

Senior Comprehensive Project Proposal: AI Powered Music Making Website for Non-Musicians

James Wang

Jwang2@oxy.edu

Occidental College

1 Introduction

The world of music creation has traditionally been dominated by individuals with technical expertise and specialized training. Producing high-quality songs or beats often requires knowledge of music theory, composition, and digital tools for editing and mixing. For those without formal musical education or experience, the barriers to entry can be daunting. The creative energy that many people possess often remains untapped due to a lack of access to simple yet powerful tools that lower the threshold for music creation.

In recent years, technology has gradually begun to democratize music production by making recording software more accessible and affordable. However, many existing solutions still cater to experienced musicians, offering features that can overwhelm beginners. There's a gap between comprehensive professional digital audio workstations (DAWs) and the creative experience that newcomers need. As a result, many people with musical ideas struggle to translate their concepts into complete compositions, leaving a wide audience under-served.

Moreover, the gap between musicians and listeners often fosters a perception that meaningful music creation is reserved for the highly skilled. However, creativity is universal and should be inclusive. Providing accessible tools can empower individuals to bridge this gap and cultivate their creative potential.

This comprehensive project seeks to bridge this divide by developing an accessible music creation software designed specifically for less experienced or non-musicians. This intuitive software will guide users through the songwriting and beat-making process, offering tools to create harmony, rhythm, and melody while simplifying the production process with features like equalizers and compressors. By allowing users to use text input to create different tracks, the project aims to walk users through music creation step by step, making the process enjoyable and engaging.

Ultimately, this project aligns with computer science objectives by offering an innovative solution that merges sophisticated algorithms for musical composition with an accessible interface. It addresses a significant societal need by closing the gap between musicians and listeners, democ-

ratizing music creation in a way that has never been done before.

2 Technical Background

To realize the vision of making music creation accessible for less experienced or non-musicians, it's essential to incorporate several foundational concepts from the fields of digital music production and computer science. This section provides an overview of the crucial technologies that will form the basis of the project. By understanding Digital Audio Workstations (DAWs), MIDI technology, Digital Signal Processing (DSP), virtual instruments, and web development technologies, we'll establish a framework that allows users to produce quality music tracks. This background will outline each of these components and describe their specific relevance to our software project.

2.1 Musical Instrument Digital Interface (MIDI)

Musical Instrument Digital Interface (MIDI) is a technical standard that describes a protocol, digital interface, and connectors, allowing a wide variety of electronic musical instruments, computers, and other related devices to connect and communicate with one another. MIDI does not transmit audio—it transmits "event messages" such as the pitch and intensity of musical notes to play. In this project, MIDI will be pivotal for translating user inputs (e.g., via text commands) into music, allowing users to generate and manipulate music sequences without needing to perform with physical instruments.

2.2 Digital Signal Processing (DSP)

Digital Signal Processing (DSP) involves the numerical manipulation of signals, primarily for the processing of audio signals in this context. DSP techniques are integral to effects such as reverb, equalization, and dynamic range compression in music production. For this project, understanding and implementing basic DSP will enable the creation of essential audio effects, which are crucial for enhancing the musical output and ensuring the produced

tracks meet a baseline quality standard expected in modern music.

2.3 Virtual Instrument

Virtual instruments are software simulations of traditional musical instruments and often play a crucial role in music production, especially when access to physical instruments is limited. In this software, virtual instruments will provide users with sound options, enabling them to compose and modify music directly within the app. The app will also enable web-based plug-ins and VSTs for use to use outside sound to create their music. This feature is particularly important for playback and review, allowing users to hear and refine their compositions in real-time.

2.4 Digital Audio Workstation

Digital Audio Workstations (DAWs) are software platforms used for recording, editing, mixing, and producing audio files. DAWs facilitate complex music production processes and are essential tools for contemporary music creation. In this project, the aim is to create a simplified, accessible version of a DAW. The software will offer key functionalities of traditional DAWs, including but no limit to track management, Virtual instrument, MIDI editing, MIDI record and playback, and effect application, but will be designed specifically for users without prior music production experience. By focusing on simplicity and accessibility, this project will help bridge the gap between amateur music enthusiasts and professional production tools.

2.5 Web Development

Integrating these components into a cohesive web-based application requires a robust understanding of web development technologies. This project will utilize HTML, CSS, and JavaScript for the frontend to create an intuitive user interface. React will be used for efficient UI management. The backend, likely built on Node.js, will handle API requests, interact with a database for storing user tracks and settings, and manage the audio processing tasks.

3 Prior Work

In the realm of music creation, several tools have significantly shaped how both amateurs and professionals engage with music digitally. Here's an exploration of established platforms as well as similar initiatives in other creative fields:

3.1 Web-Based Digital Audio Workstations (DAWs)

Several web-based Digital Audio Workstations (DAWs) have shaped how amateurs and professionals engage with music production. Soundtrap, owned by Spotify, offers an intuitive, cross-platform interface with collaborative features. Its accessibility makes it an ideal tool for educational settings and casual music enthusiasts alike. Similarly, Soundation provides powerful features that mirror traditional desktop DAWs like FL Studio and Ableton, while offering the convenience of browser-based access. Such tools are helping to blur the lines between professional and amateur music production.

3.2 Production Technologies for Non-Musicians in Music Education

Technology-based music creation tools have proven invaluable in empowering non-musicians to engage in learning and creating music. According to Pendergast and Robinson (2020) [3], secondary students who participate in music outside of formal education express a strong preference for small-group learning environments with a focus on technology-enabled music composition classes. Their research indicates that these students are more interested in courses that emphasize personalized learning and creativity, such as classes centered on music production with technology and popular music groups.

Walzer's article [5], "Software-Based Scoring and Sound Design: An Introductory Guide for Music Technology Instruction," in *Music Educators Journal* explores how virtual instruments, sequencers, loops, and software synthesizers make sound design accessible. Through structured, discovery-based learning environments, students can hone their creative skills even if they lack traditional music backgrounds.

Williams (2012) [6] identifies "non-traditional music" (NTM) students as a unique group who are largely disengaged from traditional music education but thrive in technology-based music classes (TBMCs). By incorporating virtual instruments and technology-based recording tools, TBMCs can engage NTM students in composing, recording, and performing. Many of these students have an active music life outside of formal education and aspire to careers in the music industry. Technology-based music education can thus help them transition from passive listeners to active creators.

Hillier, Greher, et al. (2015) [1] examined the impact of touch-screen technology and iPads in a music program for adolescents with autism spectrum disorders. Their findings show that the intuitive interface and tactile engagement of iPads enabled participants to express themselves musically

and reduce their stress and anxiety. This underscores the potential for technology to make music creation more inclusive and therapeutic, even for non-traditional learners.

These studies collectively emphasize that technology in music education can bridge the gap for non-musicians, providing accessible tools and learning environments that nurture creativity and promote engagement.

3.3 Other Related Fields

Democratizing creative processes is not limited to music. For instance, Tinkercad, a web-based 3D modeling and CAD tool, has made 3D design approachable for beginners through intuitive features and simple workflows. BobRossLIVE (formerly Bobosa) brings virtual sculpting to a wider audience with its tactile interface, accessible to those without formal art training. Similarly, Procreate, a digital painting app, allows both amateurs and professionals to produce digital artwork on par with desktop software. These platforms illustrate how technology can open creative expression to those without traditional training.

In the realm of art education, artificial intelligence (AI) has begun to make significant inroads, particularly in disciplines like architectural painting. According to a study by Li and Zhang (2020) [2], AI technologies, including lightweight deep learning models and the Limited Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm, have been effectively used to teach complex art concepts. This approach has not only enhanced the learning experience but also achieved high levels of accuracy in training and testing phases, demonstrating AI’s potential to significantly improve educational outcomes in art. These advancements in AI are pivotal for introducing more nuanced and precision-based teaching methodologies in art education, similar to the transformations seen in music technology.

4 Ethical Consideration

AI music generation tools, including those powered by advanced APIs like GPT (from OpenAI) or Stability, risk unintentionally perpetuating cultural biases inherent in their training data. This concern arises because these models, and their developers, often do not have complete transparency about the specifics of the data used for training. Typically, AI models are reflective of the data they are exposed to, and if this data primarily encompasses Western music or popular genres, the diversity of the music generated by the AI is inherently limited. This limitation can perpetuate and even exacerbate existing disparities within the music industry, where certain styles and communities are already marginalized. Although ethnomusicology is a significant area of study within musicology, the overwhelming focus of most AI music models tends to lean towards Western music

theory. This focus shapes the output of the AI, potentially sidelining a rich spectrum of underrepresented musical genres and traditions.

In discussing data bias in AI music generation tools, it’s essential to consider how cultural stereotypes can influence the emotional interpretation and representation of music. A pivotal study [4] that sheds light on this issue investigated the emotional responses to various music genres and their associated cultural stereotypes. The research, examining genres such as Fado, Koto, Heavy Metal, and Hip Hop, demonstrated that listeners often project stereotypical emotions of a culture onto its music. For instance, listeners associated peace and calm with Koto music and Japanese culture, and anger and aggression with Heavy Metal and its corresponding culture. This phenomenon, explained through the stereotype theory of emotion in music (STEM), suggests that an emotion filter simplifies the assessment process for unfamiliar music genres, leading to stereotyped emotional responses (Study on Emotional Responses to Music Genres and Cultural Stereotypes). This insight is critical for AI music generation, as it highlights the risk of AI systems perpetuating these stereotypes, particularly when trained on data that does not adequately represent the diversity of musical expressions and the complexities of cultural backgrounds. Such bias can distort the AI’s music generation capabilities, emphasizing the need for developers to incorporate a wide and diverse range of cultural and musical inputs to mitigate these biases.

AI music generation tools, while offering unprecedented creative possibilities, also risk perpetuating cultural biases due to the limitations of their training data. Recognizing this challenge, I aim to mitigate biases by exploring different models available on Hugging Face, reintroducing this platform as a valuable resource. Hugging Face hosts an array of pre-trained language and music models developed by diverse communities. By testing and comparing these models’ behavior, we can identify those capable of translating text prompts into MIDI while offering broader cultural representation. Our approach includes curating varied training data reflective of global musical traditions to reduce the inherent Western-centric focus. Moreover, feedback loops involving ethnomusicologists will refine the models by identifying gaps and cultural nuances overlooked in conventional datasets. This proactive methodology will ensure the AI system not only captures a richer spectrum of musical genres and traditions but also reduces the risk of reinforcing cultural stereotypes, ultimately fostering an inclusive and equitable music creation process.

5 Methods

In an era where technology is transforming creative industries, this web-based digital audio workstation (DAW)

project aims to bridge the gap between musicians and non-musicians by providing an accessible platform for music creation. Leveraging modern frameworks and APIs, the DAW will offer a unique environment that blends intuitive text-based commands with powerful virtual instruments and audio effects. By employing technologies like React for the frontend, Python for the backend, and the JUCE framework for digital signal processing (DSP), this project will deliver a versatile and scalable solution. Incorporating advanced tools, including the OpenAI API for creative suggestions and web-based plugin support, the DAW will empower users to craft tracks and apply professional-grade effects effortlessly. This proposal outlines the comprehensive methods that will be employed to design, develop, and refine this innovative software, creating an environment where musicians at every skill level can explore their creativity.

5.1 Frontend Development

The frontend of the DAW will feature an interactive chat box designed using React. This component will guide users through the music creation process in a conversational format. Initially, the chat box will prompt the user to specify the type of music they want to create. The user input can be relating to genres, like rock, pop, or jazz; or it can be ambient, experimental music. Following this, it will ask users to describe their desired melody. Users' responses will be converted into MIDI tracks by the backend. Further, the chat box will inquire about beats and rhythms, using these inputs to generate drum tracks. Lastly, it will request descriptions of chords or specific instruments like piano or guitar to create harmonies. The chat box will also wait for the user to instruct it to download the final composition, providing a streamlined and interactive user experience.

Additionally, the frontend will offer MIDI editing and track management functionalities. A dedicated track interface will list tracks according to their respective instruments, allowing users to play and stop audio playback. Clicking on individual MIDI clips within a track will enable editing of MIDI velocity, timing, and pitch, as well as randomization of velocity for creative variation. Both the chat box and track interfaces will incorporate a time grid to indicate the time signature and key signature, helping users maintain an organized workflow.

5.2 Utilizing API

System stability will be prioritized to handle the data flow seamlessly between the chat box and backend. Text inputs from the frontend will be processed via the OpenAI API to translate user prompts into MIDI information. This integration will utilize the API's language processing capabilities to interpret melodies, rhythms, and instrument re-

quests into structured MIDI commands. The backend will efficiently communicate with the API to receive suggestions for melodies, rhythms, and harmonies, helping guide users through the composition process.

Additionally, I will explore different language models on the Hugging Face platform to evaluate their efficacy in translating text prompts into MIDI data. Testing various models will help identify those best suited for accurately interpreting musical descriptions. I will experiment with fine-tuned models to understand how each behaves with varying prompt styles, ultimately choosing the one that provides the most relevant and musically accurate outputs. This comparative exploration will offer valuable insights into the diversity of language models and their potential application within the DAW.

5.3 Backend

The backend, developed in Python, will serve as the DAW's processing hub, handling user inputs and transforming them into structured musical outputs. As users interact with the chat box, their descriptions of melodies, beats, and harmonies will be received by the backend, where Python scripts will process these text inputs into MIDI commands. Using natural language processing (NLP) techniques and trained models, the backend will interpret the descriptive language to generate accurate musical elements. These elements will be combined to create MIDI sequences that correspond to the user's intent.

The backend will utilize the OpenAI API and various Hugging Face models to assist in interpreting and translating user prompts. Python will handle API communication, passing user inputs to these models and receiving structured musical suggestions in return. The backend will then convert these suggestions into a format compatible with the audio processing and synthesis modules.

Once MIDI data has been generated, the backend will apply audio effects and virtual instruments to enhance the musical output. Python libraries such as Mido or pretty_midi will be explored for MIDI processing, while the JUCE framework will handle digital signal processing (DSP). The backend will allow users to select web-based plugins for specific effects or instruments, applying them directly to the generated tracks. The backend will also securely store user compositions and plugin selections for future use, ensuring consistent data retrieval and playback. API calls and user data will be processed in secure environments, with safeguards to prevent unauthorized access. A modular design will be employed to ensure the backend can handle new feature additions or changing user demands. Each functional module (text-to-MIDI conversion, API integration, audio processing) will remain loosely coupled, promoting independent updates and scaling.

5.4 Virtual Instrument and DSP

Utilizing the JUCE framework, the project will develop a variety of virtual instruments and DSP effects. This section will elaborate on how the backend uses JUCE to create authentic sounding instruments and effects from the MIDI data generated by user inputs. The integration of JUCE enables the DAW to produce high-quality audio outputs, turning text-based descriptions into rich musical compositions.

5.5 Web-Based Plug-in

The DAW will support web-based plugins, allowing users to enhance their music with additional effects and instruments. These plugins can be accessed and managed through the chat box, providing users with options to further customize their tracks according to their creative preferences.

6 Evaluation Metrics

The evaluation of the music creation software will focus on user testing among beginner music students and obtaining feedback from music production professors. By involving beginner music students in structured user testing, the goal is to measure the effectiveness of the software in facilitating creative music production and enhancing learning outcomes. Specific metrics include the ease of understanding musical harmony, rhythm creation, and utilizing basic production tools like the equalizer and compressor. Additionally, feedback from music production professors will provide critical insights into whether the software can be a valuable teaching tool for non-musicians. Their assessment will help determine if this platform can bridge the gap between musicians and beginners, fostering a more inclusive approach to music education. Ultimately, combining student feedback with expert evaluation will ensure that the software is both accessible and effective, offering a learning tool that is beneficial for non-musicians and adaptable for classroom use.

6.1 Web-Based Plug-in Support

Enabling web-based plugin support will enhance the DAW's functionality and flexibility. Different types of plugins will be considered, including VSTs (Virtual Studio Technology) and custom audio effects. A secure API will manage plugin integration to ensure consistent performance and allow users to configure and select plugins easily from the frontend. Security measures will be employed to safeguard data while ensuring compatibility with various plugins.

6.2 Virtual Instrument and DSP

The JUCE framework will be employed to develop a range of virtual instruments and audio processing features. Specific DSP effects, such as reverb, distortion, and delay, will be explored to provide rich sonic possibilities. Synthesizers and samplers will also be implemented for unique virtual instruments. User interfaces will be customized for these instruments and effects to offer intuitive control and creative experimentation. JUCE's cross-platform support will facilitate efficient development and allow the instruments to integrate into the DAW, providing a versatile environment for users to craft unique soundscapes.

7 Timeline

Week 1-2 (06/01 - 06/14): Learn React basics and explore MIDI editor libraries.

Week 3-4 (06/15 - 06/28): Learn full stack database concepts and design API contracts for the project, and explore different models on hugging face.

Week 5-6 (06/29 - 07/12): Finalize front-end design, and implement API mockups using tools like Postman or Swagger.

Week 7-8 (07/13 - 07/26): Implement core API features like authentication and MIDI data processing, and start building a basic front end that interacts with these APIs.

Week 9-10 (07/27 - 08/09): Design the data flow and finalize the software architecture, and organize the codebase.

Week 11-12 (08/10 - 08/23): Finalize the back-end processes.

Week 13-14 (08/24 - 09/06): Finish creating the front end and back end.

Week 15-16 (09/07 - 09/20): Create virtual instruments and a sampler.

Week 17-18 (09/21 - 10/04): Conduct user testing and evaluation.

Week 19-20 (10/05 - 10/18): Conduct a second round of user testing and evaluation.

Week 21-22 (10/19 - 11/01): Update the software based on user feedback.

Week 23-24 (11/02 - 11/15): Prepare documentation and

presentation materials.

References

- [1] Hillier, Ashleigh et al. “Music, technology and adolescents with autism spectrum disorders: The effectiveness of the touch screen interface”. In: *Music Education Research* 18 (Aug. 25, 2015). DOI: 10.1080/14613808.2015.1077802.
- [2] Li, Jing and Zhang, Bingyu. “The Application of Artificial Intelligence Technology in Art Teaching Taking Architectural Painting as an Example”. In: *Computational Intelligence and Neuroscience 2022* (May 17, 2022), p. 8803957. ISSN: 1687-5265. DOI: 10.1155/2022/8803957. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9129932/> (visited on 05/10/2024).
- [3] Pendergast, Seth and Robinson, Nicole. “Secondary Students’ Preferences for Various Learning Conditions and Music Courses: A Comparison of School Music, Out-of-School Music, and Nonmusic Participants”. In: *Journal of Research in Music Education* 68 (June 22, 2020), p. 002242942093182. DOI: 10.1177/0022429420931826.
- [4] Susino, Marco and Schubert, Emery. “Cultural stereotyping of emotional responses to music genre”. In: *Psychology of Music* (2019).
- [5] Walzer, Daniel A. “Software-Based Scoring and Sound Design”. In: *Music Educators Journal* 103.1 (2016), p. 19. ISSN: 0027-4321. URL: https://www.academia.edu/27823736/Software_Based_Scoring_and_Sound_Design_An_Introductory_Guide_for_Music_Technology_Instruction (visited on 05/10/2024).
- [6] Williams, David. “The non-traditional music student in secondary schools of the United States: Engaging non-participant students in creative music activities through technology”. In: *Journal of Music, Technology and Education* 4 (Feb. 16, 2012), pp. 131–147. DOI: 10.1386/jmte.4.2-3.131_1.