Lab 1: CrossFade Product Description

Hyacinth S. Abad

CS 411W

Professor Thomas Kennedy

17 November 2023

Version 2

Table of Contents

Listing of Figures

Listing of Figures

1 Introduction

The process of music production tends to be overly complex, usually requiring a professional music producer and high-level equipment to produce music. Tools such as the mixing board, synthesizer, and the digital audio workstation are only some of the many equipment that can be found in a typical recording studio room as shown in Figure 1 (Scarlatos).

Figure 1: A typical recording studio

For the most part, entry-level musicians do not have access to this type of technology and are not knowledgeable of its functionalities. Some of the equipment is highly valuable with costs up to $10, 000 and are not beginner-friendly to use. Those who are starting out usually do not have the money to afford the costly music production tools (Zhang, 2022). Without this type of technology, it would be difficult for an aspiring musician to produce and share his body of work. In addition to entry-level musicians, practicing musicians, composers, artists, and music students would also struggle with making or sharing music without the necessary equipment. As a result, there is a need for a type of software that alleviates the technical requirements when producing and sharing music. CrossFade is the perfect product for those who lack the equipment but yet have the desire to create and share their music. Unlike the complex technology found in

recording studios, CrossFade will be readily available to the general public and will be easy to use for all aspiring musicians in all levels. Gone will be the days where there is a need for a professional music producer or an actual recording studio room in order to produce and share an individual's music.

2 CrossFade Product Description

CrossFade is a graphical user interface (GUI) application that is accessible via personal computer or laptop. The main purpose of this program is to alleviate the need for a professional music producer and the technological equipment utilized in order to create and share music. With this product, the barriers to making music are greatly reduced for aspiring musicians, music students, and artists who typically do not have access to a professional recording studio and all its equipment. The main objective of CrossFade is to allow musicians to focus on creating quality work with just a few pieces of equipment. This product takes away the technical worries of creating and sharing music that many entry-level musicians have. The musicians provide the creativity and CrossFade will do the rest.

2.1   Key Product Features and Capabilities

CrossFade boasts of multiple features that set it apart from other similar music applications. One of its features is the ability to listen to input. The type of input can be in the form of live audio, an audio file, or a Musical Instrument Digital Interface (MIDI) file. Another unique feature is its effective ability to transpose the inputted audio into sheet music format (MusicXML) or a MIDI file format. CrossFade can also recognize various instruments played solo or played in together with other instruments. Transcription correction is another CrossFade feature. With this feature, users can either choose A.I.-suggested edits to the sheet music or they can perform their own edits. Finally, CrossFade can output MIDI file and a corresponding sheet music for each body of

work produced by the musician. All of these features combined make CrossFade an innovative

application, allowing music creation and sharing with minimum and easy to use technology.

2.2   Major Components (Hardware/Software)

Unlike the complexity of professional music production technology, CrossFade is known for its

minimal hardware requirements. Besides the instruments that the user plays, only a microphone

and a computer are needed to utilize the application. The software technology included in

CrossFade include audio parsing and interpreting libraries such as Yaafe and SuperPowered.

These libraries are necessary for the application to extract and recognize the different sounds that

are inputted into the application. Likewise, MIDI input will be recognized by MusicXML support

such as libmusicxml and pymusicxml. Other libraries that CrossFade utilizes include IRMAS and

Philharmonia samples. These libraries are databases for instrument samples. With these libraries,

CrossFade has the ability to identify and differentiate various instruments played alone or

together. The basis of the application will be coded in C++ and Python and will include state-

ofthe-art artificial intelligence such as TensorFlow and mlpack. All of this hardware and software

architecture is utilized throughout the solution process as shown in Figure 2.
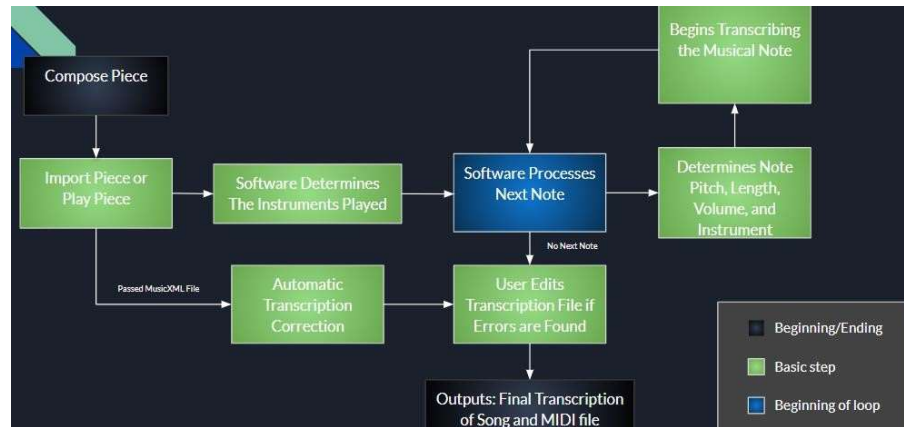
Figure 2: Solution process flow of CrossFade

3 Identification of Case Study

This product is mainly developed for entry-level musicians who usually lack the type of

technology found in most recording studios that allow for music collaboration. The lack of access

to these types of equipment can serve as a barrier for entry into the music industry. Other types of

users such as music students and practicing musicians may also benefit from this application.

Because of its key transcription feature, CrossFade can help those who lack the knowledge of

transcribing their own music such as music students. Having this feature allows students to share

their music in the form of a sheet music which can be shared with other students who might want

to collaborate together.

4 CrossFade Product Prototype Description

The CrossFade prototype will include a user interface in the form of a feature select menu and a

music editor. The feature select menu allows the user to choose an option of inputting a live

audio or inputting a MIDI file. With these inputs the user can choose an automatic or a manual

transcription. The prototype has transcription correction features found in the music editor tab

such as highlighting possible errors on the sheet music, requesting user feedback, offering

possible solutions to transcription errors, comparing single note from original to written, and

comparing a segment of music from original to written. The algorithms utilized in this

application will include an audio file and live audio transcription algorithm, MIDI file

transcription algorithm, an automatic transcription correction algorithm, a manual transcription

correction algorithm, and a musical instrument recognition algorithm. The combination of these

key features make CrossFade an innovative and useful program.

Some capabilities that were initially brought up during the brainstorming phase were either

reduced or eliminated. These include multitrack transcription, monophonic and polyphonic note

recognition, instrument recognition, and instrument distinction as shown in Figure 3. The focus

of the program's prototype is its transcription error correction as shown in Figure 5. CrossFade's

error correction features allow for AI-suggested edits and manual transcription edits which the

application's competitors seem to either lack or have poor functionalities. Certain capabilities,

such as the instrument recognition, will still be included but are reduced in order to focus more

on its transcription correction feature.

Transcription		
Feature	Real World Product	Prototype
Live Audio Transcription	✓	
MIDI Input/File Transcription	✓	
MIDI Transcription Correction	✓	✓
Multitrack Transcription	✓	
MusicXML Compatibility	✓	✓
MIDI File Export	✓	✓

Note Recognition		
Feature	Real World Product	Prototype
Monophonic Note Recognition	✓	
Polyphonic Note Recognition	✓	
Instrument Recognition	✓	
Instrument Distinction	✓	

Table 1: Transcription feature RWP vs Prototype   *Table 2: Note recognition feature RWP vs Prototype*

Transcription Error Correction		
Feature	**Real World Product**	**Prototype**
Transcription Error Correction	✓	✓
Highlight Possible Errors	✓	✓
Take User Feedback	✓	✓
Offer Possible Solutions	✓	✓
Compare Single Note from Original to Written	✓	✓
Compare a Segment from Original to Written	✓	✓

Table 3: Transcription error correction feature RWP vs Prototype

4.1   Prototype Architecture (Hardware/Software)

The architecture of the CrossFade prototype begins with live audio inputs from a microphone or music files such as an audio file, MusicXML, or a MIDI file. The libmusicxml and pymusicxml are different types of softwares that allow MusicXML files to be inputted. These inputs are directly inputted into a personal computer or a laptop that contains the CrossFade application. Yaafe and SuperPowered are programs that will be used to interpret the audio and utilize its transcription error correction features. Artificial intelligence models such as TensorFlow and mlpack are then used to identify possible transcription errors and suggest transcription edits. C++ and Python are the main languages that will display the GUI on the PC, including a main menu options of inputting live audio/audio files or transcription correction. After the audio is transcribed, the program outputs a MIDI file of the audio and a transcription sheet music. With the sheet music, the A.I. models mentioned before can identify errors in transcription and suggest edits. The users also have an option to correct the transcription themselves instead of accepting

the A.I.-suggested edits. The diagram shown in Figure 3 shows the major functional components

of the CrossFade prototype including the hardware and software architectures.
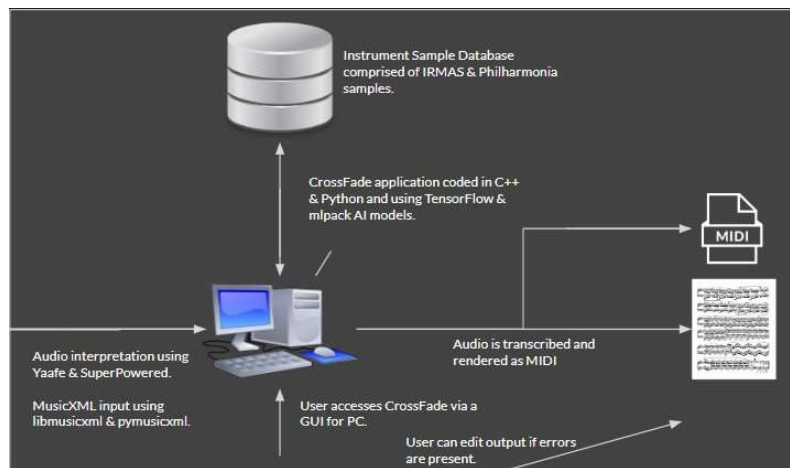

Figure 3: Major Functional Components Diagram

4.2   Prototype Features and Capabilities

Unlike the real-world product which has all the capabilities of transcription, note recognition,

and transcription error correction, the prototype will focus more on the transcription error

correction feature. This is highlighted by the multiple features that allow the transcribed sheet

music to be either automatically corrected by the A.I. models or manually corrected by the user.

Despite the features that have been reduced or minimized in the prototype, the CrossFade

application continues to provide a solution for budding musicians who are in need of technology

that allows for music sharing via MIDI files and audio transcription. CrossFade allows for music

editing and collaboration between musicians.

4.3   Prototype Development Challenges

There are three different types of challenges that are encountered when completing the prototype.

The customer risks shown in Figure 4 includes difficulties using the application by the user and

unfamiliarity with the terminology in the program. The first risk can be mitigated by including a

tutorial on how to use CrossFade which can be shown in the opening of the program. The second

risk can be mitigated by adding definitions of terms or an entire glossary in the program.
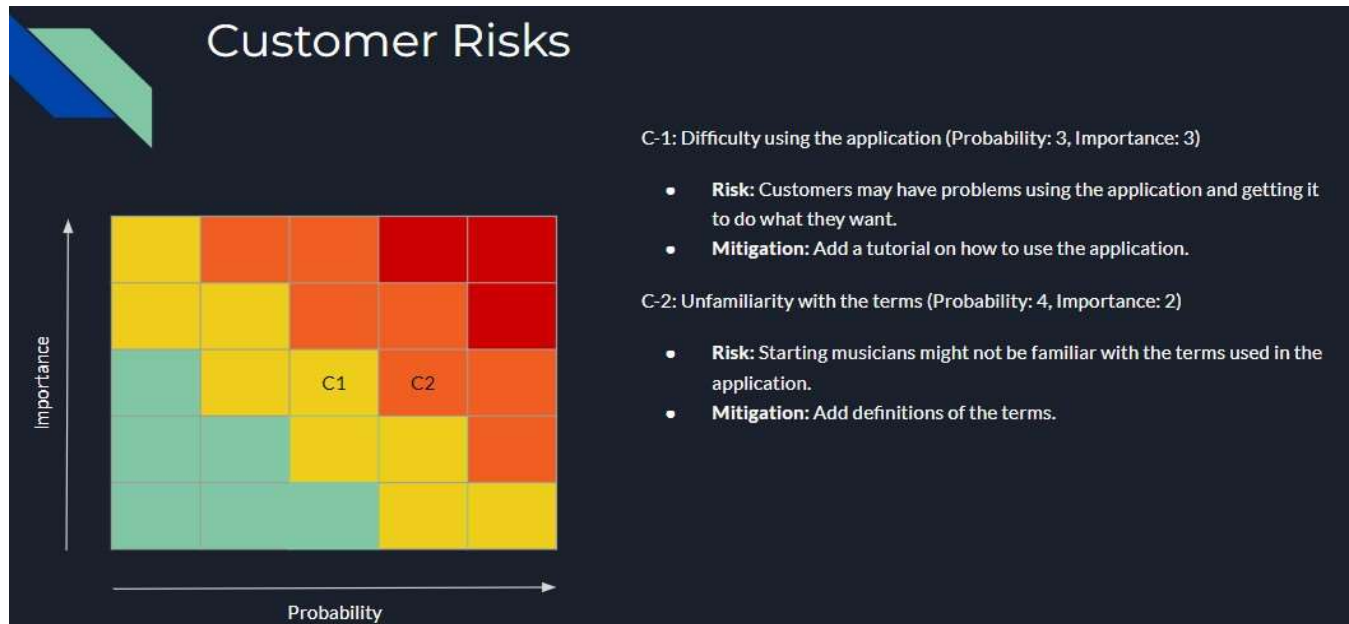

Figure 4: Customer Risks and Mitigation

The second type of challenge the prototype might encounter is its technical risk shown in Figure

5. The prototype might have a limited system performance due to musical diversity. Some music

styles might be more difficult to transcribe than others. This risk can be reduced through

multidisciplinary collaboration and algorithms tailored to specific use cases. CrossFade might

also have limited library availability meaning various databases of sounds and instruments might

be lacking resulting in the inability to recognize certain sounds and outputs. There is also a

possibility that the program will recognize background noises from the live audio input and

affect its performance to transcribe the audio. A possible solution for this would be to isolate and

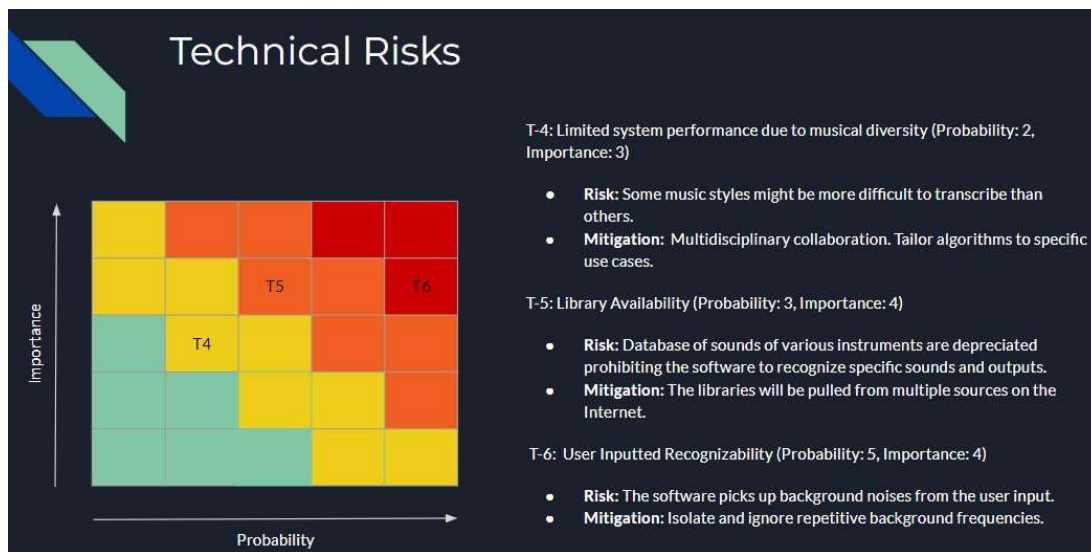ignore repetitive background frequencies.

Figure 5: Technical Risks and Mitigation

Legal risks are also some challenges that might be encountered when completing the prototype as shown in Figure 6. Unlawful product use and copyright law are some of the legal risks. Customers might attempt to steal or distribute music or sheet music using the application. This risk can be mitigated by not prohibiting the creation of an audio file of the music and adding a watermark to discourage illegal distribution. The use of copyright works in datasets used for training AI systems may have legal implications. Using non-copyrighter works on AI training models or using copyrighted material on the grounds of a license agreement can help mitigate this legal risk.
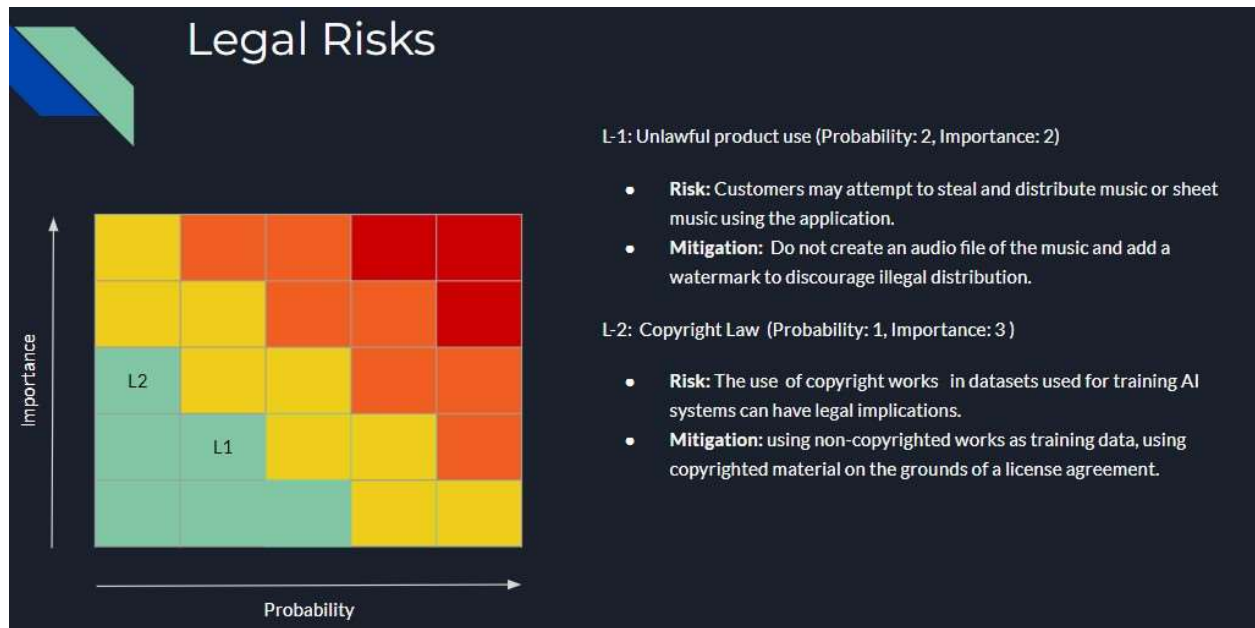
Figure 6: Legal Risk and Mitigation

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

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

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/