Electric Guitar Classification in Popular Music Recordings

Lennon Seiders
lseiders@umich.edu

Abstract

This paper investigates the feasibility of classifying electric guitar presence within polyphonic popular music recordings using deep learning methods. While the role and influence of the electric guitar on popular music are well-documented from historical, cultural, and musicological perspectives, quantifying its presence at a large scale remains challenging. I present a dataset derived from Billboard Hot 100 year-end charts spanning 1960 through 1999 and employ a two-stage classification pipeline. First, a pretrained audio event recognition model (YAMNet) transforms raw audio into low-dimensional embeddings. These embeddings are then passed to a neural network that performs binary classification on the presence or absence of electric guitar. Although the resulting accuracy of 74% leaves substantial room for improvement, this proof of concept demonstrates that commonly used deep learning architectures can be used to effectively identify electric guitar usage. The work also highlights the persistent need for better-labeled music datasets and suggests that with larger, more balanced, and accurately annotated data, a more robust model could be developed. Ultimately, this research provides a starting point for musicological analysis of the electric guitar's role in shaping popular music.

1. Introduction

The electric guitar has been a driving force behind popular music in the United States dating back to its inception in the early 1930s. In her work titled The Power of the Electric Guitar, Rebecca McSwain [13] lists a number of environmental factors that lead to the development and eventual popularity of the instrument:

- National electrification in the United States. Adequacy and accessibility of the relevant technological elements, such as the vacuum tube for amplification.
- Widespread acceptance of electricity as a power amenable to domestic use, including acceptance of the idea of using electricity to make music and musical instruments.
- The popularity of certain musical forms, specifically dance music and music based upon deviations from the standard European diatonic scale.
- Urbanization, bringing regional music and musicians into direct contact with one another and with an audience varied in ethnic, racial, and regional origins.
- The existence of American companies with experience in mass-producing and mass-marketing guitars. Nationwide radio broadcasting, which created a mass audience for alternative and regional music across ethnic, racial, and geographical barriers.

McSwain goes on to argue how these environmental factors, in this specific moment in history, resulted in the inception of an instrument we hear nearly every day of our lives. This concept that certain environmental factors may significantly influence an individual's potential to create any particular thing is not one that is disputed, however what these exact factors may be in a given scenario is always up to debate.

In the context of music, one might ask why a particular sound finds its way into popularity– could this be the result of cultural, economic, or political factors subconsciously affecting the creativity of the artist and the taste of the consumer? Most of the time, these factors are not entirely quantifiable, but we are often aware of them and able to explain their presence. Because the electric guitar has played such a significant role in popular music, a crucial part of studying the sound of popular music is studying the electric guitar. Just like the environmental factors that have affected the electric guitar, is the sound of the electric guitar an environmental factor that then helps shape the sound of popular music? If this is the case, we must understand how the sound of the electric guitar has changed and developed, which begs the question: How can we define and/or identify the sound of the electric guitar? This project attempts to explore the sonic identity of the electric guitar through a (seemingly) simple machine learning classification. Do the characteristics of the electric guitar's sound lie in some quantitatively-identifiable place, or just like studying the environmental influence one thing has on another, is a guess the best we can do?

Because of their purely analog nature, more traditional instruments such as the piano or drums are limited, or at least more consistent in their sound-making capabilities. Drums rely on some kind of percussive contact against their head to make a noise, and a piano will play one of twelve notes defined by its keys. Depending on the tunings of each instrument, an electric guitar may play the same twelve notes that a piano does, however it can express these notes in just as many ways as there are to strike a drum. Its strings may be picked, strummed, or contacted in any possible way to create a unique noise. On top of this, its electrification allows for another layer of endless possibilities: feedback noise, distortion from amplification, or effects pedals, for example. Because of these endless possibilities, an electric guitarist may express themselves in any way they may see fit. The result of this, however, is the lack of a consistent timbral or sonic identity for the instrument. In other words, it may be difficult to put the electric guitar into a well-defined box.

This is of course, only in regards to the sound of the electric guitar in a vacuum. In reality, the electric guitar oftentimes finds itself as one of many voices in an ensemble. This poses further questions: Has the electric guitar always maintained the same identity in polyphonic music? It is likely that the advent of sophisticated recording technologies, audio effects, or new play styles have changed how the unique voice of the electric guitar is used in a recording. A proper starting place to explore this question would be the sound itself: What, for example, makes an electric guitar sound different from a drum kit, and how can we observe that difference through data analysis?

A spectrogram, which creates a heatmap of the dominant frequencies of a sound, can display this difference between the two. The spectrograms found in Figure 1 are taken from the beginning of six different Billboard Top 100 songs from different years, where a singular instrument is playing. One can observe that the points of highest intensity on the guitar spectrograms appear in a horizontal pattern, while those of the drums appear as more circular peaks. These differences suggest that there are quantitatively-definable patterns that differentiate between the two instruments.

Because we now can identify which sound is the electric guitar in this scenario, we have the ability to better understand its role in the music, as we can answer questions regarding where and when it is used, how it is used (maybe the spectrogram of an electric guitar being used in the chorus looks very different from its use during a verse), and how this role has changed over time.

Of course, this is an over-simplified toy example. In the real world, trying to separate the sound of an instrument

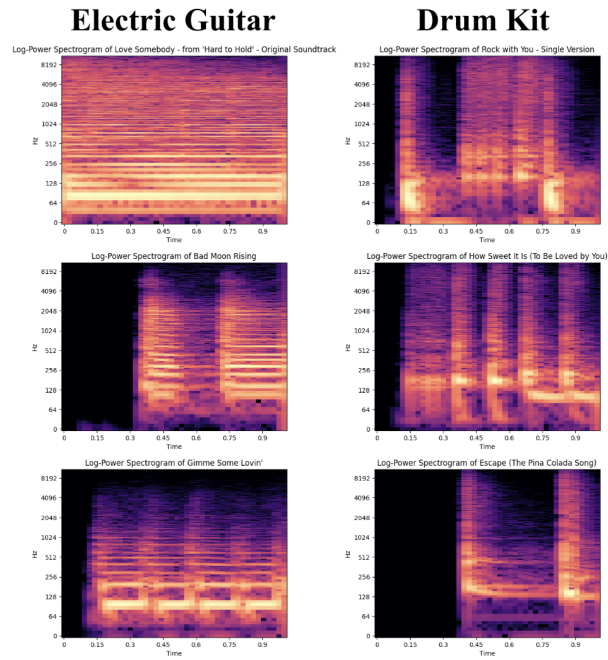

Figure 1. Caption

from polyphonic music, especially the electric guitar, is extremely difficult. The 2023 ISMIR Demixing Challenge [5] demonstrates that even the state-of-the-art performance of this task is far from perfect. In this example, the 1st place team was able to achieve a 7.46 mean signal-to-distortion ratio across all four instruments in the challenge. A signal being only seven times more present than the present distortion in a recording is far from high-fidelity. Additionally, these source separation and classification tasks rarely give specific attention to the electric guitar, despite the crucial role that it plays in western music. This project attempts to address the relative lack of research and lack of understanding of the electric guitar through its physical audio features. As illustrated above, a better understanding of the instrument's sonic identity will lead to a more sophisticated musicological analysis of the instrument and the music it contributes to. It is important to clarify that this paper does not set out to give a comprehensive set of answers to the questions I've asked above. Rather, they are being asked to provide context for the significance of exploring machine learning tasks such as this one pertaining to the electric guitar and the study of popular music.

2. Related Work

The task of instrument recognition or classification in polyphonic music has existed for a long time in the field of music information retrieval. Many machine learning methods have been employed to achieve varying accuracies across many datasets for this task, such as hierarchical clustering [4], varying neural network implementations and SVMS [7]. Most recently, deep neural networks have proven to be the most effective method to achieve state-of-the-art results for this task [1][8].

These methods however lack specificity, which can prove to be an issue for an instrument with a complex sound such as the electric guitar. Existing music information retrieval work for the electric guitar centers around the tasks of transcription and source separation. Typically, transcription involves using audio features such as spectrograms, MFCCs, or raw audio waveforms to identify and classify specific guitar notes being played [3][14]. Similarly, source separation for the electric guitar has been primarily performed using convolutional and recurrent neural networks to separate the electric guitar's specific voice from a mix [9][11].

This project uses a custom dataset comprised of *Billboard* [10] year-end popular music charts from 1960 through 1999. This dataset was created due to the lack of popular music datasets with comprehensive instrument labels paired with complete audio and song metadata. Similar data sets do exist however, such as MusiClef [16], which pairs the Musicbrainz [17] music metadata database with audio and editorial data, and the Million Song Dataset [2], which consists of approximately one million popular music recordings, their metadata (also extracted from Musicbrainz), and extracted features. Unfortunately neither these nor other available datasets were adequate for this task, because of their lack of instrument labeling and also due to their sheer size, which made them impossible to process or label with instrument data due to a lack of computational resources.

3. Methodology

3.1. Dataset: *Billboard* Year-End Charts

Because the goal of this project is to identify the sound of the electric guitar in polyphonic popular music, this called for the curation of an original dataset comprised of labeled popular music recordings. Additionally, this dataset had to be of a size significant enough to produce meaningful results while taking into account the limited computational resources at my disposal. Year-end *Billboard Hot 100* charts proved to be an ideal solution to this problem, because they were able to provide an approximate snapshot of the state of popular music in any given year without taking up an unwieldy amount of space. The final version of the dataset used in this project consists of 100 songs from each year from 1960 through 1999, resampled to 16000 Hz. In each given year, there were typically one or two songs that were not downloadable for this dataset, which resulted in a total of 3917 songs instead of the expected 4000. These songs were downloaded as .mp3 files with metadata including the song title, release name, artist and year. Preprocessing involved storing each song as three random ten-second audio samples from the recording with its corresponding metadata as a class object and adding a binary label denoting usage of the electric guitar.

To label the recordings based on their usage of the electric guitar, I initially intended to make use of the Musicbainz database and API, which contains instrument metadata. Unfortunately, the instrument labeling is far from comprehensive for the list of popular recordings from *Billboard*, so I turned to the usage of an LLM trained on vast amounts of internet data to complete the labeling task. OpenAI's GPT-4o model [15] was selected due to its broad training on extensive textual data, including music journalism, song analyses, studio notes, reviews, interviews, and reference materials that often detail the instrumentation of well-known tracks. GPT-4o was used along with the OpenAI API to label each of the songs with the following prompt:

> Song details:
> Title: {title}
> Artist: {artist}
> Album: {album}
> Year: {year}
> Does this recording use the electric guitar?
> Answer with a single word: ”Yes” only if it contains electric guitar, and ”No” otherwise.

This labeling scheme was run on the dataset multiple times to estimate the labels, resulting in an estimated 3242 total songs labeled as containing the electric guitar with a 95% confidence interval of $(3211, 3273)$. Figure 2 shows these estimates of electric guitar usage labels by year.

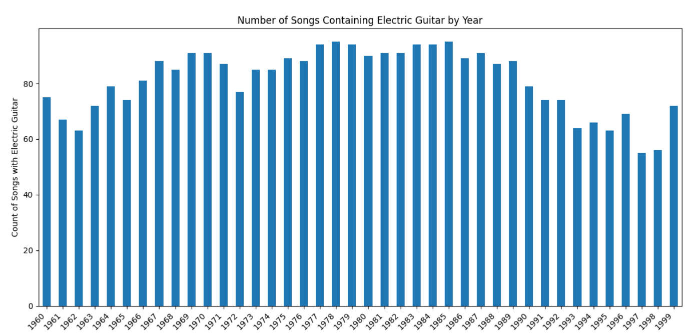

Figure 2.

Despite the dataset's approximate labeling, the sheer quantity of significant popular music recordings provides a competent deep learning architecture with ample information to make coherent inference on the presence of electric guitar in a polyphonic recording. This is demonstrated in the later Results and Discussion section after the introduction of the machine learning implementation in the following section.

3.2. Binary Classification

After preprocessing the database as raw audio samples, labels, and metadata, an initial pretrained deep neural network, YAMNet [12] is used to generate embeddings for the raw audio, which is then fed into a second neural network for binary classification. Figure 3 illustrates a high-level view of the model pipeline.

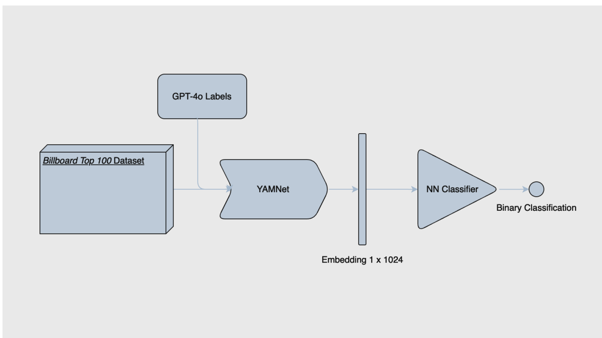

Figure 3.

A pretrained model is used to generate audio embeddings from the raw audio data for this binary classification task. This is because raw audio data is high-dimensional and complex, and a deep neural network can learn representations of the data automatically, capturing subtle patterns in time and frequency domains. Because effective deep neural networks for audio already exist and my computational resources are limited, YAMNet proves to be an ideal model to interpret the raw audio data. YAMNet is a pretrained deep net that predicts 521 audio event classes based on the AudioSet-YouTube [6] corpus using the MobilenetV1 convolution architecture. YAMNet takes Tensors of audio sampled at 16000 Hz and passes them through 27 convolutional layers and one fully-connected layer to output a 1024-dimensional embedding vector. Each song in the dataset has three ten-second samples, and each sample is converted into embeddings by YAMNet as an individual datapoint.

The vector of embeddings is then input to a neural network with three hidden layers, and finally a single binary output predicting the presence of the electric guitar in the audio. Because the dataset is not extremely large, and the number of labels denoting electric guitar greatly outweighs the number of labels denoting no electric guitar, overfitting and model specificity were two key concerns. Specificity is defined as

$$\frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}},$$

which in the context of this classification problem is a metric that measures the model's performance on songs without electric guitar. This is also called the true negative rate. A specificity score close to 1.00 would suggest that the model still does a good job of predicting songs without electric guitar as songs without electric guitar, despite being trained on less data for this case. To avoid a low specificity score and to avoid the model overfitting the training data, l2 regularization, batch normalization, and a dropout layer were employed.

4. Results and Discussion

Figure 4.

Observed in Figure 4, the highest validation accuracy achieved by the model was 74.38%, after 60 training epochs. This score however, does not balance performance on both classes very well, as it achieves a validation specificity score of 55.58% at this epoch, which suggests that the model performs only marginally better than the naive approach on songs without electric guitar. The large gap between training and validation performance also suggests overfitting. Taking this into account, the model balances these metrics more effectively around training epoch 40, where model validation specificity scores 66.90% and validation accuracy is 68.50%. The trade-offs between these metrics is illustrated in Figure 5, where it can be observed that the model eventually begins to prefer accurate classification of songs with electric guitar over songs without, which is likely due to the mismatch in data for each.

For a generic classification task, these are not high accuracy scores. Despite this, there are no real state-of-the-art results to compare the results of this project to. It can be stated however, that this model was able to somewhat successfully identify features that distinguish the presence of an electric guitar in a recording containing many instruments. Because the implementation of this model was relatively lightweight, it is reasonable to hypothesize that a deeper model trained on more data would be able to perform this task significantly more effectively, potentially to the level at which a human is capable of performing this task.

From the perspective of usefulness, in its current state this model would not be able to effectively contribute to a

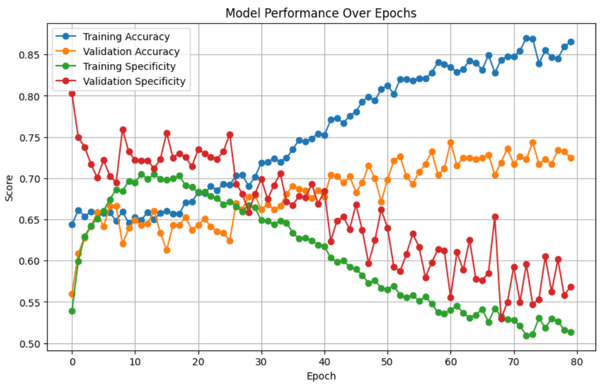

Figure 5.

musicologist's work. It does however, demonstrate that a useful model likely does exist. With an accurate classifier for the electric guitar in polyphonic music, studying the instrument could be made easier, as tasks such as archiving a large corpus of recordings could be automated.

5. Future Work

Because the model created in this project performs a simple binary classification task, its use cases are limited. Be that as it may, more complex models that perform this classification in the temporal domain of audio recordings could serve many purposes. Future work supplemental to a deeper neural network and a larger dataset could include this implementation, where the model can determine exactly when the electric guitar is playing in a song by performing the same binary classification on each moment in an audio recording. This could be used to research more of the questions discussed at the beginning of the paper, by providing answers as to when the electronic guitar is used in the song structure of popular music recordings, or how its use has changed through the decades.

This project also addresses a significant shortcoming of many song datasets: the lack of instrument metadata. I assume this is the case because other dataset curators have experienced the same issue as I did when creating my dataset, which was the lack of a method to accurately label songs on their instrument content. Musicbrainz has thousands of recordings matched to instruments, but this is simply not enough information if one wishes to train a model for a specific type of music, even chart-topping popular music. Future work to better the field of organology through music information retrieval would include labeling more recordings for their contributing instruments.

6. Conclusion

This study demonstrates the potential, yet current limitations, of machine learning-based electric guitar recognition in polyphonic popular music recordings. By assembling a custom dataset of historical Billboard Hot 100 tracks, I implemented a binary classification model to detect electric guitar presence, utilizing pretrained embeddings to handle the large amount of data given my limited computing resources. The lightweight model implemented in this project demonstrates promise that an effective classifier may be attainable given better labeling, a larger dataset, and a deeper neural network. The absence of thorough instrument annotations in large-scale music datasets remains a critical bottleneck for tasks such as this one, which leads this project to highlight the importance of richer instrument metadata. Ultimately, refining a model and dataset such as the ones introduced in this paper will provide a valuable resource for musicologists and organologists to systematically explore how the electric guitar has influenced popular music's evolving sound.

References

[1] Kleanthis Avramidis et al. "Deep convolutional and recurrent networks for polyphonic instrument classification from monophonic raw audio waveforms". In: *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2021, pp. 3010–3014.

[2] Thierry Bertin-Mahieux et al. "The million song dataset". In: (2011).

[3] Yu-Hua Chen et al. "Towards Automatic Transcription of Polyphonic Electric Guitar Music: A New Dataset and a Multi-Loss Transformer Model". In: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2022, pp. 786–790. DOI: 10.1109/ICASSP43922.2022.9747697.

[4] Slim Essid, Gaël Richard, and Bertrand David. "Instrument recognition in polyphonic music based on automatic taxonomies". In: *IEEE Transactions on Audio, Speech, and Language Processing* 14.1 (2005), pp. 68–80.

[5] Giorgio Fabbro et al. "The Sound Demixing Challenge 2023

$$\backslash$$

unicode $\{x2013\}$ $MusicDemixingTrack$". In: *arXiv preprint arXiv:2308.06979* (2023).

[6] Jort F Gemmeke et al. "Audio set: An ontology and human-labeled dataset for audio events". In: *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE. 2017, pp. 776–780.

[7] Philippe Hamel, Sean Wood, and Douglas Eck. "Automatic Identification of Instrument Classes in Polyphonic and Poly-Instrument Audio." In: *ISMIR*. Citeseer. 2009, pp. 399–404.

[8] Agelos Kratimenos et al. "Augmentation methods on monophonic audio for instrument classification in polyphonic music". In: *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE. 2021, pp. 156–160.

[9] Wen-Hsing Lai and Siou-Lin Wang. "Separation of Electric Guitar Sound based on stacked Recurrent Neural Network". In: *2020 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*. 2020, pp. 52–55. DOI: 10.1109/TAAI51410.2020.00018.

[10] Billboard Magazine. *Billboard Hot 100 Year-End Charts*. URL: https://www.billboard.com/charts/.

[11] Ethan Manilow, Gordon Wichern, and Jonathan Le Roux. "Hierarchical Musical Instrument Separation." In: *ISMIR*. 2020, pp. 376–383.

[12] Dan Ellis Manoj Plakal. *YAMNet*. 2024. URL: https://github.com/tensorflow/models/tree/master/research/audioset/yamnet.

[13] Rebecca McSwain. "The power of the electric guitar". In: *Popular Music & Society* 19.4 (1995), pp. 21–40.

[14] Meinard Müller and Tech Anssi Klapuri. "Automatic transcription of bass guitar tracks applied for music genre classification and sound synthesis". PhD thesis. Ph. D. Dissertation. Gustav-Kirchhoff-Straße 1, 98693 Ilmenau, Germany, 2014.

[15] OpenAI. *GPT-4*. https://openai.com/research/gpt-4. 2023. URL: https://openai.com/research/gpt-4.

[16] Markus Schedl et al. "A professionally annotated and enriched multimodal data set on popular music". In: *Proceedings of the 4th ACM Multimedia Systems Conference*. 2013, pp. 78–83.

[17] Aaron Swartz. "Musicbrainz: A semantic web service". In: *IEEE Intelligent Systems* 17.1 (2002), pp. 76–77.