

# **Proposal - Platform for Intelligent Music Library Management and Audio Processing**

## **Motivation and Project Overview**

The landscape of music production and audio engineering is continuously shaped by innovations in technology. Since the early days of Hip-Hop and Rap music, artists have been making use of existing musical elements to re-envision genres. In recent years, with artificial intelligence becoming widely available, it has increasingly found application in various aspects of music production and other creative fields. From sound-design, over assisting in mixing and mastering, all the way to aiding with compositional tasks. AI integration offers novel methods and tools, which can be helpful when it comes to streamlining processes and opening new horizons for artistic exploration.

The project centers around developing a distributed platform, tailored to music professionals, audio engineers and hobbyists, facilitating access to state-of-the-art neural networks for audio processing and music library management. It aims to provide a tool that simplifies the process of managing a library of music and samples, as well as extracting musical information and isolated musical elements from those audio files.

From a technical perspective, this project is an opportunity to explore and learn about key aspects of modern computing and AI in the context of audio processing. One of the primary challenges is the architectural design of a distributed system specifically for audio processing. This involves understanding and implementing a framework that can efficiently handle large-scale audio data, ensuring that the system is both robust and adaptable to different processing needs. Another significant component is the deployment of trained neural networks, specifically since running network inference is inherently resource-intensive. The challenge is to develop a system where these networks are deployed as containerized services, enabling scalability and maintenance efficiency. Simultaneously, the system must effectively utilize parallel processing, particularly with GPUs, to ensure that the intensive demands of neural network inference do not compromise processing time or resource efficiency. Lastly, the UI/UX design of the platform is an important consideration. It needs to be intuitive and user-friendly, ensuring that users can easily navigate and utilize the advanced features of the system without unnecessary complexity.

## **Features and Project-Scope**

### **Music / Sample Collection**

- Users can upload music
- Users can manage uploaded library
- Users can manage extracted stems and samples
- Users can export collection for local usage, e.g. DAW
- Users are able to play back music, extracted stems and samples

### **Source Separation / Stemming**

- Users can separate uploaded music into individual stems

- Using [Demucs v4](#) (Facebook/Meta Research - Defossez)

### **Sample Slicing / Classification**

- Users can slice extracted stems further into individual samples (drum hits, vocal chops, synth hits etc.)
- Sliced samples are classified/tagged automatically
- What network/pre-trained model will be used is still tbd, either build/train own, or use transfer-learning with existing models

### **Audio to Midi (Vocals/Bass/Melodic)**

- Users can convert extracted stems to midi
- Using [Basic Pitch](#) (Spotify)

### **Audio to Text (Vocals)**

- Users can extract lyrics / text from vocals
- Using [OpenAI Whisper](#) or one of the several other open-source models

### **Implementation Strategy**

The implementation strategy for this project has already begun with exploring various processing approaches and determining an optimal system architecture, likely leaning towards a microservice architecture for the backend. This will involve setting up individual services for each processing task, integrating a task queue for efficient workload management, and establishing an API Gateway to facilitate access of the web-client to the services. As for the front-end, the initial focus will be on creating a minimal viable product that meets all technical requirements listed above. Once a functional foundation has been established the UI/UX will be refined.

### **Future Work and Outlook**

Upon completion, this project opens up possibilities for further enhancements. Potential improvements could include adding advanced music information retrieval tasks and integrating VST/Clap audio plugins with the existing backend infrastructure. Another area for exploration could be scaling the system to support more users and handle larger data volumes. A particularly interesting aspect of scaling involves considering cloud infrastructure deployment and evaluating its cost implications, given their significance in neural network production deployment.

Additionally, the project could serve as a foundation for my master's thesis or future projects. Potential areas of exploration might include investigating general approaches and optimizations for distributed systems in audio processing, or the development of novel neural networks for audio analysis.