

Emotion Recognition from Composed Music using LMMS and Audio Signal Processing

1. Abstract:

Music can express a wide range of human emotions through subtle changes in rhythm, speed, tone, and intensity. In this work, short instrumental pieces are created using *Linux MultiMedia Studio (LMMS)* to represent four basic moods — happiness, sadness, calmness, and excitement. The audio signals from these compositions are examined with basic techniques from *Digital Signal Processing (DSP)* to obtain key features such as amplitude variation, frequency spread, tempo, and spectral brightness. These parameters are then compared across the four emotional categories to study how sound properties influence emotional feeling.

The study also highlights how LMMS can serve as a simple yet effective platform for experimenting with sound analysis in an academic setting. Observations show clear contrasts among the samples: lively tracks tend to have higher tempo and energy, while slower pieces show lower intensity and limited frequency range.

Such findings form a small step toward developing automatic systems that can recognize emotion from music using only low-level signal features, with the possibility of expanding the approach through machine-learning or brain-signal studies in the future.

2. Introduction:

Music has always been connected with the way people feel things.

Every change in sound — whether in speed, tone, or beat — can change the mood of a listener.

Over the years, researchers have tried to find out how these small sound variations can actually show emotions like joy or sadness. This idea forms the base of what is known as *music emotion recognition*. When we look at music from a technical side, it is nothing but a signal that changes with time. So, by using simple digital-signal-processing (DSP) methods, it becomes possible to study how a sound behaves. Features such as amplitude, frequency, and phase give useful clues about how a tune feels to the human ear. For example, faster beats often sound happy, while slow and low tones sound calm or sad. To test this relation, I have used *Linux MultiMedia Studio (LMMS)*, which is a free and open-source music-making tool. It allows anyone to create small melodies by changing tempo, pitch, and different sound instruments. In this work, I have made short instrumental pieces for four basic moods — happy, sad, calm, and energetic.

After creating them, the signals are checked for their basic DSP properties such as loudness, tempo, and frequency range. By comparing these values, the difference between each emotion can be observed in a clear and simple way. This small study mainly focuses on understanding how music signals and human emotions are related, and how open-source tools like LMMS can help students learn DSP concepts through real sound experiments.

3.Literature Survey

Research on music emotion recognition has grown rapidly in recent years. Different authors have explored machine-learning models, deep-learning systems, and various audio features to understand how music expresses emotions. This section reviews the main ideas from recently published papers and highlights the strengths and limitations found in their work.

3.1 Kang et al. (2024)

Kang and colleagues presented an overview of recent work in music emotion recognition. They explained how different studies use different labeling schemes, which makes the results difficult to compare. The paper also pointed out that most researchers depend on large datasets collected from the internet. Because the recordings come from many sources, their sound quality and style vary a lot. The authors suggested that controlled audio samples may give clearer insights into how musical features relate to emotions.[1]

3.2 Louro et al. (2024)

This work compared traditional machine-learning methods with modern deep-learning models. The authors used features such as MFCCs and spectrograms. They reported that deep networks performed better in most cases, but they required a large amount of training data. The study also mentioned that these models often behave like “black boxes,” and it is not easy to explain why they predict a certain emotion. This becomes a limitation when someone wants to study the features themselves.[2]

3.3 Jia et al. (2022)

Jia and co-authors built an emotion-classification system using neural networks. Their experiments were conducted on popular public datasets. While the accuracy was good, the authors also mentioned that the music samples in these datasets were produced in very different environments. Because of this variation, it was hard to understand whether the emotion came from tempo, pitch, or some other factor. The paper did not include experiments using controlled or self-made musical pieces.[3]

3.4 “Towards Unified Emotion Recognition Across Datasets” (2025)

This study examined whether a system trained on one dataset can work reliably on another. The authors tested several combinations and found that performance dropped heavily when switching datasets. They argued that differences in production style, recording quality, and emotional labelling make cross-dataset research difficult. This highlighted the need for more uniform or controlled data.[4]

3.5 Multimodal MER Survey (2025)

This survey discussed approaches that combine audio with lyrics or video to identify emotions. While multimodal systems usually perform better, they also become more complex. The authors pointed out that many papers focus on songs with lyrics. Pure

instrumental emotion recognition, which depends only on audio features, is still not explored deeply.[5]

3.6 Deep-Learning Model for MER (2021)

This paper introduced a deep-learning pipeline using CNN and LSTM layers. The experiments showed strong results on the selected dataset. However, the authors mentioned that the model required considerable computation and long training time. The paper also did not explore how specific audio features influence the final prediction, so interpretability remained limited.[6]

3.7 ADFP: Deep Feature Fusion Approach (2022)

In this study, the researchers used classical machine-learning methods and clustering algorithms. They extracted features such as spectral contrast, tempo, and MFCCs. The results changed noticeably when the dataset was normalized differently. The authors stated that the variety in musical styles made clustering less stable. This again highlighted the issue of uncontrolled data.[7]

3.8 Music Emotion Recognition: Robust Standards Study

This paper presents a detailed discussion on the problems and future directions in music emotion recognition. The authors explain that many existing MER systems depend heavily on specific datasets and fixed taxonomies, which do not always match how listeners actually perceive emotions. They highlight that emotional responses vary from person to person and also depend on the listening context. Because of this, models trained on one dataset may not work well in real situations. The paper also points out that several studies use inconsistent labels, making it difficult to compare results across researchers. The authors suggest that emotion recognition should move toward more personalized and context-aware approaches, with clearer standards for annotation and evaluation.[8]

3.9 Real-Time Music Emotion Recognition using Feature Fusion (2025)

This paper presented a system that tries to recognize the emotion of a music clip in real time. The authors combined several types of audio features, such as MFCCs, energy levels, spectral patterns, and rhythm-based features. They used a Bi-LSTM model to process these features together. The system gave good accuracy on the test dataset, but the authors also mentioned that the approach was quite complex. It required fast processing, a trained model, and more computational power than a simple machine-learning setup. Another limitation was that the music used for training came from multiple online sources, which made the data uneven in terms of sound quality and recording style. The paper did not test any small, controlled audio samples, so it was difficult to understand how individual features contribute to emotion.[9]

3.10 Research Gap

Across the studies reviewed, a common issue appears again and again. Most authors worked with large collections of professionally produced music, which often come from many

different sources and styles. Because these recordings are not controlled, it becomes difficult to understand how individual features such as tempo, loudness, or spectral brightness influence the emotional impression of a track. Several papers also relied on deep-learning models that gave good accuracy but did not clearly explain why the prediction was made, making the systems hard to interpret. Very few studies created their own musical samples or tested emotions using simple, self-composed clips. This leaves a clear gap for experiments that use controlled music created specifically for analysis. In this project, short emotion-based clips are composed in LMMS so that tempo, pitch, and overall sound can be adjusted intentionally, making it easier to study how basic DSP features relate to different emotional states.

4. Problem Statement

Most research on music emotion recognition uses large collections of songs taken from different sources. These songs are professionally produced and include many instruments, effects, and recording styles. Because of this variation, it becomes difficult to understand which part of the signal actually creates the emotional feeling. Many studies also focus mainly on accuracy using complex models, but they do not clearly show how basic sound features, such as tempo or loudness, influence the emotion itself. Very few works try to analyse music that is created in a controlled way, where the composer decides the exact tempo, pitch, and overall tone. This makes it hard to study emotions at a simple signal level.

In this project, the main problem addressed is the lack of controlled audio samples for understanding emotion through basic DSP features. By creating short musical clips manually in LMMS, it becomes possible to adjust each element of the sound and then observe how these changes affect the emotional character. This small gap—between complex datasets and simple controlled examples—is what the project tries to explore.

5. Objective

The work in this project was carried out with a few simple goals in mind. These goals are written below in a straightforward way:

1. To make a few short music clips in LMMS that express different emotions.
The idea is to keep the clips simple so that the effect of tempo and tone is easy to see.
2. To observe basic DSP features of each clip, such as how loud the sound is, how fast the beat moves, and how the frequency changes across time.
3. To compare these features across the four emotions and note any clear differences that can be linked to the mood of the music.

4. To understand how small changes in tempo, pitch, or loudness affect the emotional feel of a tune when the music is created in a controlled way.
5. To show that a free tool like LMMS can be used for learning DSP concepts in a practical manner, even without using big datasets or heavy algorithms.

6. Software used

For this project, I mainly worked with two tools, LMMS and Audacity, and honestly, I used them because they were free and easy to handle. I didn't want to deal with very advanced software since the goal was only to make a few small music clips and then look at their signal behaviour. So, these two options felt practical for the kind of work I needed to do.

When I opened LMMS for the first time, the screen looked a little confusing, but after clicking around for a bit, I started understanding what goes where. I used it mostly for drawing notes in the piano roll and changing the tempo. For each emotion, I just picked a simple sound and tried to adjust the pattern until it felt right. The software let me control things like beat speed and instrument type without needing much technical knowledge. After some trial and error, the four clips started sounding different enough for the emotional categories I wanted.

Once the clips were done, I exported each one as a WAV file and moved to Audacity. I didn't edit anything there; I only used it to check how the wave looked and how the frequencies were spread out. Switching between waveform and spectrogram was simple, and I took screenshots whenever I saw a clear pattern. Most of the time, I just zoomed in or out and tried to understand what the visual changes meant for the sound. Audacity basically helped me "see" the music instead of only listening to it.

Overall, both tools did what I needed. LMMS helped me create the sounds, and Audacity helped me look at them. Nothing too complicated, but enough to finish the experiment properly.

7. Methodology

The way I worked on this whole thing was honestly pretty simple. I started LMMS without fully knowing what I was doing, and I slowly figured things out. First, I just opened a new project and tried to see what sound comes from which instrument. After clicking around for a while, I learned how to open the piano roll. So, I began placing notes here and there and listened every time to check if it matched the emotion I wanted. I didn't follow any fixed rule. If something didn't sound right, I moved the note or changed the instrument and tried again.

For every emotion, I set the tempo first because that was the easiest thing to adjust. Happy needed fast tempo, sad needed slow, calm somewhere in between, and energetic was obviously fast. After tempo, I kept adding a few notes until it somewhat felt okay. I sometimes deleted whole patterns and made them again because it didn't match the emotion

at first. This was more like trial and error than any proper method. Once each clip sounded close to what I wanted, I exported it as a WAV file. I did this four times, one for each emotion.

After making the WAV files, I opened them in Audacity. I didn't edit the sound there. I only looked at how the waveform appeared, mostly to see if the loudness was high or low. Then I switched to the spectrogram view because it showed how the frequencies were spread. I zoomed in and zoomed out a couple of times so I could see the colours clearly, and then I took screenshots. I repeated this for all four clips. While doing this, I made small notes for myself, like "this one looks brighter" or "this one has mostly low frequencies," so that I could compare them later.

This is basically the whole process I followed: make clips in LMMS, export them, open in Audacity, look at them carefully, and take screenshots. There was no complicated setup. I just kept doing one step after another until everything was done.

8. Results and Discussion

After exporting the four clips from LMMS, each audio file was opened in Audacity to observe how its waveform and spectrogram looked. The purpose was to see whether the emotional character of each clip showed up in its basic signal behavior. Even without detailed numerical measurements, the visual patterns gave a clear sense of how the sound changed from one emotion to another.

8.1 Happy Clip

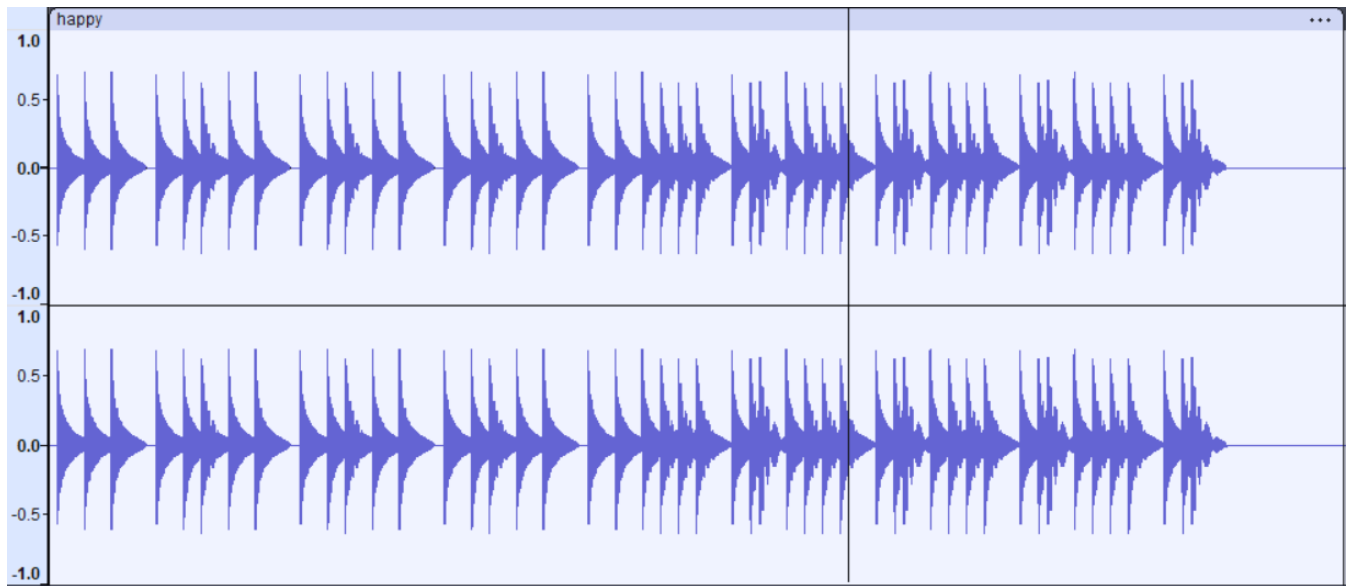


Figure 1: Waveform of Happy Clip

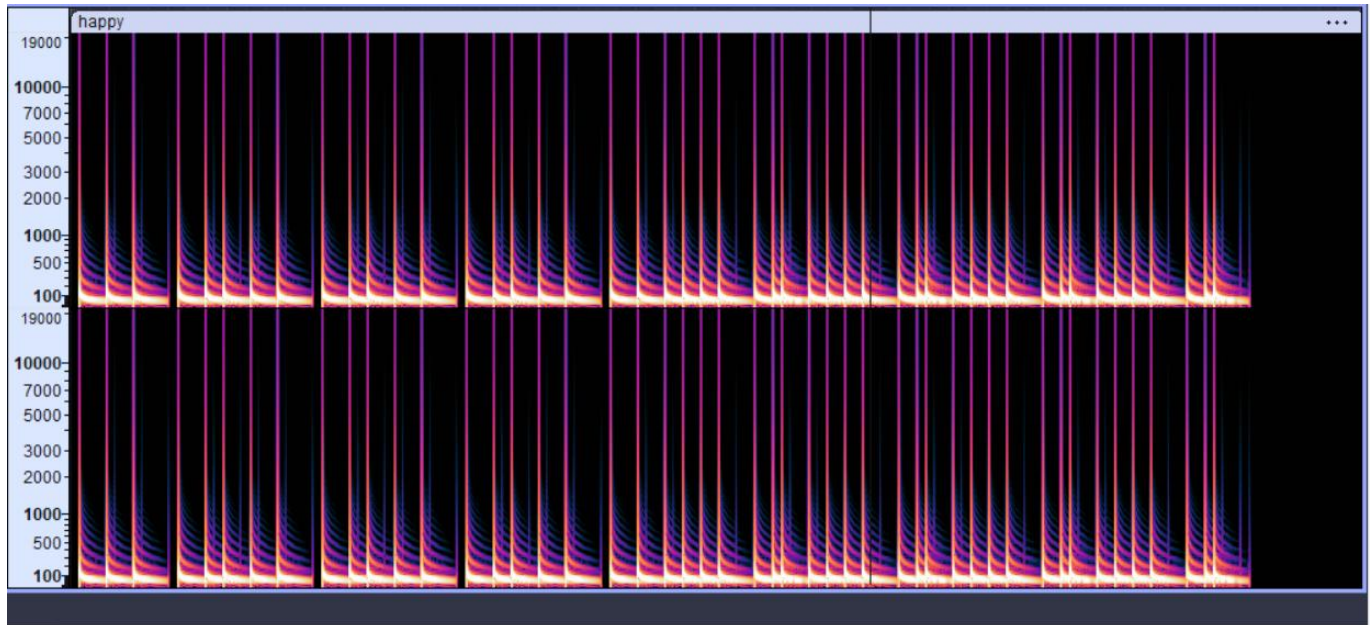


Figure 2: Spectrogram of Happy Clip

For the **happy** clip, the waveform showed regular rises and falls, giving it a lively shape. The loudness was slightly higher, and the repeated peaks suggested an active rhythm. In the spectrogram, brighter colors appeared in the mid and higher frequency ranges. The pattern looked energetic and well-spread, which matched the cheerful feeling intended during composition.

8.2 Sad Clip

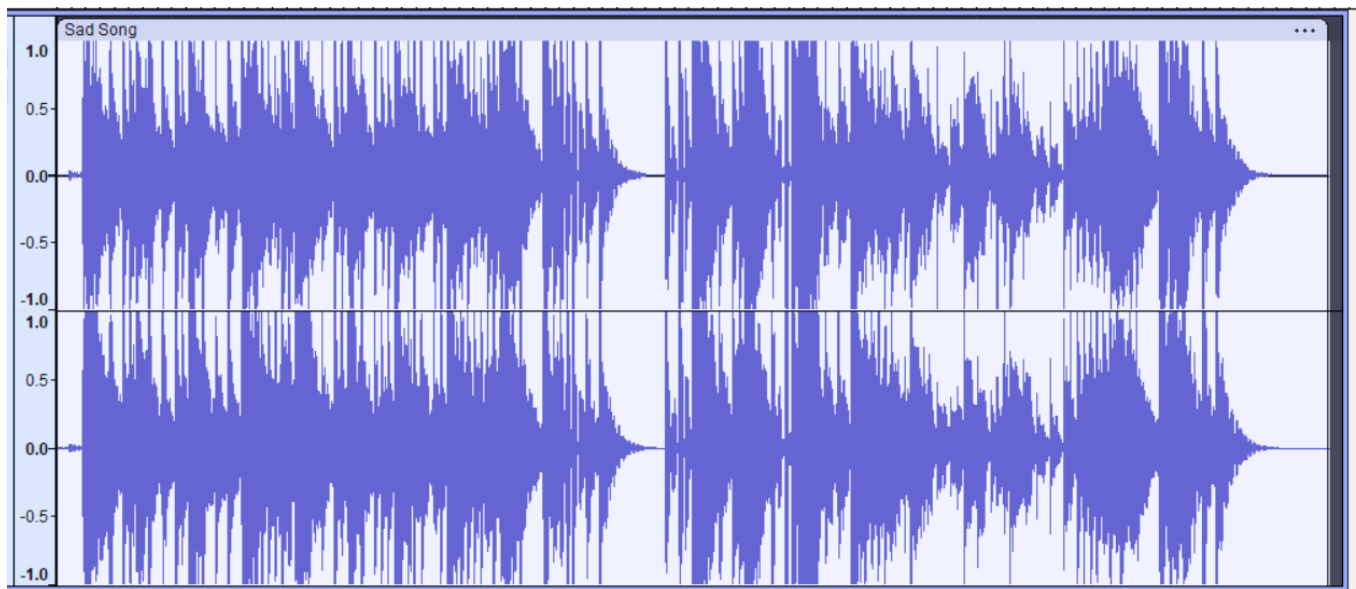


Figure 3: Waveform of sad Clip

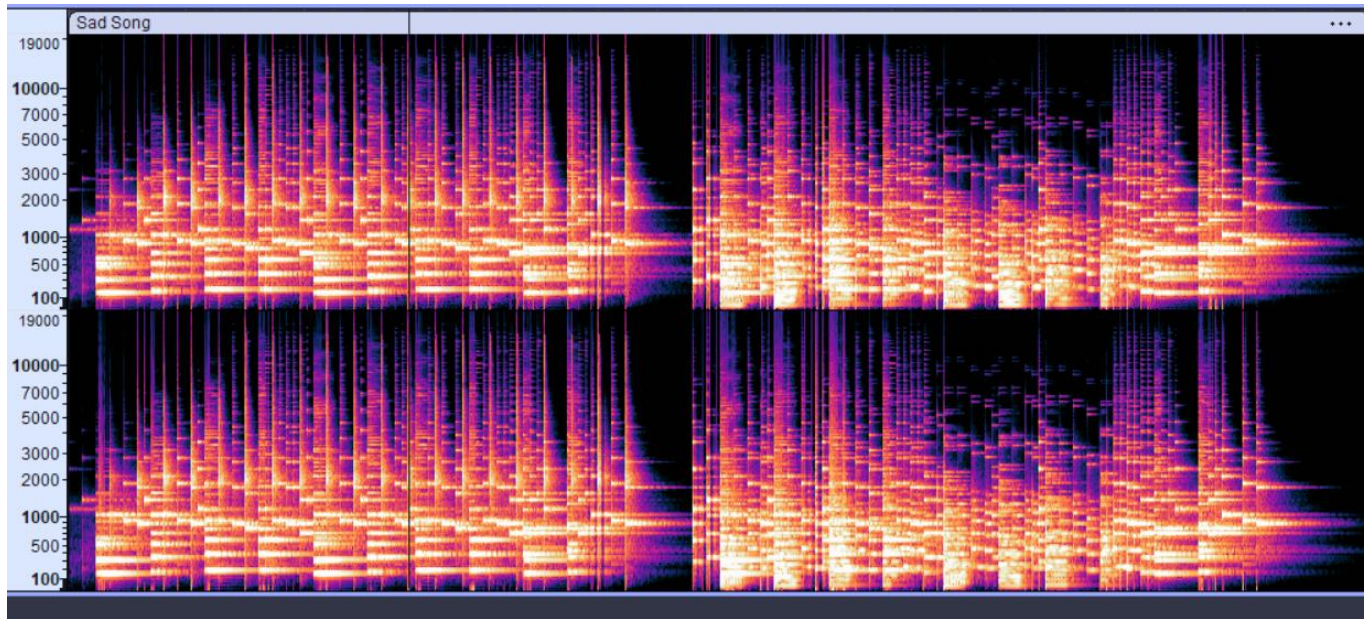


Figure 4: Spectrogram of Sad Clip

The **sad** clip displayed a smoother and heavier waveform. There were fewer sharp peaks, and the loudness stayed at a moderate level throughout. Its spectrogram mostly showed darker shades in the lower and mid frequencies, with minimal activity in the upper region. This created a more muted appearance, which fits the emotional tone of sadness.

8.3 Calm Clip

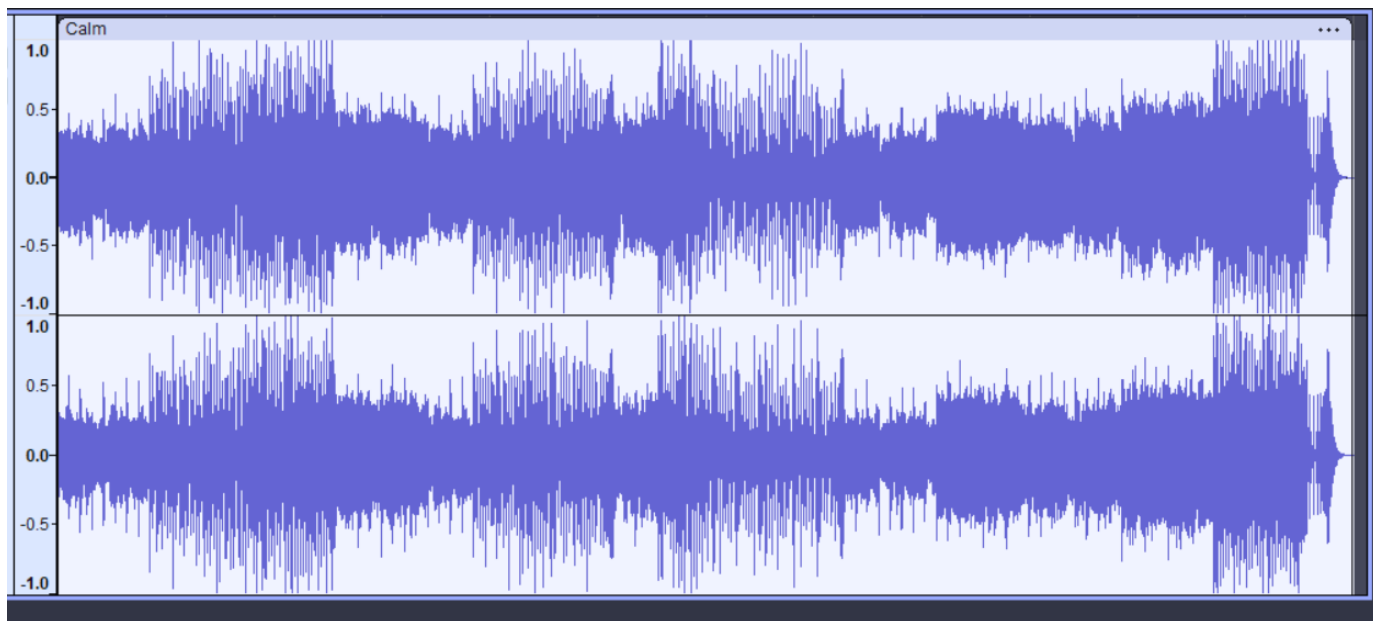


Figure 5: Waveform of Calm Clip

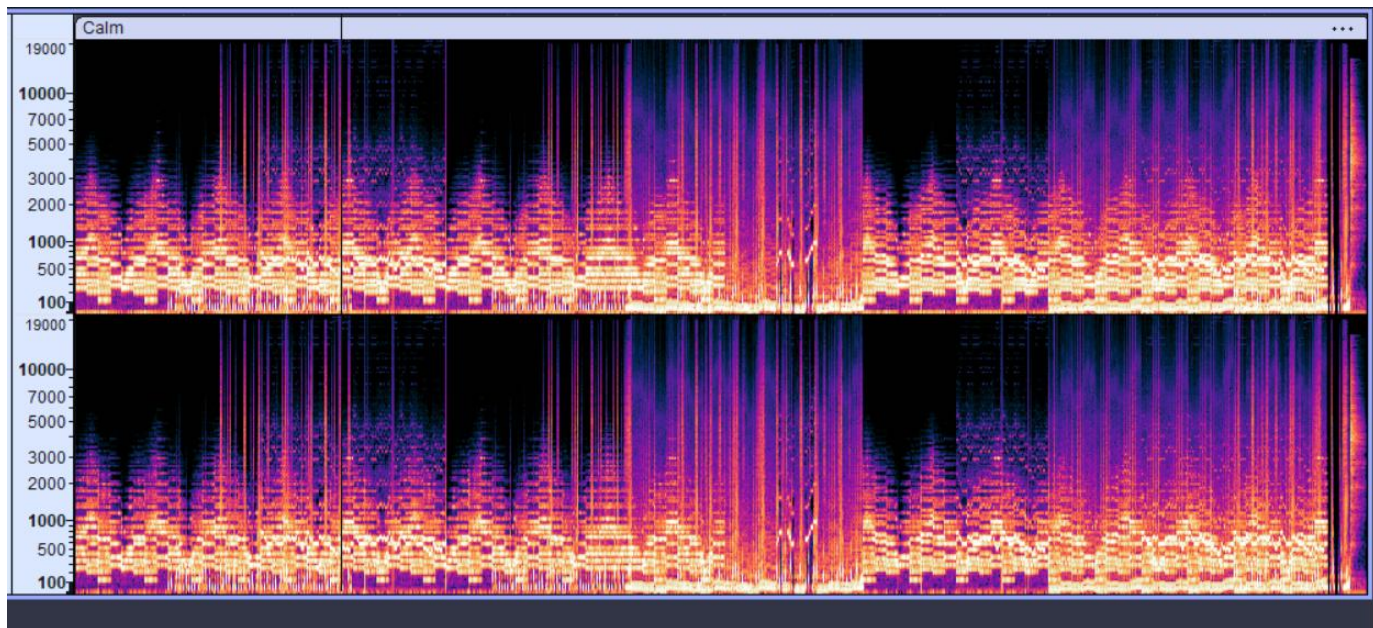


Figure 6: Spectrogram of Calm Clip

The **calm** clip had a waveform that looked even and steady. There were no sudden jumps, and the sound felt controlled. In the spectrogram, most of the color remained in the mid-range with gentle transitions. It did not show strong brightness, and the overall pattern appeared soft, aligning well with the calm mood.

8.4 Energetic Clip

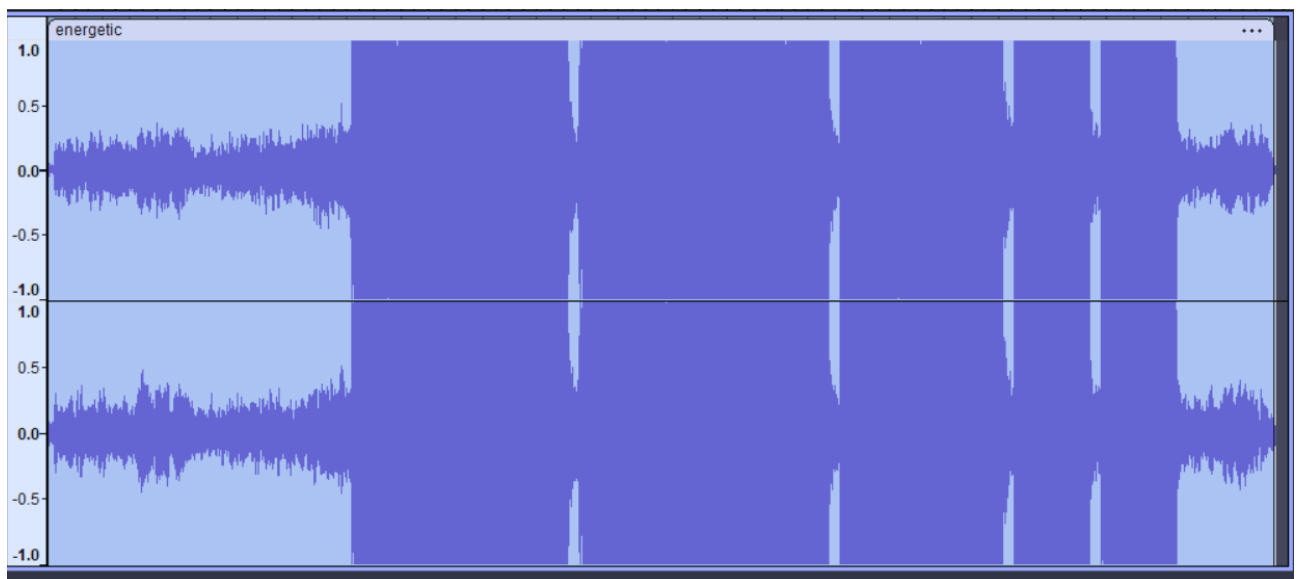


Figure 7: Waveform of Energetic Clip

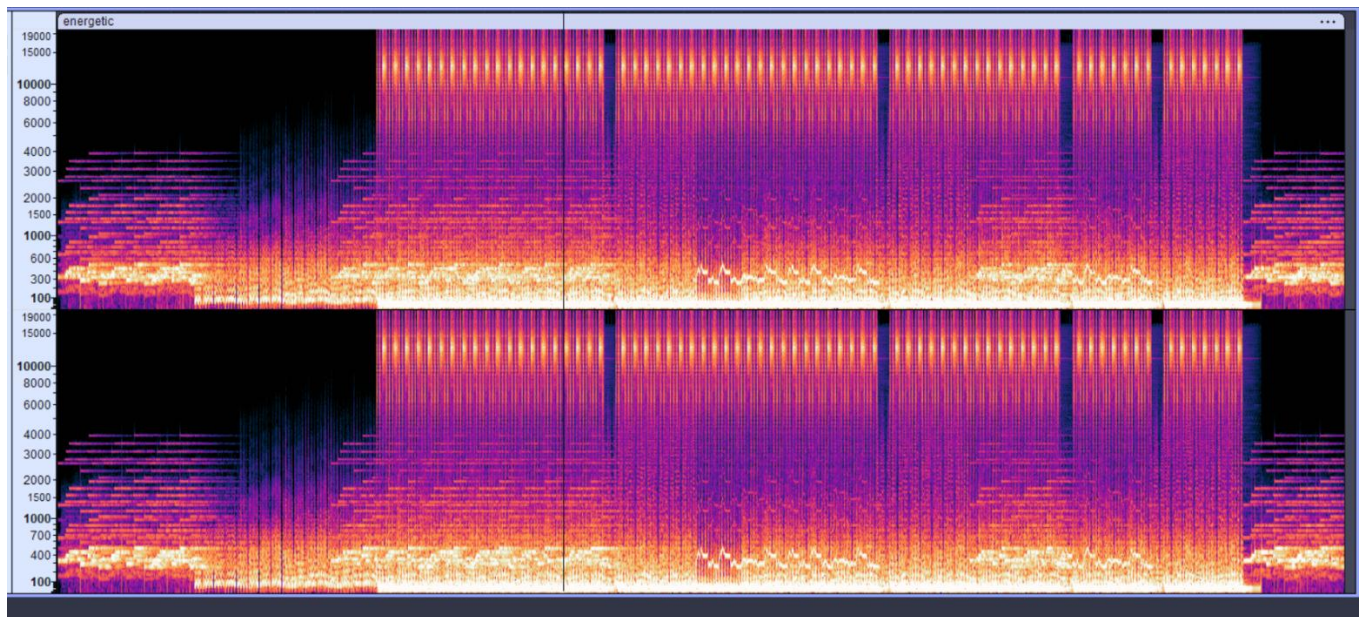


Figure 8: Spectrogram of Energetic Clip

The **energetic** clip stood out the most. Its waveform showed tall peaks and stronger loudness compared to the other clips. The spectrogram contained bright bands across a large part of the frequency range. Frequent vertical patterns indicated fast and forceful beats. This made the clip look visually intense and matched the intended high-energy emotion.

8.5 Summary Table

Emotion	Loudness Pattern	Frequency Spread (Spectrogram)	Brightness / Energy	Overall Interpretation
Happy	Medium to High, Repeating Peaks	Wide spread (Low to High frequencies)	Bright colors, active	Lively and cheerful tone
Sad	Mostly steady, fewer peaks	Mostly low-mid frequencies	Low brightness	Soft, heavy and emotional
Calm	Smooth and even	Mid-range focused	Moderate brightness	Relaxed and peaceful
Energetic	Strong tall peaks	Very broad (full band)	Very bright, dense	Fast, powerful and intense

8.6 Overall Interpretation

When all four results were compared together, the differences became quite clear. Happy and energetic clips had wider and brighter frequency activity, while sad and calm clips were more restricted and smoother. These observations show that simple changes in tempo, loudness, and instrument choice directly influence how the signal behaves, and these changes match the emotional intention of each clip.

9. Comparison With Existing Work

Compared with earlier studies, the approach used in this project is much simpler and more controlled. Many existing works rely on large music datasets collected from different sources, and several authors have pointed out that such datasets bring a lot of variation in recording style and emotional labels. **Kang and Herremans** [1] and **Jia** [3] both highlighted that this variation makes it difficult to understand the exact features that relate to emotion. In contrast, the clips in this project were created manually in **LMMS**, so the tempo, loudness, and instrument type were directly controlled, making the emotional differences clearer in the waveform and spectrogram.

Several studies, such as **Louro et al.**[2] and Huang and Zhang [6], used deep-learning models that often behave like black boxes. These models can classify emotions but do not clearly show which part of the signal influences the output. The present work focuses only on observable **DSP** features, which makes the results easier to explain. Other works, including the **ADFF** method [7] and **real-time fusion model**[9], rely on complex pipelines. In comparison, this project uses a simpler setup that still reveals clear distinctions between emotions. The controlled clips used here support the idea suggested in **Gómez-Canon et al.** [8], who noted the need for more consistent and interpretable emotion-analysis frameworks.

10. Conclusion

At the end of this project, I mainly learned how small music clips behave when they are treated as signals instead of just sounds. I made four short clips in LMMS for different emotions and opened them in Audacity to see how they looked. Most of the work was just trying things out—changing the tempo, moving notes, and listening again. I didn't follow any fixed steps, but once all the screenshots were placed together, the differences between the clips started becoming easier to notice.

When the four clips were viewed side by side, the changes were quite clear without needing any complex analysis. The happy and energetic clips showed more movement in their waveforms and looked brighter in the spectrograms. The sad and calm clips were quieter and mostly stayed in the lower or middle areas. Their shapes were much smoother too. Even though the clips were short and simple, the emotional changes still appeared in the signal in a very direct way. From this small exercise, I got a basic understanding of how things like tempo and sound choice affect the feel of a piece of music. The project was simple, but it still gave me some useful insights, and it can be expanded later with more music or a bigger method if needed.

11. References:

- [1] J. Kang and D. Herremans, “Are We There Yet? A Brief Survey of Music Emotion Prediction Datasets, Models and Outstanding Challenges,” *IEEE Trans. Affect. Comput.*, pp. 1–16, 2025, doi: 10.1109/TAFFC.2025.3583505.
- [2] P. L. Louro, H. Redinho, R. Malheiro, R. P. Paiva, and R. Panda, “A Comparison Study of Deep Learning Methodologies for Music Emotion Recognition,” *Sensors*,

vol. 24, no. 7, pp. 1–17, 2024, doi: 10.3390/s24072201.

- [3] X. Jia, “Music Emotion Classification Method Based on Deep Learning and Explicit Sparse Attention Network,” *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/3920663.
- [4] J. Kang and D. Herremans, “Towards Unified Music Emotion Recognition across Dimensional and Categorical Models,” 2025, [Online]. Available: <http://arxiv.org/abs/2502.03979>
- [5] R. Liyanarachchi, A. Joshi, and E. Meijering, “A Survey on Multimodal Music Emotion Recognition,” *ACM Comput. Surv.*, vol. ?, no. ?, 2025, [Online]. Available: <http://arxiv.org/abs/2504.18799>
- [6] C. Huang and Q. Zhang, “Research on Music Emotion Recognition Model of Deep Learning Based on Musical Stage Effect,” *Sci. Program.*, vol. 2021, 2021, doi: 10.1155/2021/3807666.
- [7] Z. Huang, S. Ji, Z. Hu, C. Cai, J. Luo, and X. Yang, “ADFF: Attention Based Deep Feature Fusion Approach for Music Emotion Recognition,” *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2022–Septe, no. 1, pp. 4152–4156, 2022, doi: 10.21437/Interspeech.2022-726.
- [8] J. S. Gomez-Canon *et al.*, “Music Emotion Recognition: Toward new, robust standards in personalized and context-sensitive applications,” *IEEE Signal Process. Mag.*, vol. 38, no. 6, pp. 106–114, 2021, doi: 10.1109/MSP.2021.3106232.
- [9] X. Hao, H. Li, and Y. Wen, “Real-time music emotion recognition based on multimodal fusion,” *Alexandria Eng. J.*, vol. 116, no. January, pp. 586–600, 2025, doi: 10.1016/j.aej.2024.12.060.