

Lab 2 - CrossFade Product Specification

Lab 2 - CrossFade Product Specification

Arnaud Blay

CS411W

Thomas J. Kennedy

CrossFade

12/1/2023

Version 2

Table of Contents

1 Introduction.....	2
1.1 Purpose.....	2
1.2 Scope.....	3
1.3 Definitions, Acronyms, and Abbreviations.....	4
1.4 References.....	5
1.5 Overview.....	7
2 Overall Description.....	7
2.1 Product Perspective.....	7
2.2 Product Functions	9
2.3 User Characteristics	10
2.4 Constraints	11
2.5 Assumptions and Dependencies.....	11

List of Figures

Figure 1: Prototype Major Functional Components Diagram	9
---	---

1 Introduction

CrossFade democratizes the process of music creation as it transforms audio recordings into MIDI and Sheet Music with remarkable ease. This innovative platform not only simplifies music transcription and error correction but also enhances the collaborative and creative potential of its users.

1.1 Purpose

CrossFade is a music production software solution that facilitates the creation and distribution of music by transforming live audio or pre-recorded audio files into MIDI and Sheet Music. Aimed at novice musicians, it leverages advanced audio parsing libraries and AI technology to detect notes and instruments, simplifying the music production process by negating the need for extensive technical expertise or high-end equipment. CrossFade's user-friendly platform not only encourages musical creativity and expression but also fosters inclusivity within the music industry by making music production more accessible.

At a very high level, CrossFade will provide musicians the ability to convert audio recordings directly into MIDI files and sheet music, facilitating the sharing and editing of their compositions. It will offer error correction for transcriptions, using AI to enhance the accuracy of the musical notation produced. Additionally, the product will support user interaction, allowing for manual corrections that contribute to the system's learning and improvement over time. However, CrossFade will not replace the comprehensive functionalities of advanced music production software that requires deep technical knowledge. It will not focus on teaching music theory or providing in-depth music education. Instead, it focuses on transcription and error correction to streamline the music creation process for its users.

1.2 Scope

The application of CrossFade centers on providing a seamless transition for novice musicians into music production. Its objective is to lower the barriers to music creation by converting audio to MIDI and sheet music, bypassing the need for extensive technical knowledge or equipment.

The relative benefits include the democratization of music production, enabling a broader range of artists to share and refine their music. The goals are to enhance accessibility, foster creativity, and streamline the music-sharing process.

At a high level, the prototype of CrossFade will serve as a proof of concept to demonstrate the product's capabilities. It will utilize advanced machine learning algorithms to analyze audio inputs and transcribe them into MIDI and MusicXML formats. The prototype will highlight features such as real-time error detection and correction, user-assisted transcription, and automatic music notation improvement based on user feedback. It will also showcase the application's ability to learn and evolve from manual corrections, ultimately aiming to display how technology can solve the challenges of accurate music transcription.

[This space intentionally left blank]

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.

[This space intentionally left blank]

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>

Team Crystal (2023, April 24). Lab 1 - CrossFade Product Description. Retrieved November 11,

2023 from <https://arnaud-odu.github.io/CS411-Crystal-CrossFade/lab1.html>

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-](https://www.toptal.com/algorithms/shazam-it-music-processing-fingerprinting-and-recognition)

[fingerprinting-and-recognition](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>

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/>

[This space intentionally left blank]

1.5 Overview

This product specification provides the hardware and software configuration, external interfaces, capabilities and features of the Product the CrossFade prototype. The information provided in the remaining sections of this document includes a detailed description of the hardware, software, and external interface architecture of the CrossFade prototype; the key features of the prototype; the parameters that will be used to control, manage, or establish that feature; and the performance characteristics of that feature in terms of outputs, displays, and user interaction.

2 Overall Description

With its intricate balance of hardware and software components, and a robust architecture using Python and C++, CrossFade is at the forefront of music technology, simplifying transcription processes and encouraging innovation and exploration in music creation.

2.1 Product Perspective

CrossFade is an innovative music production solution tailored to address the technical challenges faced by novice musicians in the creation and distribution of their music. The application stands out by enabling users to generate MIDI and Sheet Music through the analysis of live audio or audio files, discerning the notes played with precision. This functionality is particularly beneficial for aspiring musicians who may not possess extensive technical knowledge or access to high-end equipment, thereby democratizing music production.

Key Product Features and Capabilities include the ability to input audio files, which CrossFade processes to identify instruments and convert the music into MIDI format. This feature is important for musicians without access to soundboards or keyboards capable of MIDI output, simplifying music production and sharing processes. Moreover, CrossFade integrates AI-assisted editing and correction tools. These tools not only automate but also learn from manual

corrections, enhancing the software's capability over time. Such features are instrumental in aiding musicians to navigate the learning curve and challenges of the industry.

CrossFade is comprised of hardware and software elements. As seen in Figure 1, Hardware components include audio MP3 and MusicXML or MIDI files as inputs, and a computer to run the application. On the software front, CrossFade leverages audio parsing and interpreting libraries such as Yaafe and SuperPowered, implemented in C++ and Python. The software accesses a database of instrument samples from remote servers, utilizing resources like IRMAS and Philharmonia samples. Furthermore, it supports MusicXML through libmusicxml and pymusicxml. Advanced AI techniques, powered by TensorFlow and mlpack, are integral to enhancing CrossFade's functionality, making it a cutting-edge tool in music technology.

Written predominantly in Python with C++ for complex music algorithms, CrossFade encompasses a range of algorithms for music transcription. These include live audio and audio file transcription, MIDI file transcription, and both automatic and manual transcription correction. This comprehensive approach in its architecture makes CrossFade an invaluable platform for emerging musicians, simplifying complex processes and fostering musical innovation and exploration.

[This space intentionally left blank]

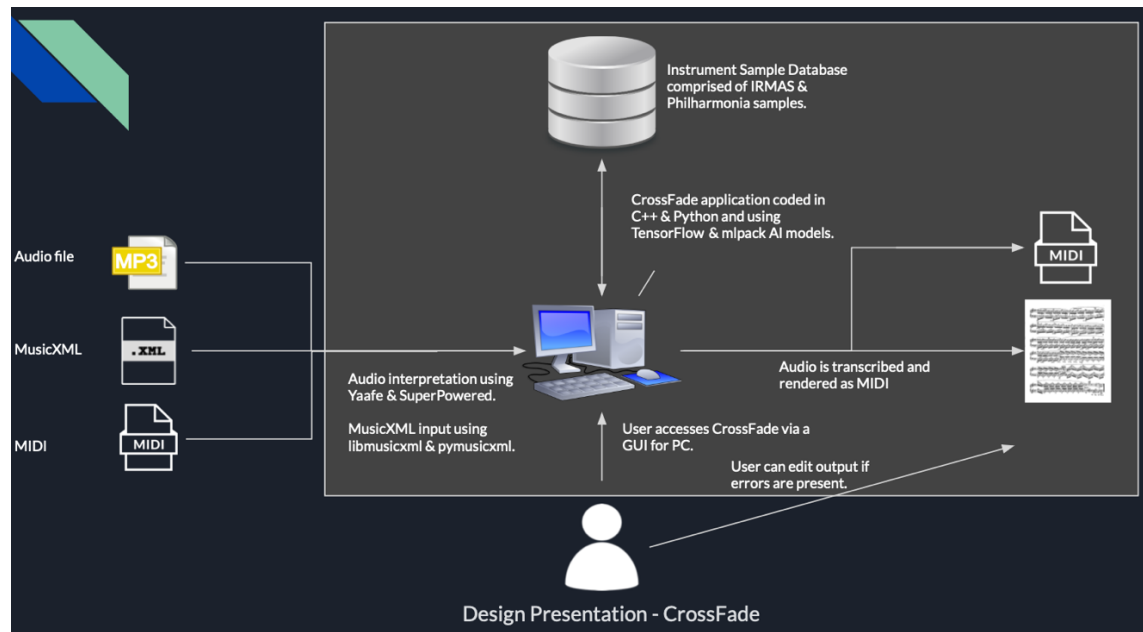


Figure 1: Prototype Major Functional Components Diagram

2.2 Product Functions

The major functional components of the CrossFade prototype include the following:

- **Song and Metadata Editing Interface:** This function allows the user to edit song information (time signature, tempo, key) and metadata (song title, artist(s) name, album) through an intuitive interface.
- **Track and Note Editing Interface:** Facilitates the editing of individual track information (track name, instrument) and note details (pitch, duration, articulation), enhancing the depth and detail of music editing.
- **Edited Music Playback System:** This module enables users to listen to either the complete musical piece or a selected section of the song being edited. It provides real-time auditory feedback of the edits made, enhancing the user's editing experience.

- **Music File Import and Export System:** Supports importing music files in MusicXML and MIDI formats and saving/exporting modified music files. This ensures compatibility with common music file formats and facilitates the sharing and distribution of edited work.
- **MusicXML File Parsing and Dataset Generation:** Parses MusicXML files for trainer datasets, essential for the application's machine learning and algorithmic functions.
- **Transcription Error Detection and Correction:** Identifies and corrects transcription errors, including rhythm, articulation, and dynamics discrepancies. This function is pivotal for maintaining the accuracy and quality of the musical content.
- **System Performance Optimization:** Guarantees efficient file upload times and swift updates of the sheet music page, optimizing the overall user experience and application performance.
- **Assumptions and Constraints Handler:** Operates under specific assumptions (e.g., only single-instrument MusicXML files input) and utilizes specialized libraries (such as music21) to manage these constraints effectively, simplifying the music processing tasks.

2.3 User Characteristics

The user demographic for CrossFade primarily consists of individuals venturing into the music industry, including a range of users from novices to experienced professionals. The prototype caters to practicing musicians who require a tool for transcribing their compositions into sharable sheet music or MIDI files. Music teachers represent facilitators who would employ CrossFade as a teaching aid for their students. Composers form another critical user segment, utilizing the prototype's advanced features for creating and modifying musical pieces. Artists can leverage CrossFade's capabilities for personal projects and professional endeavors. Collectively, these user

groups define the application's target audience, each with unique requirements and expectations from the software, shaping its development and feature set.

2.4 Constraints

N/A

2.5 Assumptions and Dependencies

N/A