Lab 2 - CrossFade Product Description

Hyacinth Abad

CS411W

Thomas J. Kennedy

12/01/2023

Section 1 & 2

1. Introduction

Music production is often a complex endeavor, typically demanding the expertise of a professional music producer and the use of high-end equipment such as a mixing board, synthesizer, and digital audio workstation.

In general, aspiring musicians at the entry level lack access to such advanced technology and may not be familiar with its functionalities. Moreover, some of this equipment can be quite costly, reaching up to $10,000, and is not user-friendly for beginners. Consequently, individuals starting out in music production often find it challenging to afford these tools. This financial barrier poses difficulties for emerging musicians to create and share their work.

This challenge extends beyond entry-level musicians to encompass practicing musicians, composers, artists, and music students. Without the essential equipment, these individuals also face obstacles in the creation and sharing of their musical endeavors. Recognizing this need, there is a demand for software that simplifies the technical aspects of music production and sharing.

CrossFade is ideal for those lacking expensive equipment but harboring the passion to create and share music. Unlike the intricate technology found in traditional recording studios, CrossFade is designed to be accessible to the general public and user-friendly for musicians at all skill levels. With CrossFade, gone are the days when a professional music producer or a dedicated recording studio was a prerequisite for producing and sharing one's music.

1.1 Purpose

The purpose of this document is to outline the development of CrossFade, an innovative software solution designed to address the high entry barriers and technical challenges faced by aspiring musicians in the music industry. CrossFade aims to provide a user-friendly platform

accessible to musicians at all levels, eliminating the need for expensive and complex music production equipment. This program intends to empower musicians, composers, artists, and music students by offering a simplified yet powerful tool for creating and sharing music. The primary goal is to revolutionize the music production landscape, making it more inclusive and removing the dependency on professional expertise and costly recording studio setups.

1.2 Scope

The music industry presents significant barriers for aspiring musicians, as entry often demands technical equipment like MIDI-producing tools and soundboards, which many lack. The complexity of music production, requiring expensive equipment and professional knowledge, contributes to burnout among newcomers. CrossFade aims to address these challenges by offering accessible software that simplifies the music production process. Unlike traditional recording studios with costly equipment, CrossFade is designed for musicians at all levels, providing an affordable and user-friendly alternative for producing and sharing music, eliminating the need for extensive technical expertise or substantial financial investment.

1.3 Definitions, Acronyms, and Abbreviations

Convolutional Neural Network (CNN) - Deep learning algorithm which can differentiate one image from another by assigning weights and biases to different aspects of the images. It is used in audio to differentiate different frequencies in a visual format.

Deep Learning - Subfield of machine learning which uses neural networks to solve complex problems. Learning comes directly from the data, instead of being hand-engineered by humans.

Keyboard - An electronic piano used to produce sound and MIDI information.

Monophony - A phrase of music in which only a single voice is played at a time.

Musical Instrument Digital Interface (MIDI) - A communications protocol used to connect physical and virtual music devices and instruments. MIDI files store note information which can be used to trigger instruments and devices.

MusicXML - A markup language format used to interchange and distribute digital sheet music.

Polyphony - A phrase of music in which more than a single voice is played at a time

1.4 References

Automatic Music Transcription and Ethnomusicology: A user study. (n.d.). Retrieved from https://diva-portal.org/smash/get/diva2:1474663/FULLTEXT01.pdf

H. Takeda, T. Nishimoto and S. Sagayama, "Rhythm and Tempo Analysis Toward Automatic Music Transcription," 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, Honolulu, HI, USA, 2007, pp. IV-1317-IV-1320, doi: 10.1109/ICASSP.2007.367320.

Huang, Z., Jia, X., & Guo, Y. (2019, June 29). State-of-the-art model for music object recognition with Deep Learning. MDPI. Retrieved February 8, 2023, from https://www.mdpi.com/2076-3417/9/13/2645

Jovanovic, J. (2015, February 2). How does Shazam work? music recognition algorithms, fingerprinting, and processing: Toptal®. Toptal Engineering Blog. Retrieved February 8, 2023, from https://www.toptal.com/algorithms/shazam-it-music-processing-fingerprinting-and-recognition

Scarlatos, L. L. (n.d.). Continuous media. Audio. Retrieved March 1, 2023, from https://www3.cs.stonybrook.edu/~lori/classes/GUI/sound.htm

Solanki, A., & Pandey, S. (2019, January 30). Music instrument recognition using deep convolutional neural networks - International Journal of Information Technology. SpringerLink. Retrieved February 8, 2023, from https://link.springer.com/article/10.1007/s41870-019-00285-y

Team Crystal. (2023, November 13). Lab 1 - Crossfade Product Description. Retrieved November 13, 2023 from https://www.cs.odu.edu/~411crystal

Zhang, X. (2022, March 11). Aided recognition and training of music features based on the internet of things and Artificial Intelligence. Computational intelligence and neuroscience. Retrieved February 8, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8933112/

1.5 Overview

In Section Two, there will be an overview of the system that contains a Product Perspective. It will feature an in-depth look at the functions of CrossFade, an overview of its characteristics, and some constraints that will be contained within the program. Finally, there will be an overview of the assumptions and dependencies of CrossFade. Following those topics is a look at the Specific Requirements of the program. The remainder of this SRS contains two main sections: section two gives an overview of the system while in section three, the following content is organized by features that will describe the specific requirements of the program.

2 Overall Description

CrossFade is a user-friendly application revolutionizing music creation and sharing by eliminating the reliance on professional producers and complex equipment. Tailored for entry-level musicians, music students, and artists without access to traditional studios, CrossFade enables users to focus on quality work with minimal equipment. Its unique features include processing live audio, audio files, or MIDI files, transposing them into sheet music or MIDI format, recognizing solo or ensemble performances, and offering transcription correction capabilities. The application's simple hardware requirements—just a microphone and a computer—make it accessible to a wide user base. Powered by cutting-edge technologies like TensorFlow and mlpack, CrossFade is positioned as an innovative and user-friendly solution for collaborative music creation.

2.1 Product Perspective

CrossFade is an application that is accessible via personal computer or laptop and is designed to eliminate the need for professional music producers and complex technological equipment, making music creation and sharing more accessible for aspiring musicians, music students, and artists without access to traditional recording studios. The primary goal of CrossFade is to enable musicians to focus on creating quality work with minimal equipment, alleviating the technical concerns that often burden entry-level musicians. It features the ability to listen to various inputs, including live audio, audio files, or MIDI files, and uniquely transposes the input into sheet music format or a MIDI file. The application can recognize solo or ensemble performances, offers transcription correction capabilities, and outputs MIDI files and corresponding sheet music for each musical composition. CrossFade's hardware requirements are simple, requiring only a microphone and a computer, while its software utilizes

audio parsing and interpreting libraries such as Yaafe and SuperPowered, as well as MusicXML support like libmusicxml and pymusicxml. The application is coded in Python, incorporating state-of-the-art artificial intelligence tools such as TensorFlow and mlpack to enhance its capabilities. This combination of features and technology positions CrossFade as an innovative and user-friendly solution for music creation and sharing.

2.2 Product Functions

CrossFade distinguishes itself with unique features, including the ability to process live audio, audio files, or MIDI files and transpose them into sheet music or MIDI format. The software recognizes solo and ensemble instruments, offering transcription correction features where users can choose AI-suggested edits or perform manual corrections. CrossFade can output both MIDI files and corresponding sheet music for each musical composition, making it an innovative application for music creation and sharing. The product includes a user interface with a feature select menu and a music editor, allowing users to input live audio or MIDI files with options for automatic or manual transcription. Key algorithms cover audio and live audio transcription, MIDI file transcription, automatic and manual transcription correction, and musical instrument recognition. The program emphasizes transcription error correction, distinguishing itself with features like error highlighting, user feedback, and AI-suggested edits, while reducing some capabilities like multitrack transcription and instrument recognition to focus on enhancing transcription correction functionalities.

2.3 User Characteristics

The primary user role for CrossFade is the entry-level musician who faces barriers to accessing the technology prevalent in recording studios, hindering their entry into the music industry. This target user group typically lacks professional equipment for music collaboration.

Additionally, music students, lacking expertise in transcribing their compositions, form another user role. CrossFade's key transcription feature is beneficial for students, enabling them to easily transcribe their music into sheet format for sharing and collaboration. Practicing musicians, who may encounter challenges in sharing their work due to technical barriers, also fall within the user base. Overall, CrossFade caters to individuals at various stages of musical development, providing a user-friendly platform for collaborative music creation and sharing, with a specific emphasis on breaking down barriers for entry-level musicians and aiding music students in their learning journey.

2.4 Constraints

CrossFade faces three main types of constraints: customer risks, technical risks, and legal risks. Customer risks include potential difficulties in using the application and unfamiliarity with program terminology. Technical risks involve potential limitations in system performance due to musical diversity, which could impact the transcription of certain music styles. Additionally, limited library availability may hinder the recognition of certain sounds, and background noises from live audio input might affect performance. Legal risks encompass concerns about unlawful product use and copyright violations, with potential issues related to music or sheet music theft.

2.5 Assumptions and Dependencies

The system will leverage the music21 library to efficiently parse MusicXML files which will streamline the process of extracting information from these files without the need for manual intervention. This is a key dependency that will eliminate the manual effort associated with extracting MusicXML file details. An assumption underlying the system's functionality is that only MusicXML files featuring a single instrument will be accepted as input. This assumption

simplifies the task by limiting the complexity associated with handling different frequencies from multiple instruments, ensuring a more focused and streamlined parsing process.