



CROSSFADE



Dylan Hopper

CS 411W Group Crystal

Professor Thomas J. Kennedy

11/17/2023

Version 2

Table of Contents.

| | |
|--|---|
| 1. Introduction..... | 2 |
| 2. Product Description..... | 2 |
| 2.2 Major Functional Components | 2 |
| Hardware..... | 3 |
| Software..... | 3 |
| 3. Usage Case..... | 4 |
| 4.1 Prototype Architecture..... | 5 |
| 4.2 Prototype Features and Capabilities..... | 5 |
| 4.3 Prototype Development Challenges..... | 6 |
| 5. Glossary | 6 |
| 6. References..... | 7 |

1. Introduction

The barrier to entry for the music industry can be too high for newcomers. Both musical equipment and software can be expensive, and the time it can take to learn to read sheet music can take too long. A study by Ruth et al. showed that 50% of music students quit their musical lessons and other musical activities by the time they have reached age 17, some at an earlier age. The study notes that cost, loss of support, and the quality of the musical program are among some of the reasons for students losing interest. CrossFade software makes it easier for the novice musician to learn music and to share their ideas; helping to combat some of the previously mentioned issues facing students.

2. Product Description

CrossFade is a music transcription software that allows musicians to transcribe their music into sheet music and to save it as a MIDI file. Unlike other transcription software, CrossFade does not require a USB or MIDI device to interface with the software. These traditional methods can still be used, but CrossFade also supports input via a microphone which allows for any analog instrument to be used as well. This reduces the cost and need to learn the piano for musicians that desire to have their music automatically transcribed. This also allows musicians who are not familiar with transcription and sheet music to share their music with other people. CrossFade supports sheet music and MIDI output.

2.2 Major Functional Components

CrossFade supports input from any sound producing instrument (including the voice) via a microphone, MIDI keyboard input, or audio file input. The audio is then transcribed to both a MIDI file and to MusicXML where it is rendered as a visible sheet. Audio input via a microphone uses state-of-the-art artificial intelligence trained on a massive dataset of IRMAS

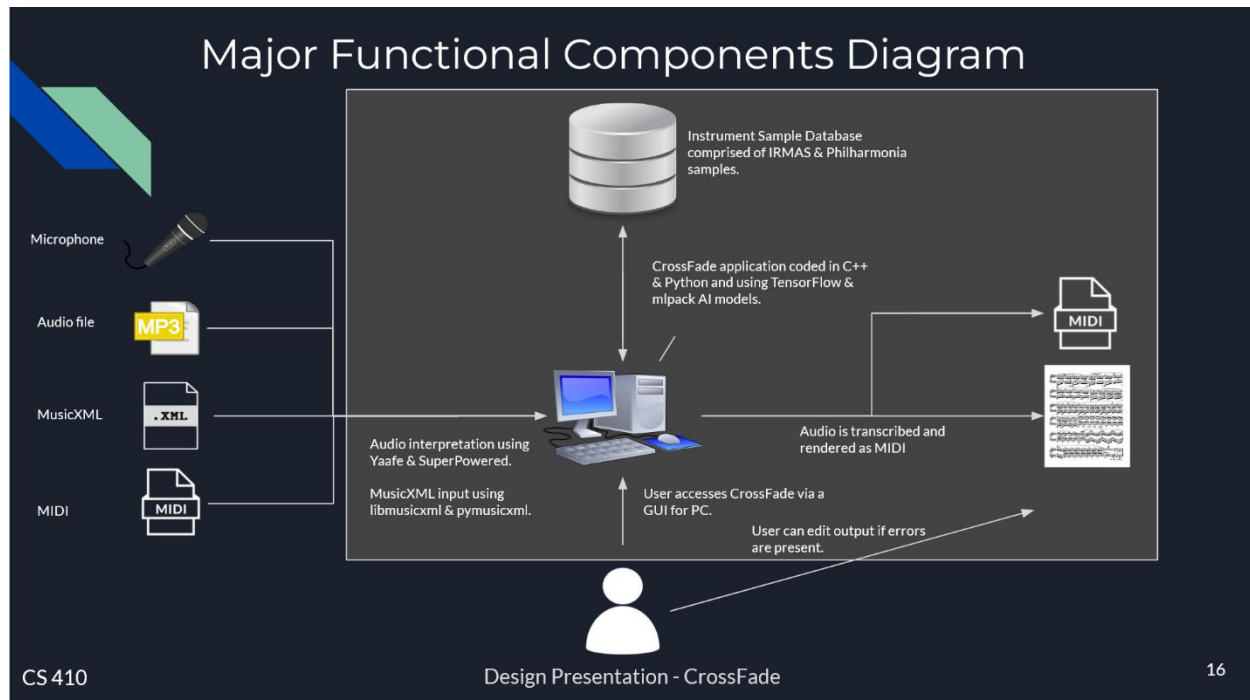
and Philharmonia instrument samples. The recognized audio is then interpreted using the Yaafe and SuperPowered libraries. This interpreted audio is then converted to MIDI, and to MusicXML using the libmusicxml and pymusicxml libraries, finally it is rendered as a visible sheet using the music21 toolkit for Python.

Hardware.

- Instrument(s)
- Microphone
- Computer

Software.

- Audio Parsing and Interpreting Libraries
 - Yaafe
 - SuperPowered
- C++ and Python
- Database of Instrument Samples
 - IRMAS
 - Philharmonia samples
- MusicXML Support
 - libmusicxml
 - pymusicxml
- State-of-the-Art Artificial Intelligence
 - TensorFlow
 - mlpack



3. Usage Case

CrossFade is designed to be used by learning musicians to reduce the usual investment cost associated with entering the field. The software aids the beginner musician in learning sheet music, transcribing audio, and songwriting. These are crucial in learning music as it helps the beginner digest the music in a more understandable form. Additionally, institutional music instructors will also find CrossFade useful when teaching their students due to its ability to transmute audio into a readable form more easily, taking the normal time and effort of having to transcribe audio by hand out of the equation.

4. CrossFade Prototype Description

The CrossFade prototype will focus more on transcription correction using AI as the driving mechanism behind error detection. A GUI will be designed for a desktop application. Audio, MIDI, and MusicXML will be parsed and interpreted using external libraries. On the

backend, state-of-the-art artificial intelligence will be used to detect instrumentation by comparing to samples, as well as detecting any errors present in current MusicXML transcription. Live audio capture will not be supported in the prototype model, however prerecorded audio files can still be used for input. The final product will be output to a MusicXML file.

4.1 Prototype Architecture

CrossFade will be designed as a desktop application using Python for the GUI with embedded C++ modules for library interfacing. Audio will be interpreted and parsed using the Yaafe and SuperPowered libraries. MusicXML is implemented using the pymusicxml and libmusicxml libraries, these will be used for both input and output. MIDI libraries are to be determined. A database containing IRMAS and Philharmonia instrument samples will be used for training and reference purposes for our AI models. As for the AI models themselves we will be using TensorFlow and mlpack.

4.2 Prototype Features and Capabilities

The prototype demonstrates the main goal of this software which is providing both beginner musicians, and songwriters & educators an easy ability to transcribe music and/or correct previously transcribed music containing errors. The omitted live audio capture is unnecessary in showcasing the capabilities of CrossFade's transcription-focused design, and because the live audio capture is a large undertaking, building the prototype without it can help provide investors and users with a more successful deliverable.

4.3 Prototype Development Challenges

The biggest hurdle when developing the prototype will be training our machine learning and artificial intelligence models to recognize audio, midi, and MusicXML sheet music, and then correct them if errors are present. This will most likely take a large dataset with many different parameters.

Another issue will be learning and implementing all the third-party libraries that are present. Many of them have little documentation & support, and it will take time to fully understand how to use them, and make them play nicely together.

5. Glossary

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.

6. References

Ruth, N., & Müllensiefen, D. (2021). Survival of musical activities. When do young people stop making music?. *PloS one*, 16(11), e0259105. <https://doi.org/10.1371/journal.pone.0259105>