

The classification of emotion and music of SongComposer Project

Bohan Xu

Introduction

Song generation through artificial intelligence has emerged as a groundbreaking trend, revolutionizing how creative professionals approach musical composition. The advent of platforms like Suno has democratized music creation, enabling users to generate unique tracks through AI-powered custom prompts. For content creators in fields such as comics, video games, and multimedia storytelling, this technology offers an unprecedented opportunity to craft soundscapes that precisely match the emotional landscape of their narratives.

Our study addresses a critical gap in AI-generated music research. Building on Gao's 2024 study, which identified emotion words' positive impact on music generation, we aim to systematically analyze the specific musical features corresponding to emotional states. By examining these characteristics, we seek to provide insights into how artificial intelligence can nuance musical generation to convey distinct emotional experiences and develop guidelines for creating emotion-driven datasets.

Baseline Model

The SongComposer model (Ding et al., 2023) offers versatile music generation capabilities, including lyrics-to-melody, melody-to-lyrics, song continuation, and music-to-sound transformations. Our research specifically explores the text-to-melody component, which translates emotional input into musical compositions. By focusing on this feature, we aim to understand how AI can systematically convert emotional descriptors into distinct melodic structures.

Method

We classified five primary emotions: joy, sadness, anger, disgust, and fear. Using the SongComposer model, we generated music for each emotional category. Our analysis involved parsing the generated lyrics into distinct features, conducting comprehensive data analysis, and comparing the statistical characteristics across different emotional states.

The choice of emotions—joy, sadness, anger, disgust, and fear—is grounded in practical content creation needs. We intentionally avoided complex emotion taxonomies, focusing instead on fundamental emotional states that creators most frequently seek to capture in musical compositions.

This approach prioritizes practical applicability over theoretical complexity, ensuring the research remains directly relevant to content creators' requirements for emotionally nuanced music generation.

Hypothesis

This research examines four key music features:

1. Sing/Rest Ratio
2. Characters per Song Length
3. Average Pitch
4. Pitch Range

Predicted Emotional Characteristics:

- Joy: Longer singing time, more characters
- Sadness: Lower average pitch, more rest time
- Anger: Fewer characters, high pitch
- Fear: Shorter singing time, highest pitch range
- Disgust: Smallest pitch range

The hypotheses systematically explore how different emotions might manifest in musical generation, providing a structured approach to understanding the relationship between emotional input and musical output.

Result

Emotion	Note Percent	Rest Percent	Average Pitches	Average Pitch Spread	Average words per second
joy	0.899662988	0.100337012	63.96706989	11	2.56902529
sadness	0.859049467	0.140950533	61.87694974	11	2.448324927
anger	0.862595823	0.137404177	63.95542775	9.888888889	2.447362356
fear	0.878714817	0.121285183	62.4427673	9.517241379	2.625720835
disgust	0.869540514	0.130459486	64.41785252	11.86206897	2.546113307

The analysis of music features across different emotions reveals nuanced variations. Joy demonstrates the highest note percentage (0.900) and lowest rest percentage (0.100), while sadness shows the opposite trend. Average pitches range from 61.88 (sadness) to 64.42 (disgust), with pitch spread varying between 9.52 (fear) and 11.86 (disgust). Words per second are relatively consistent, spanning from 2.45 (anger) to 2.63 (fear), suggesting subtle but distinct emotional characteristics in AI-generated music across these fundamental emotional states.

Summary

Key Observations

1. **Pitch and Note Range Characteristics** The AI-generated music demonstrated significant variations in pitch and note range across different emotional states. This variance suggests that the model can effectively translate emotional inputs into distinct musical features, capturing the nuanced characteristics of different emotions.
2. **Linguistic Content** The number of words per second remained relatively consistent across emotional categories, indicating that linguistic density has minimal correlation with the emotional expression of the generated music.

Statistical Feature Validation

Our analysis confirmed that the AI music generation model produces statistically distinguishable features across different emotions. The variations in note percentage, pitch, and rest time provide empirical evidence of the model's ability to capture emotional nuances in musical composition.

Future Research Recommendations

Instead of relying solely on direct emotional labeling, we propose an innovative approach for future research. By providing training data that meets specific statistical feature criteria, researchers could potentially develop more sophisticated emotion-recognition capabilities in AI music generation models.

References

Gao, X., Chen, D., Gou, Z., Ma, L., Liu, R., Zhao, D., Ham, J. (2024). AI-Driven Music Generation and Emotion Conversion. In: Shuichi Fukuda (eds) Affective and Pleasurable Design. AHFE (2024) International Conference. AHFE Open Access, vol 123. AHFE International, USA. <http://doi.org/10.54941/ahfe1004679>

Ding, S., Liu, Z., Dong, X., Zhang, P., Qian, R., He, C., Lin, D., & Wang, J. (2024, February 27). *Songcomposer: A large language model for lyric and melody composition in song generation*. arXiv.org. <https://arxiv.org/abs/2402.17645>