Lab 1 - CrossFade Product Description

Arnaud Blay

CS411W

Thomas J. Kennedy

CrossFade

11/17/2023

Version 6

Table of Contents

List of Figures

List of Tables

1 Introduction

The music industry presents a significant challenge for aspiring musicians seeking to showcase their talent and share their music with the world. Technical equipment and knowledge of music production are essential but can be expensive, making it difficult for novice musicians to create and distribute their music. While MIDI files are the standard for music production, the necessary equipment can be inaccessible to many. Moreover, starting musicians may not be familiar with sheet music, which is vital for sharing music in written form. This problem has persisted for years, hindering the growth of budding musicians. A solution to this problem must be affordable, user-friendly, and accessible. CrossFade is a new product that aims to address this problem by simplifying the technical requirements for starting musicians, allowing them to produce and share music with ease. Leveraging AI technology, CrossFade converts audio recordings into MIDI files, enabling novice musicians to transcribe their music into a written format even without knowledge of sheet music. With CrossFade, musicians can easily share their compositions in a readable and editable format, freeing them to focus on their creativity and produce more music without feeling overwhelmed. CrossFade makes the music industry more inclusive and accessible to all musicians.

2 Product Description

CrossFade is a music production solution designed to alleviate the technical obstacles faced by novice musicians when creating and distributing their music. The application allows users to generate MIDI and Sheet Music by analyzing live audio or audio files and determining the notes being played. CrossFade's advanced capabilities empower aspiring musicians to create music without requiring extensive technical expertise or costly equipment. As a result, CrossFade

offers an accessible platform that promotes creativity, artistic expression, and inclusivity within the music industry.

2.1 Key Product Features and Capabilities

CrossFade presents a comprehensive solution to the technical challenges faced by novice musicians who lack soundboards or keyboards capable of generating MIDI output. By enabling users to input live audio or audio files, CrossFade can detect the instruments being played and convert the music into MIDI format. This feature simplifies the music production process for musicians and enables them to share their creations without the need for costly equipment. Furthermore, CrossFade incorporates AI-assisted editing and correction capabilities that can help aspiring musicians overcome the hurdles of learning and achieving success in the music industry. These corrections can be done automatically and manually, allowing it to learn and improve overtime with each manual corrective action.

2.2 Major Components (Hardware/Software)

The CrossFade application is reliant on both hardware and software components to operate effectively. As seen in Figure 1, the hardware components comprise instruments, a microphone, audio MP3 files, music XML or MIDI files (the inputs), and a computer capable of running the CrossFade application.

[This space intentionally left blank]

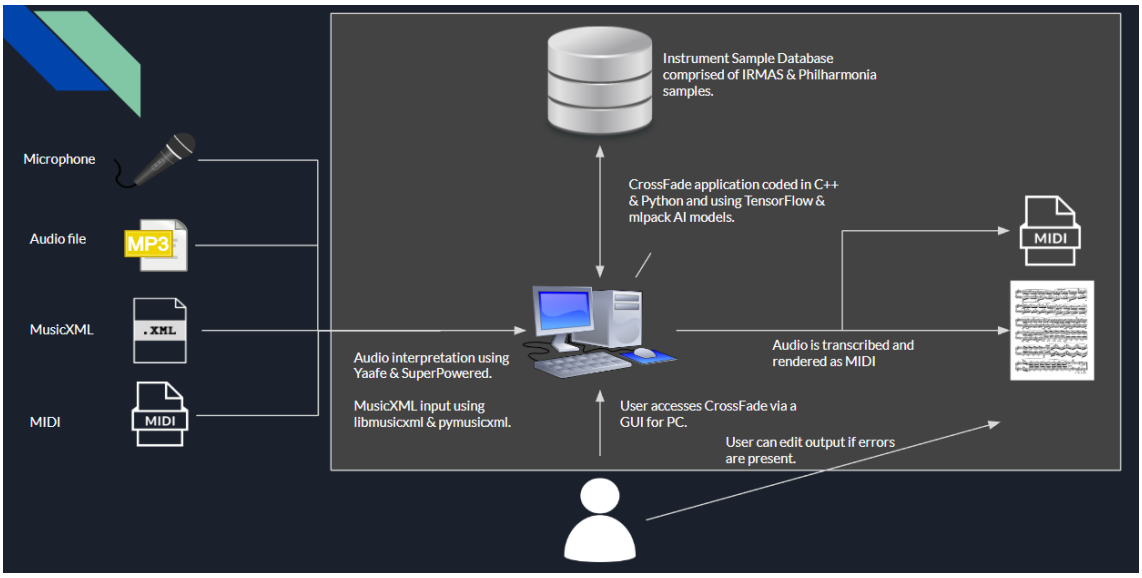

Figure 1: Major Functional Components Diagram

The software components of the application include audio parsing and interpreting libraries, such as Yaafe and SuperPowered, which are programmed in C++ and Python. The program interacts with a database of instrument samples, such as IRMAS and Philharmonia samples, located on a remote server, and also includes MusicXML support through libmusicxml and pymusicxml. CrossFade utilizes cutting-edge artificial intelligence techniques through TensorFlow and mlpack to enhance its functionality.

CrossFade is written using the Python programming language and uses C++ for more complex music algorithms. The program incorporates a range of algorithms to enable musicians to transcribe their music, including live audio and audio file transcription, MIDI file transcription, and automatic and manual transcription correction.

3 Identification of Case Study

CrossFade is a software solution that is targeted towards individuals who are entering the music industry, including aspiring musicians, composers, and artists. The program is specifically designed to provide a user-friendly platform that alleviates the technical challenges that beginner

musicians often face when trying to create and share their music. In addition, practicing musicians may find CrossFade's AI-assisted editing and correction features useful for improving their existing music files. Music teachers may also find CrossFade beneficial for their students, as it enables them to produce and share music without needing extensive technical knowledge or expensive equipment. Overall, CrossFade is an inclusive and accessible platform that encourages creativity, expression, and learning for anyone interested in music.

4 Product Prototype Description

The prototype is a software application that utilizes advanced machine learning algorithms to analyze and correct music transcription errors. As we can see in the following tables, the prototype will focus on some aspects of transcription (as shown in Table 1), all aspects of transcription error connection (refer to Table 2), and we will not focus on note recognition (as illustrated in Table 3).

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

Table 1: Real World VS Prototype – Transcription

[This space intentionally left blank]

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 2: Real World VS Prototype - Transcription Error Correction

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

Table 3: Real World VS Prototype - Note Recognition

4.1 Prototype Architecture (Hardware/Software)

Crossfade is capable of accepting MIDI and XML files as inputs and automatically identifying and correcting common errors such as incorrect rhythms, articulations, and dynamics. It utilizes a neural network trained on a vast library of musical scores to accurately recognize and correct errors. In addition to highlighting corrected errors in real-time, the software also offers alternative solutions for each error detected. The software also provides the ability to play what is written for the user's convenience. Furthermore, the software has a feedback system that

allows the user to suggest improvements and corrections, which can be used to improve the overall quality of the transcribed music.

4.2 Prototype Features and Capabilities

The prototype demonstrates the effectiveness of machine learning algorithms in analyzing and correcting music transcription errors. By utilizing a neural network trained on a vast library of musical scores, the software accurately recognizes and corrects errors in real-time. This is significant in showing how the problem of inaccurate music transcription can be solved through the use of advanced technology. The software not only reduces the time and effort required for musicians to transcribe music accurately, but it also improves the overall quality of transcribed music.

The transcription features of the software include MIDI transcription correction, MusicXML compatibility, and MIDI file export. The software also offers automatic error detection, with errors found being highlighted, and possible solutions offered for each error. The software provides real-time feedback to the user, allowing for correction and comparison of single notes or segments from the original to the written version. The software also takes user feedback into account, helping to improve the accuracy of the transcription.

4.3 Prototype Development Challenges

Developing a music transcription error correction prototype is a challenging task that requires a comprehensive and diverse music library to train the neural network accurately. The complexity of music notation, unique styles, varying interpretations, and intricate rhythms can pose difficulties for the machine learning algorithm. Technical challenges, such as the requirement for advanced computing power, significant memory, and storage, can also be encountered.

4.3.1 Customer Risks

Customer risks encompass the potential issues encountered by consumers during the utilization of Crossfade. A visualization of those risks can be found in Table 4.

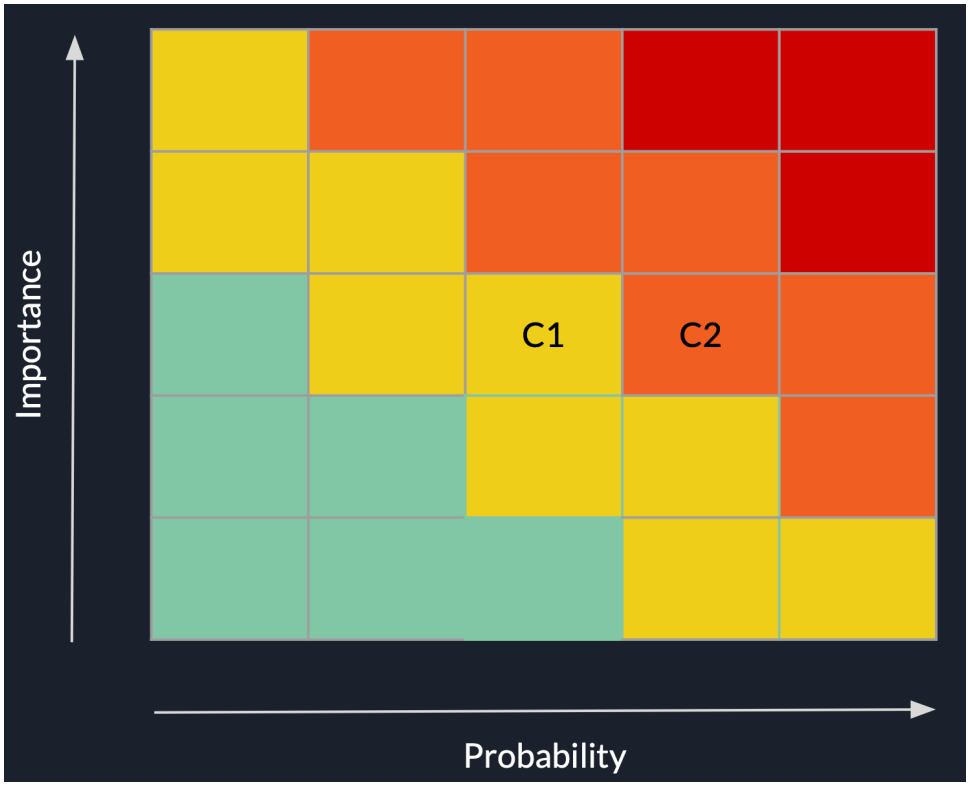

Table 4: Customer Risks

4.3.1.1 C-1: Difficulty Using The Application

One potential difficulty with using the CrossFade application is that some customers may struggle to use it effectively and achieve their desired outcomes. To mitigate this risk, a tutorial will be added to the application that provides step-by-step guidance on how to use the different features and functionalities of the program. This will help users better understand how to use the application and achieve their desired results, ultimately improving the overall user experience.

4.3.1.2 C-2: Unfamiliarity With Some Terms

In the CrossFade application, there is a risk that new musicians may not be familiar with the terms used within the application. To mitigate this risk, the software includes definitions of the terms used in the application. These definitions are readily accessible through the user interface and provide clear and concise explanations of the various musical terms used in the application. This feature helps to ensure that all users, regardless of their level of musical expertise, can understand and use the application effectively.

4.3.2 Technical Risks

Technical risks pertain to the potential challenges and uncertainties arising from the development and deployment of Crossfade. A visualization of those risks can be seen in Table 5.

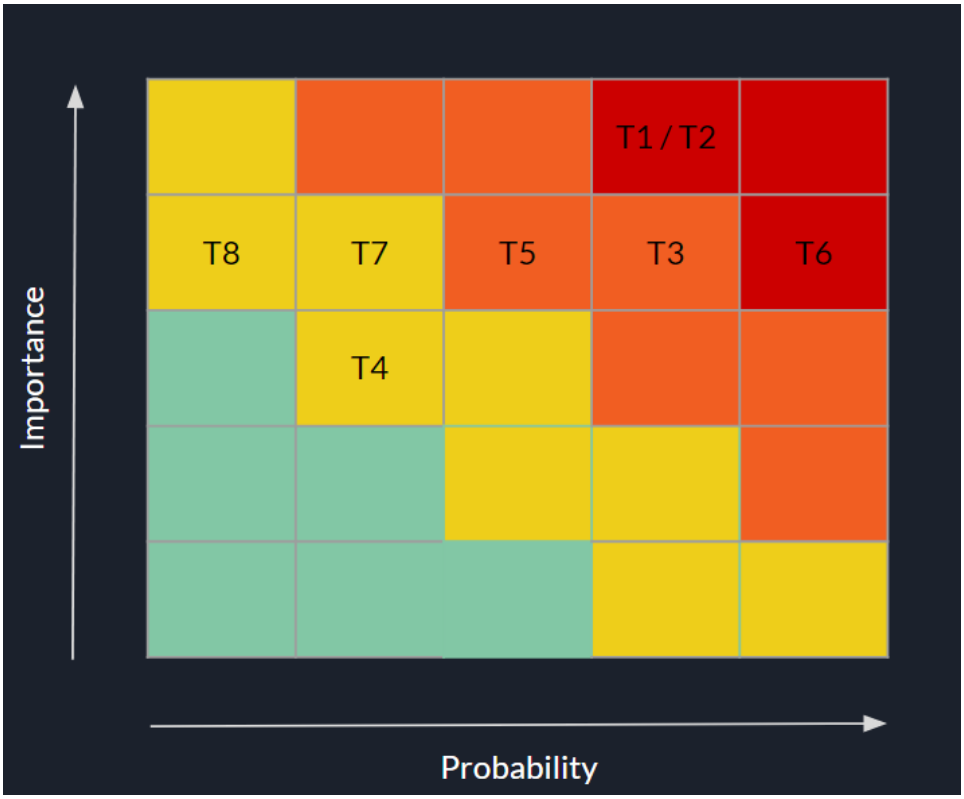

Table 5: Technical Risks

4.3.2.1 T-1: Multi-Pitch Detection Feature Difficulty

Multi-pitch detection is a complex feature in the application that allows the program to recognize and distinguish between multiple instruments playing simultaneously. However, there is a risk that the program may not be able to accurately identify which instruments are playing in certain parts of the music tracks. To mitigate this risk, the program presents the user with the most probable options for the instruments present in a specific part of the input. This allows the user to choose which one is correct and ensure that the correct instrument is being analyzed and corrected.

4.3.2.2 T-2: Note Tracking Feature Difficulty

Mitigating the risk of note tracking difficulty might be more difficult to achieve than we imagine, the application can implement user-assisted transcription to provide input that otherwise could not be inferred easily using algorithms. This will allow the user to manually input notes or musical features that the software might have difficulty detecting. Additionally, integrating independent estimators for various musical aspects such as rhythm and pitch can help increase the accuracy of the Note Tracking feature. These measures will help to ensure that the feature works effectively and provides accurate results.

4.3.2.3 T-3: Overestimating System Performance

To mitigate the risk of overestimating the performance of the system, it is important to create training data that accurately reflects the limitations of the program. One approach is to align existing musical scores with the input audio to create a more comprehensive dataset. This can help to identify areas where the program may struggle and highlight opportunities for improvement. Additionally, it is important to establish clear performance metrics and benchmarks to objectively evaluate the program's capabilities and progress over time.

4.3.2.4 T-4: Limited System Performance Due to Musical Diversity

To address the risk of limited system performance due to musical diversity, a multidisciplinary collaboration is necessary. Collaboration with musicians, musicologists, and experts in various musical styles can help in tailoring algorithms to specific use cases. In addition, the system can be trained on a diverse range of musical styles and genres to improve its accuracy and reduce the risk of errors. This can involve creating a diverse dataset that includes a wide variety of musical styles and subgenres. By training the system on a broad range of musical styles, the system can be better equipped to handle the diversity of music and improve its overall performance.

4.3.2.5 T-5: Library Availability

To mitigate the risk of a database of sounds of various instruments being depreciated, which could prohibit the software from recognizing specific sounds and outputs, the libraries will be pulled from multiple sources on the internet. By using multiple sources, the software can ensure that there is a variety of sounds available, reducing the impact of any single library becoming depreciated.

4.3.2.6 T-6: User Inputted Recognizability

To mitigate the risk of the software picking up background noises from user input, the system can isolate and ignore repetitive background frequencies. This can be done by analyzing the input for frequency patterns that are present throughout the input and filtering them out. By ignoring these repetitive frequencies, the system can focus on the more unique and recognizable musical signals in the input.

4.3.2.7 T-7: Applying Incorrect Correction Rules When Correcting Music Notation

To mitigate the risk of the program applying incorrect correction rules that cause it to make incorrect changes to the original music, users should be allowed to submit positive or negative feedback on each corrected note to help improve the ML model. This feedback can then be used to retrain the model to make more accurate corrections in the future.

4.3.2.8 T-8: Bias In The Correction Process

The program might prioritize certain musical styles or genres over others which would introduce errors in the output sheet music. To mitigate the risk of bias in the correction process, it is important to ensure that the machine learning (ML) model is designed to be free from any bias. This can be achieved by selecting a diverse dataset of music from various genres and styles, and ensuring that the ML model is trained equally on each type of music.

4.3.3 Security Risks

Security risks refer to the potential vulnerabilities and threats associated with the protection of data, systems, and infrastructure. A visualization of those risks can be seen in Table 6.

[This space intentionally left blank]

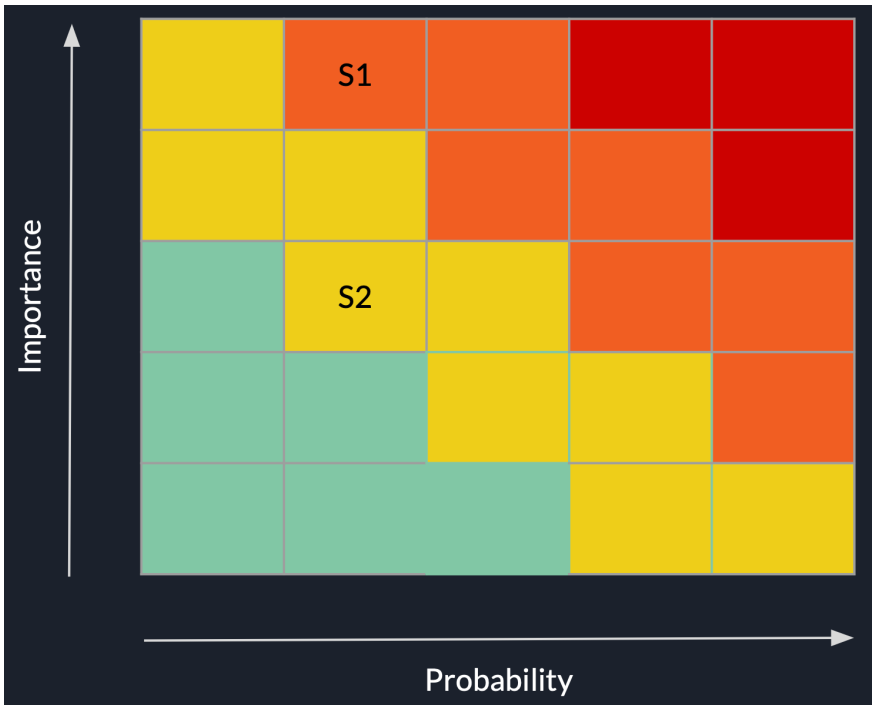

Table 6: Security Risks

4.3.3.1 S-1: Intellectual Property

To prevent data leaks and unauthorized access to unreleased content, the software should implement strong measures for data privacy and security. These measures should include encryption of data at rest and in transit, access controls to limit user access to only necessary data. By implementing these measures, the software can mitigate the risk of data leaks and protect the intellectual property of content creators.

4.3.3.2 S-2: Application Vulnerability

To mitigate the risk of application vulnerability, it is essential to implement best security practices. This includes regular security audits, code reviews and testing to identify any weaknesses in the system. The application must be designed with security in mind and follow industry-standard protocols for secure communication and data storage.

4.3.4 Legal Risks

Legal risks denote the potential challenges and exposures arising from non-compliance with laws and regulations. A visualization of those risks can be seen in Table 7.

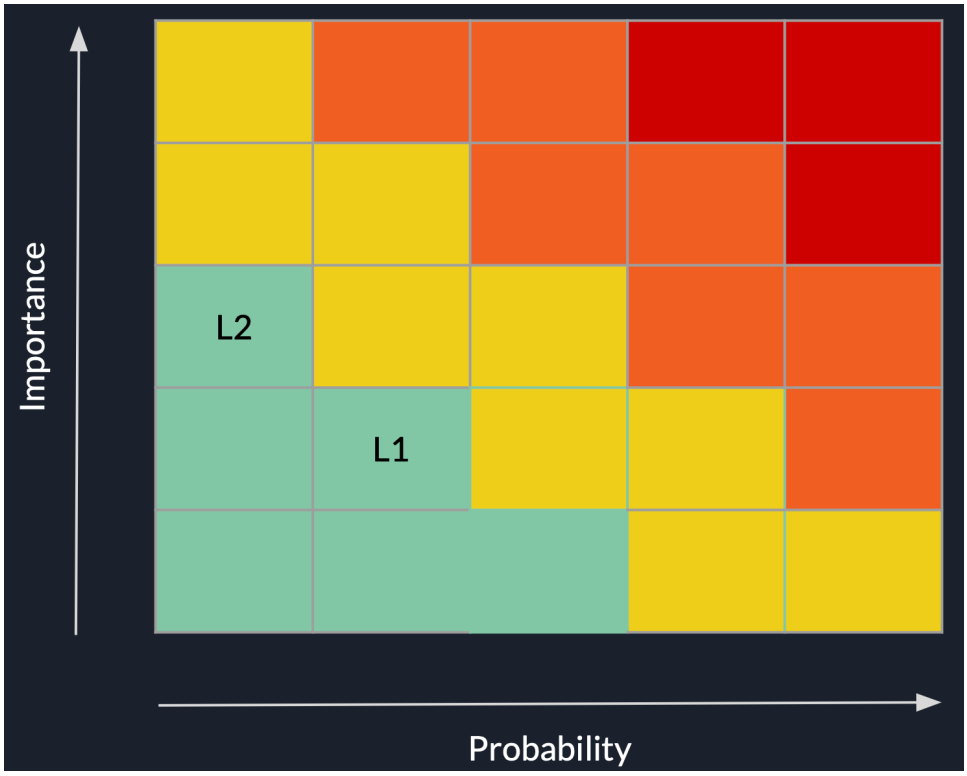

Table 7: Legal Risks

4.3.4.1 L-1: Unlawful Product Use

Unlawful product use of the application can be a significant risk, as customers may try to steal and distribute music or sheet music using the software. To mitigate this risk, it is important to take measures to discourage such behavior. One way to do this is to avoid creating an audio file of the music, as this could be used to illegally distribute the music. Adding a watermark to the sheet music can make it more difficult to use the music for unlawful purposes, as the watermark would indicate the source of the sheet music. By taking these precautions, the risk of unlawful product use can be significantly reduced.

4.3.4.2 L-2: Copyright Law

The use of copyright works in datasets used for training AI systems can have legal implications. There is a risk that the use of copyrighted material could lead to a copyright infringement lawsuit. To mitigate this risk, using non-copyrighted works as training data can be a safe alternative. In cases where the use of copyrighted material is necessary, it is important to obtain the necessary licenses or permissions from the copyright holders to use the material for training. The licenses or permissions can be granted through agreements with the copyright owners.

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/