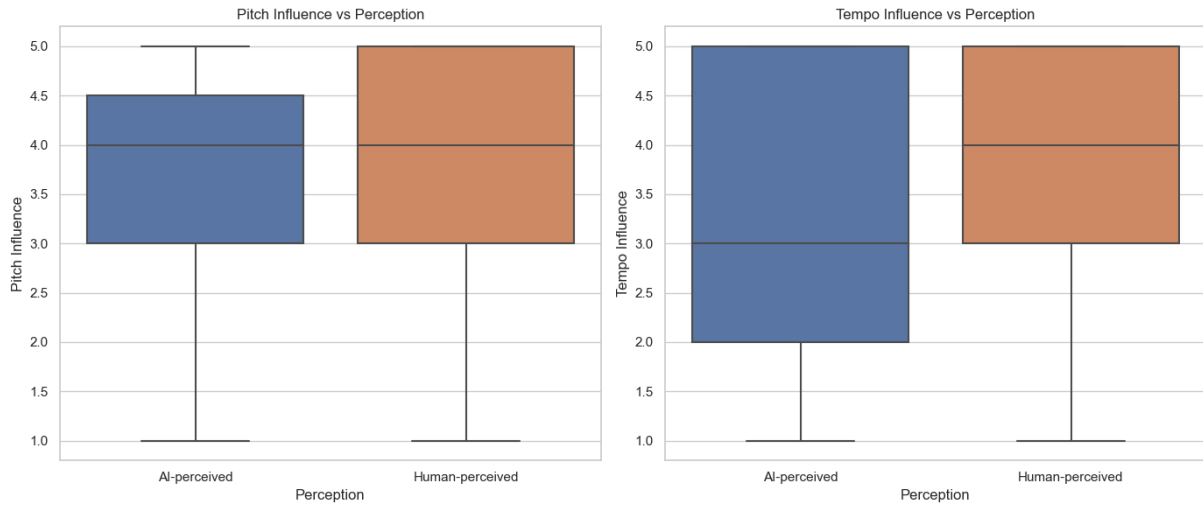


Figure 2: Effect of pitch and tempo in authorship perception

4.4 Discussion

In this study we discovered two findings. First, while listeners generally recognized the intended emotion of the clips, *tempo* systematically distorted both affect and authorship judgments. High-tempo sad clips were sometimes heard as happy, echoing EEG evidence from Yang et al. [19] that tempo can override musical mode in shaping perceived emotion.

Second, and more novel, tempo predicted whether a clip was heard as human-made: faster clips were more often attributed to a human composer. This result has direct implications for how generative audio is deployed in pervasive and wearable computing settings, adaptive in-car audio, wearable music players, and ambient meeting soundtracks such as those proposed by Suzawa et al. [14]. Listeners tend to rate music as less emotionally engaging when they believe it was AI-generated [5], particularly when they view creativity as inherently human. If specific acoustic features such as tempo prominent among them, systematically shift attribution toward “human-made”, designers can deliberately tune those features so that listener expectations align with the affective intent of the system. One plausible mechanism is that faster tempi carry cues of energy and expressivity at a sub-conscious level, evoking human agency. Combining such cues with carefully balanced predictability and novelty [9] may improve user satisfaction with AI-generated audio without misrepresenting its origin.

We see these results as a first step toward an empirically grounded vocabulary for designing emotionally and socially appropriate AI audio in ubiquitous contexts.

5 Limitations and Future Work

While this study provides significant insights into the perception of AI-generated music, certain methodological constraints should be noted. First, the stimulus duration was restricted to five seconds. Although this brief window was sufficient to capture perceptual evaluations, it may have provided limited musical context to evoke the full emotional depth found in longer compositions. Nevertheless,

the presence of significant results within such a short timeframe suggests that tempo and pitch serve as highly salient cues in human auditory perception. Additionally, the presentation order in the online survey was not counterbalanced, as stimuli were grouped by emotional category, with all “sad” songs were followed by all “happy” compositions. This structure may have introduced ordering effects, potentially influencing participant responses through contrast or fatigue. However, the robustness of the observed trends indicates that the manipulated variables exerted a primary influence that persisted despite the non-randomized sequence.

These limitations provide a clear path for subsequent research. Future studies should employ a counterbalanced design and longer musical excerpts to further elucidate the role of pitch and tempo in emotional responses to music. Furthermore, exploring why high-tempo music specifically triggers associations with human agency will further clarify the psychological mechanisms of auditory perception.

6 Conclusion

In conclusion, although we use the necessary wording and explanations in text prompt-based AI-generators, the intended emotion may not be perceived as such by the user. Our study revealed that while tempo and pitch had a significant impact on emotional perception, tempo had a stronger effect on both emotional perception and author attribution of the pieces. People tend to classify songs as happy if they have a relatively higher tempo, and they are prone to think that high-tempo music is associated with human-made music. This association might be coming from the phrases we use about happiness, which indicates that happiness is related to human connection. More research on this would help us to understand the underlying reasons for this phenomenon.

Acknowledgement

Claude Opus 4.7 was utilized to improve the texts and writings of this manuscript.

References

- [1] Andrea Agostinelli, Timo I. Denk, Zolán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank. 2023. MusicLM: Generating Music From Text. arXiv:2301.11325 [cs.SD] <https://arxiv.org/abs/2301.11325>
- [2] Alexander Bick, Adam Blandin, and David J. Deming. 2026. The Rapid Adoption of Generative AI. *Management Science* 0, 0 (2026), 1–20. doi:10.1287/mnsc.2025.02523 Published online January 20, 2026.
- [3] E. Cano and M. Morisio. 2017. Music Mood Dataset Creation Based on Last FM Tags. In *Proceedings of the International Conference*. 15–26. doi:10.5121/CSIT.2017.70603
- [4] Wenqian Cui, Pedro Sarmiento, and Mathieu Barthet. 2024. MoodLoopGP: Generating Emotion-Conditioned Loop Tablature Music with Multi-granular Features. In *Artificial Intelligence in Music, Sound, Art and Design*, Colin Johnson, Sérgio M. Rebelo, and Iria Santos (Eds.). Springer Nature Switzerland, Cham, 97–113.
- [5] Rohan L. Dunham, Gerben A. van Kleef, and Eftychia Stamkou. 2025. The Threat of Synthetic Harmony: The Effects of AI vs. Human Origin Beliefs on Listeners' Cognitive, Emotional, and Physiological Responses to Music. *Computers in Human Behavior: Artificial Humans* 6 (2025), 100205. doi:10.1016/j.chbah.2025.100205
- [6] Sir Harold Jeffreys and Sir Harold Jeffreys. 1998. *The Theory of Probability* (third edition, third edition ed.). Oxford University Press, Oxford, New York.
- [7] Eiman Kanjo, Luluah Al-Husain, and Alan Chamberlain. 2015. Emotions in context: examining pervasive affective sensing systems, applications, and analyses. *Personal and Ubiquitous Computing* 19, 7 (2015), 1197–1212.
- [8] Andrey Lovakov and Elena R. Agadullina. 2021. Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology* 51, 3 (2021), 485–504. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/ejsp.2752> doi:10.1002/ejsp.2752
- [9] Ernest Mas-Herrero and Josep Marco-Pallarés. 2025. Predictive processes shape individual musical preferences. *Proceedings of the National Academy of Sciences* 122, 29 (2025), e2500494122. arXiv:<https://www.pnas.org/doi/pdf/10.1073/pnas.2500494122> doi:10.1073/pnas.2500494122
- [10] Atte Oksanen, Anica Cvetkovic, Nalan Akin, Rita Latikka, Jenna Bergdahl, Yang Chen, and Nina Savela. 2023. Artificial intelligence in fine arts: A systematic review of empirical research. *Computers in Human Behavior: Artificial Humans* 1, 2 (2023), 100004. doi:10.1016/j.chbah.2023.100004
- [11] Shivani Pahuja, Sonal Kukreja, and Akansha Singh. 2025. Comprehensive Review of Generative artificial Intelligence: Mechanisms, Models, and Applications. *Procedia Computer Science* 258 (2025), 3731–3740. doi:10.1016/j.procs.2025.04.628 International Conference on Machine Learning and Data Engineering.
- [12] Valorie N Salimpoor, Mitchel Benovoy, Kevin Larcher, Alain Dagher, and Robert J Zatorre. 2011. Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature neuroscience* 14, 2 (2011), 257–262.
- [13] E Glenn Schellenberg, Ania M Krysciak, and R Jane Campbell. 2000. Perceiving emotion in melody: Interactive effects of pitch and rhythm. *Music Perception* 18, 2 (2000), 155–171.
- [14] Haruki Suzawa, Ko Watanabe, Andreas Dengel, and Shoya Ishimaru. 2025. Augmenting Online Meetings with Context-Aware Real-time Music Generation. In *Proceedings of the Augmented Humans International Conference 2025 (AHs '25)*. Association for Computing Machinery, New York, NY, USA, 487–490. doi:10.1145/3745900.3746116
- [15] Haruki Suzawa, Ko Watanabe, Masakazu Iwamura, Koichi Kise, Andreas Dengel, and Shoya Ishimaru. 2023. Supporting Smooth Interruption in a Video Conference by Dynamically Changing Background Music Depending on the Amount of Utterance. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers (Cambridge, United Kingdom) (UbiComp/ISWC '22 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 299–302. doi:10.1145/3544793.3560384
- [16] Francisco Tigre Moura. 2023. Artificial intelligence, creativity, and intentionality: The need for a paradigm shift. *The Journal of Creative Behavior* 57, 3 (2023), 336–338.
- [17] V. Vijayalakshmi, Pooja Shrivastav, and Gomathi Thiagarajan. 2025. Facial Expression-based AI System for Personalized Music Recommendations. In *Proceedings of the International Conference on Advances and Applications in Artificial Intelligence (ICAAAI 2025)*. Atlantis Press, 273–285. doi:10.2991/978-94-6463-738-0_23
- [18] Aku Visuri, Kennedy Opoku Asare, Elina Kuosmanen, Yuuki Nishiyama, Denzil Ferreira, Zhanna Sarsenbayeva, Jorge Goncalves, Niels van Berkel, Greg Wadley, Vassilis Kostakos, Sarah Clinch, Oludamilare Matthews, Simon Harper, Amy Jenkins, Stephen Snow, and m. c. schraefel. 2018. Ubiquitous Mobile Sensing: Behaviour, Mood, and Environment. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (Singapore, Singapore) (UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 1140–1143. doi:10.1145/3267305.3274142
- [19] Zengyao Yang, Qiruo Su, Jieren Xie, Hechong Su, Tianrun Huang, Chengcheng Han, Sicong Zhang, Kai Zhang, and Guanghua Xu. 2025. Music tempo modulates emotional states as revealed through EEG insights. *Scientific reports* 15, 1 (2025), 8276.
- [20] Yujia Zhao, Mingzhi Yang, Yujia Lin, Xiaohong Zhang, Feifei Shi, Zongjie Wang, Jianguo Ding, and Huansheng Ning. 2025. AI-Enabled Text-to-Music Generation: A Comprehensive Review of Methods, Frameworks, and Future Directions. *Electronics* 14, 6 (2025). doi:10.3390/electronics14061197