Sheet Music Transcriber

Company 1

HappyJam

Overview

- Team Introduction
- Introduction
- Background
- Technical Case
  - Algorithm
  - UI Functionality
  - UI Appearance
- Proof of Concept
  - Algorithms
  - UI
- Testing and test cases
  - Algorithms
  - UI
- Business Cost
- Market Research
- Engineering Standards
- Schedule and plan for ENSC 440
- Technical design and research
- Risk Management
- Team Reflection
- Conclusion
- Questions

Team Introduction

Company 1

Matthew Marinets	CEO and Software Engineer	Algorithms and Rendering
Polina Bychkova	CCO and Software Engineer	UI functionality development
Akaash Parajulee	Electronics Engineer	Algorithms development
Jaskirat Arora	Systems Engineer	UI development
Haoran "Harry" Hu	Software Engineer	UI development
Avital Vetshchaizer	Software Engineer	UI development

Introduction

- Some of our members are musicians and like to play music
- Learning an existing song is normally done through sheet music, which is music written down in a document called a score
- Playing or learning music from a score is straightforward, but transcribing music (going from listening to a score) is much more difficult
- We are developing the SMT, which will import an audio file to:
  - Create visual displays to help keep track of the musical content
  - Automatically detect notes and their qualities, such as pitch, timbre, duration, and volume
- This will immensely speed up the process of transcription, and will automatically create a working score that can be further modified in any music notation software

Background

- Computer audio is just a list of samples – a digital signal
- We are looking to detect pitches (frequencies) at specific times, so we need a time-frequency representation of the signal
    - We use a spectrogram – a two-dimensional representation of a signal's power, with time on one axis and frequency on another
- From there we can apply image processing techniques, such as thresholding, peakfinding, 2D convolution, or the Hough transform
- This can generate a list of Points of Interest (POIs) which may be percussive notes, pitched notes, or harmonics
- From there, we can group these points into regions classify and associate these regions into notes
    - Percussive notes take up a wide frequency space with a bandpass-noise structure
    - Pitched notes and harmonics are narrowband, with harmonics occurring at integer frequency multiples of the fundamental frequency (the root of the pitched note)
- Note pitch is determined by the frequency of the fundamental, note time and duration by the position and length of the note region along the time-axis
- Note timbre may be determined by looking at the relative strengths of the harmonics

Background - a spectrogram of a single note

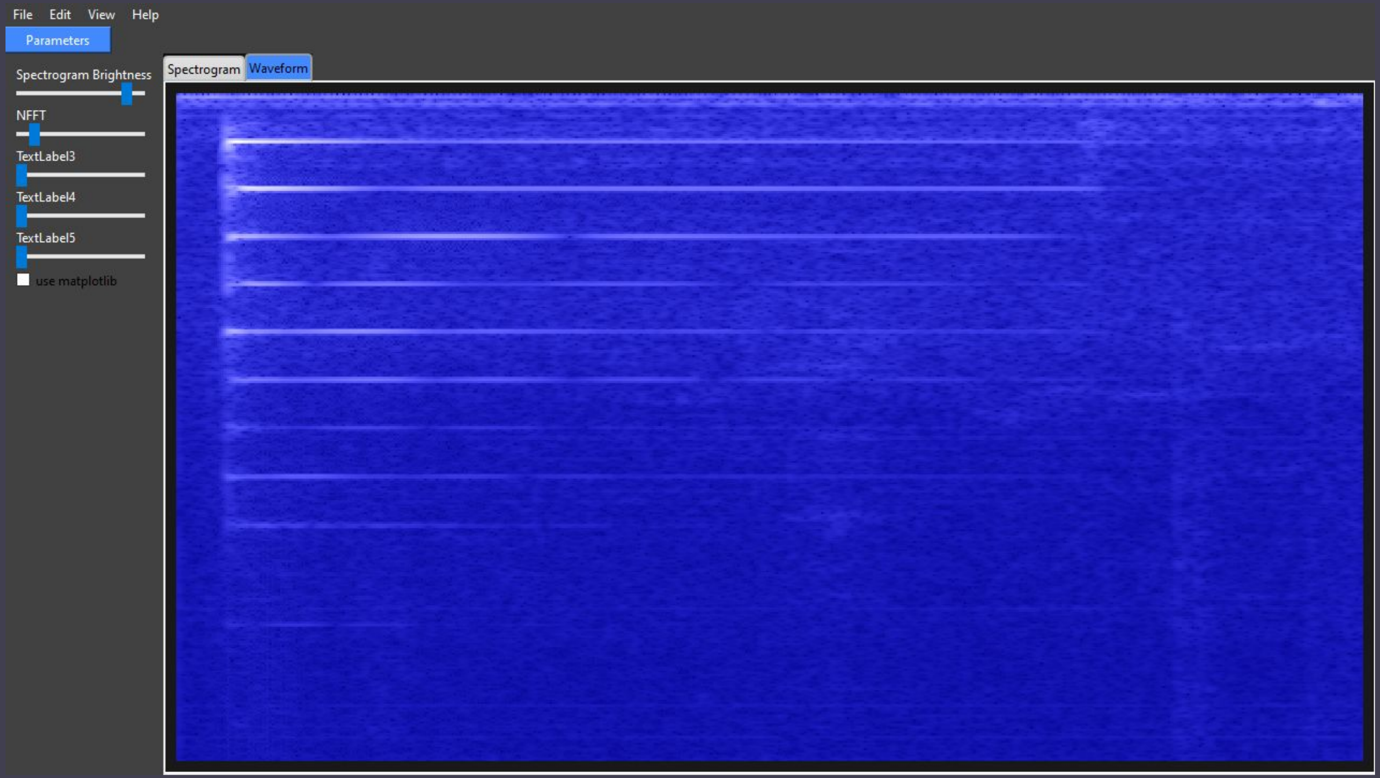


Figure 1: Spectrogram of a single note

Technical Case — UI Appearance

- Figma is used for the concept design
- Pyside 6 and QT are used for graphical components
- Python 3 used for animating widgets

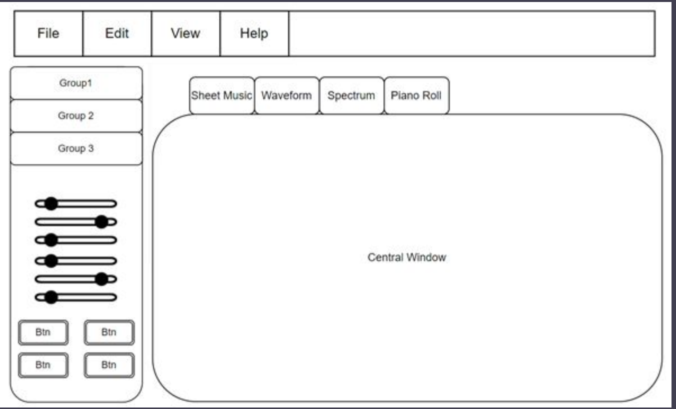


Figure 4: The aimed UI design for SMT

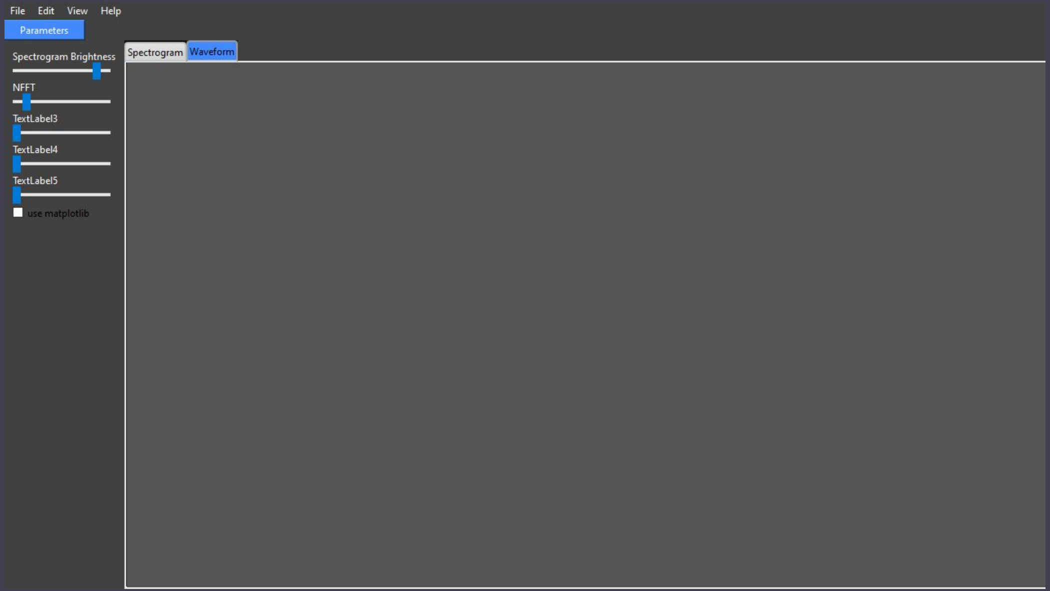


Figure 5: The achieved UI design for SMT

Technical Case – Algorithms

- We have prototypes showing note regions in a spectrogram
- We have additional functionality for finding likely note onsets
- We are currently researching separating percussion from pitched components with local variance or Hough transform
- Regions will then be chunked

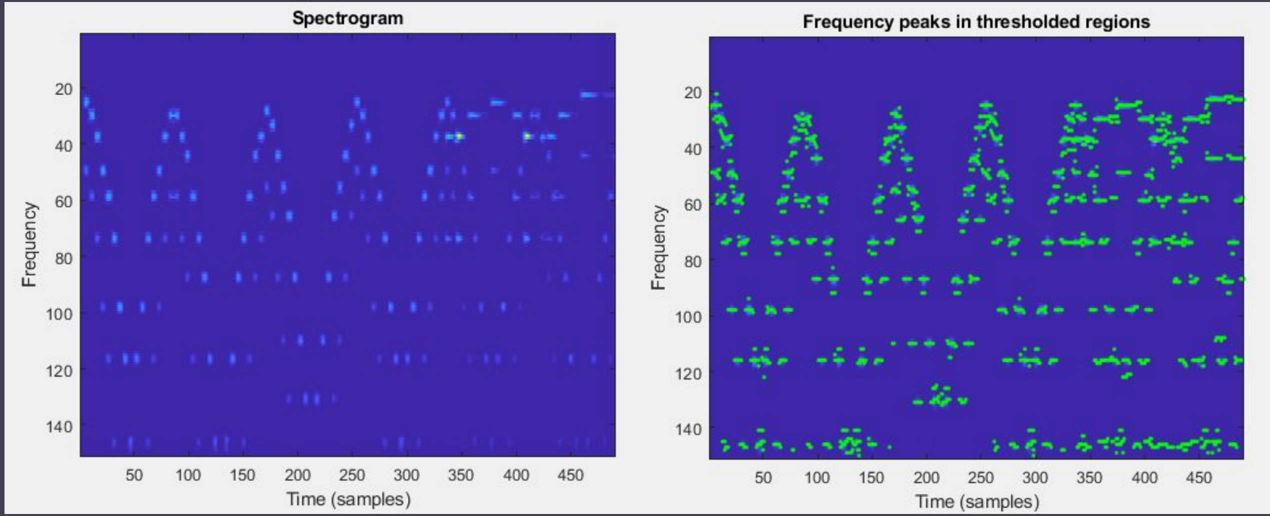


Figure 2: A spectrogram with points of interest marked

Technical Case – UI Functionality

- Backend functionality of the proposed solution
  - A user uploads .wav into Main Window along with analyses parameters
  - The file is processed by the algorithms, considering user's input parameters
  - After the processing, the result is displayed within the Viewport
- Project Modules
  - UI
  - Backend: Algorithms and State Management
  - Disk: storage of intermediate result
- Materials Utilized:
  - Python 3, PySide 6 QT binding, OpenGL potentially
- Changes in Scope, Design and Functionality of PoC prototype for Phase A
  - Visually identified notes display pitch, timing, and duration of notes moved to Phase B
  - Saving and loading project state

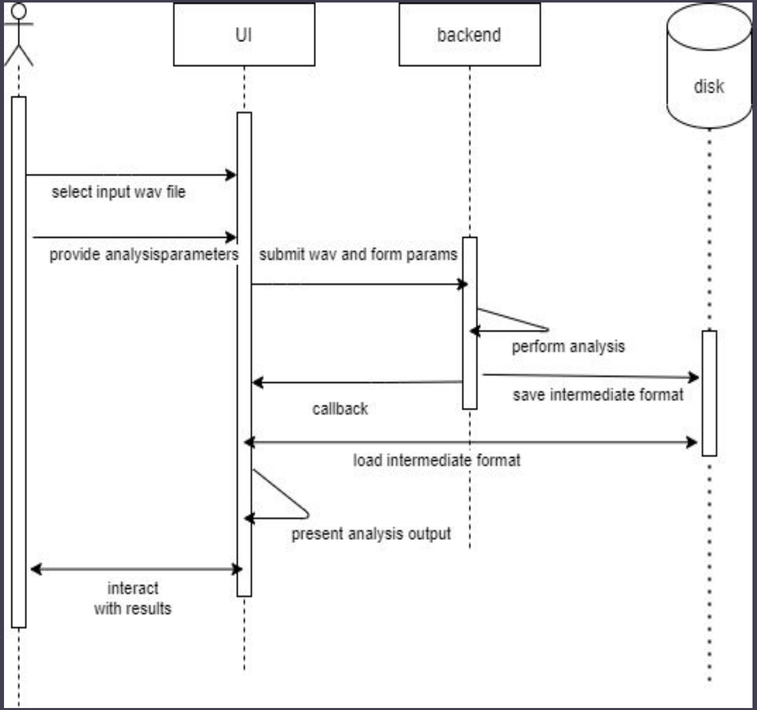


Figure 3: Program Workflow

Proof of Concept - Algorithms

- Timbre analysis as a concept is not fully proven; note detection has already been done by other products
- We can generate a spectrogram, find points of interest, and compile that into a chart
- There is still a lot of chatter and unwanted signal, and timbre analysis will not happen until ENSC440
- Algorithm development has been paused until the UI is sufficiently functional to offer interactive sliders and improved display
  - Improved interactivity will allow us to better understand which algorithms aren't working and which simply require specific parameter inputs
- Many algorithms are on the to-do list, including:
  - The Hough Transform
  - Local variance (thanks, Rodney)
  - Modified convolution feature-identification
- Neural networks remain a backup, with Ivan Bajic recommended as the person to go to

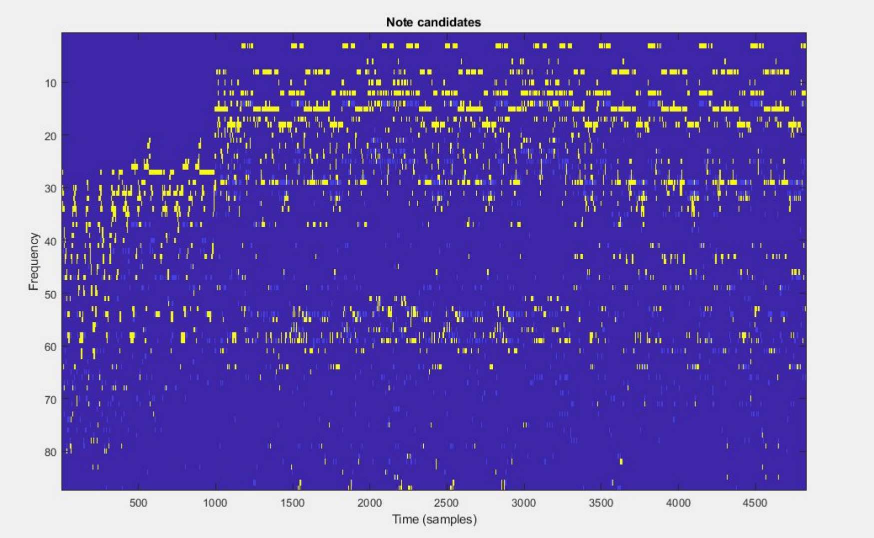
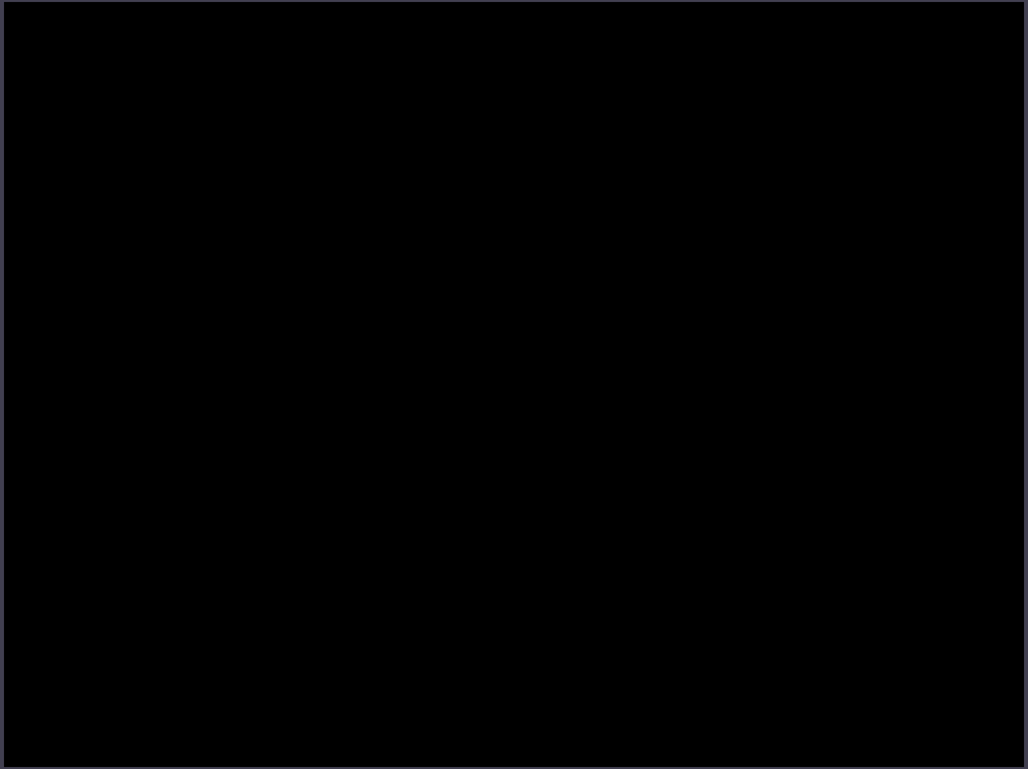


Figure 6: A plot of detected note candidates

Proof of Concept – UI

Recorded demo of the prototype: .wav file upload, Spectrogram navigation and visual note identification, UI parameter values adjustment



Testing and Test Cases - Algorithm

Testing ID	Description	Result	Comments
D-A.3.1.1-A	The SMT identifies the times at which an instrument plays a note in an audio sample	Success*	Provable in Matlab, TBI
D-A.3.1.2-A	The SMT identifies the pitches at which an instrument plays a note in an audio sample	Success*	Provable in Matlab, TBI
D-A.3.1.3-B	The SMT identifies the lengths of a notes played by an instrument in an audio sample	-	
D-A.3.1.4-B	The SMT identifies the tempo of the sample audio	-	
D-A.3.1.5-B	The SMT identifies the relative volume of a notes played by an instrument in an audio sample	-	
D-A.3.1.6-V1	The SMT identifies a tempo changes in the audio sample	-	
D-A.3.1.7-A	The SMT algorithm packages data processed by the algorithm in a format usable by other portions of the program and the user	Partial	Underlying data structure exists

Table 1: Test Cases for Algorithm 1

Testing ID	Description	Result	Comments
D-A.3.2.1-A	The SMT identifies and processes notes from a single instrument audio sample	Success*	Provable in Matlab, TBI
D-A.3.2.1.1-A	The SMT identifies the notes of a piano audio sample and displays them accurately	Success*	Provable in Matlab, TBI
D-A.3.2.1.2-A	The SMT identifies the notes of a guitar audio sample and displays them accurately	To be tested	Will tested after UI integration
D-A.3.2.1.3-B	The SMT identifies the notes of a drum-set audio sample and displays them accurately	-	
D-A.3.2.2-A	The SMT identifies the overtones of a single instrument and eliminate them in an audio sample	Partial	Overtones identified
D-A.3.2.2.1-A	The SMT identifies which instrument is playing in an audio sample based on a timbre profile	Not attempted	Pushed to next semester

Table 2: Test Cases for Algorithm 2

Testing and Test Cases - UI

Testing Item ID	Description	Result
D-5.1.2-A	User opens the software by double clicking on the program icon	Success
D-5.1.3-A	User should be able to use dark theme	Success
D-5.1.4-A	User can use the menu bar to manipulate files	Success
D-5.1.4.1.4-A	User can import a wav (.wav) file	Success
D-5.1.5-A	User can use the left side panel to adjust parameters	Success
D-5.1.6-A	User can use top bar to edit views	Success
D-5.1.8-A	User can see different views in the central window	Success
D-5.1.8.2-A	User can see waveform view	Success

Table 3: Test Cases for the user interface 1

Testing Item ID	Description	Result
D-5.1.8.3-A	User can see the spectrum view	Success
D-5.1.8.5-A	User should be able to zoom in and out of the various views	Success
D-5.1.8.7-A	User should be able to close views	Success
D-5.1.8.8-A	User should be able to swap views	Success
D-5.1.9-A	User can stop the software and close the window	Success
D-5.1.9.1-A	User should be able to see a popup window warning the user when exit	Pending
D-5.1.9.1.1-A	User should be able to disable the popup window and exit directly	Pending
D-5.1.10-A	User can minimize the window	Success

Table 4: Test Cases for the user interface 1

Business Cost

Our project has no cost at the proof of concept stage since there is no hardware associated with it at this stage. Our final project does come with a microphone recommendation for the users that will be optimum for audio pickup.

The following is a table of commonly used recording devices:

Microphone Pick-up Pattern	Estimated Cost (CAD)
Omnidirectional	26 [1]
Cardioid	31 [2]
Hypercardioid	32 [3]
Supercardioid	22 [4]
Unidirectional	29 [5]
Total Cost for Microphones	140

Table 5: Cost for different pick-up pattern microphones

Market Research

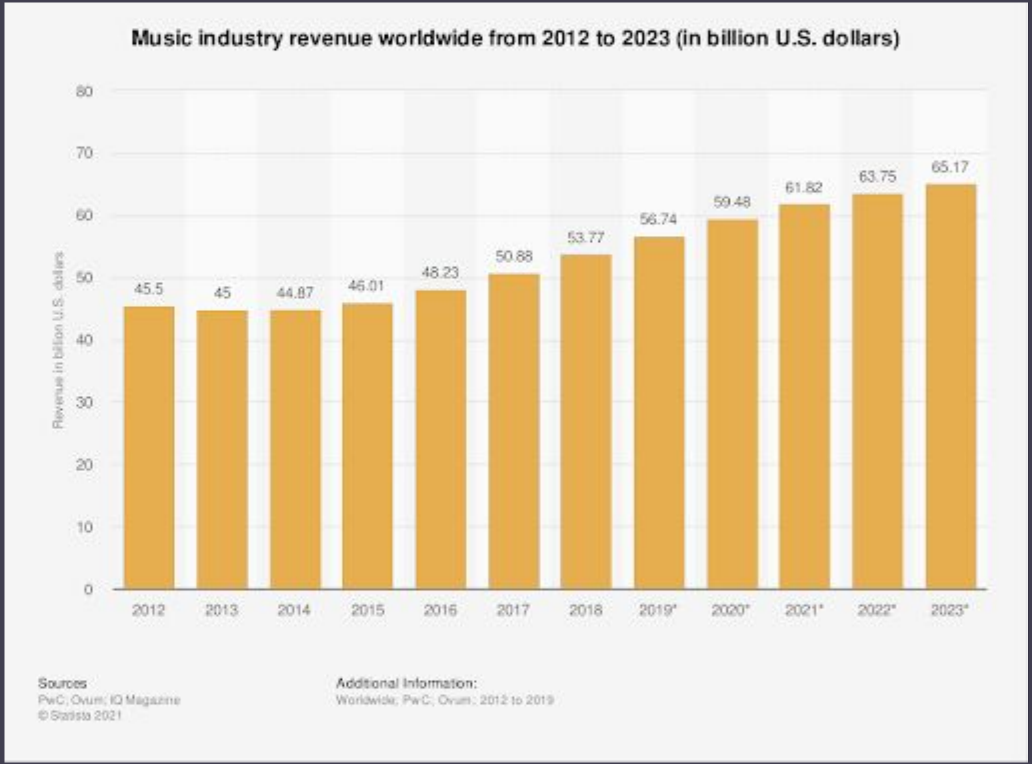


Figure 7: Music Industry Revenue Worldwide from 2012 to 2023 [6]

Market Research - continued

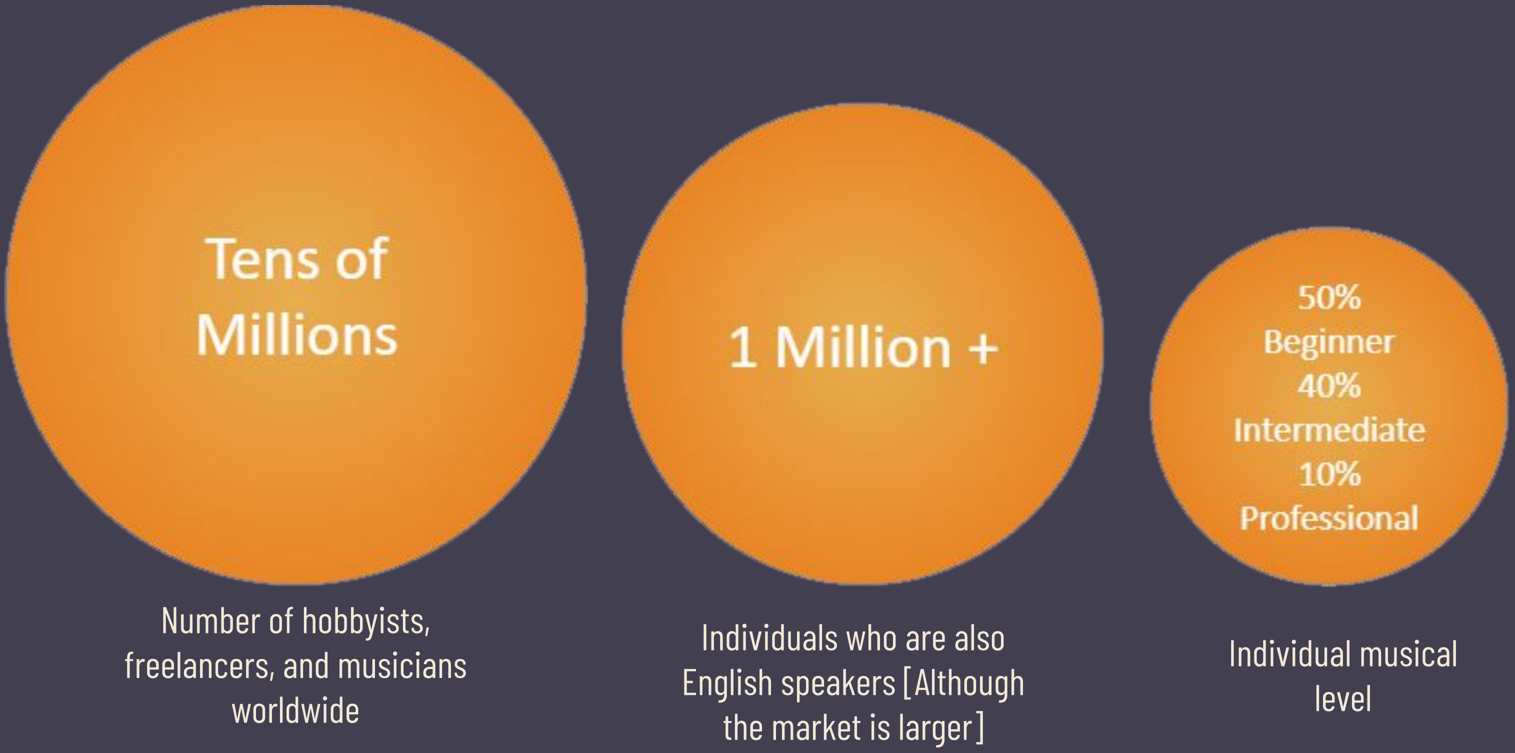


Figure 8: Market Size [7]

Market Research - continued

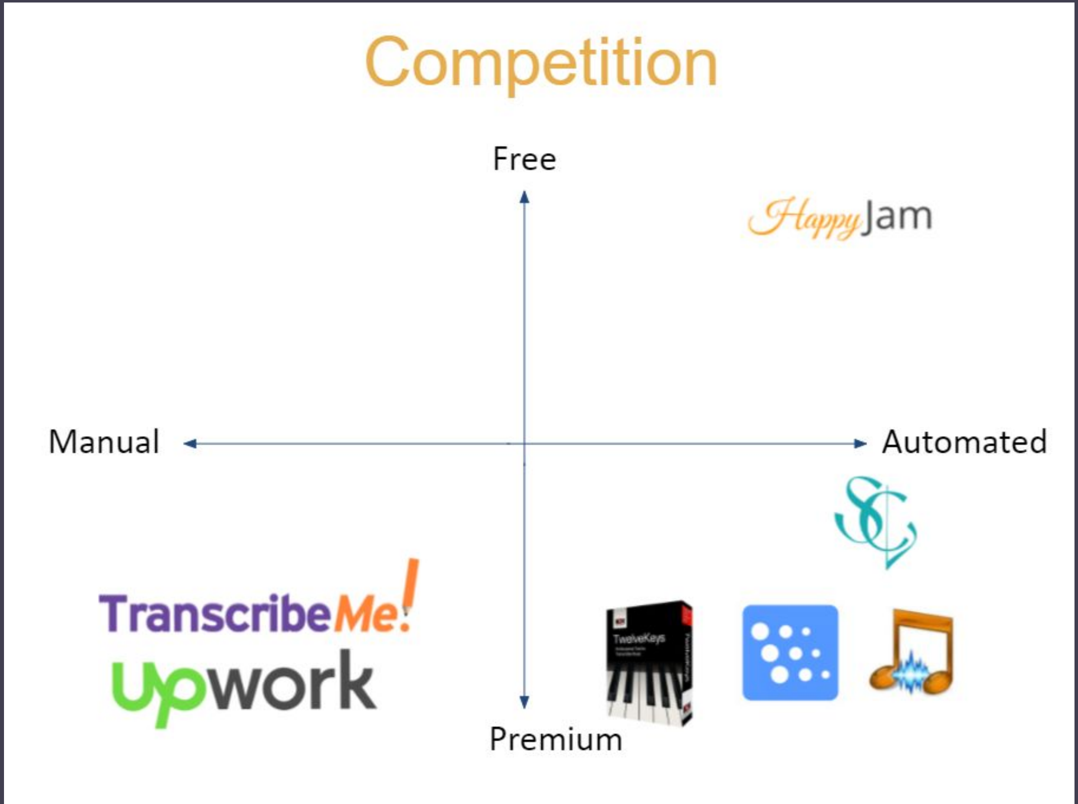


Figure 9: Competitive Softwares

Engineering Standards

Engineering Standards	Description
ISO/IEC 14496-23:2008	Information technology — Coding of audio-visual objects — Part 23: Symbolic Music Representation
ISO/IEC 23000-12:2010	Information technology — Multimedia application format (MPEG-A) — Part 12: Interactive music application format
ISO 10957:2009	Information and documentation — International standard music number (ISMN)
IEC 60417	Graphical Symbols for Use on Equipment
ISO/IEC 24752-8:2018	Information technology — User interfaces — Universal remote console — Part 8: User interface resource framework
ISO/IEC TR 29119-11:2020	Software and systems engineering — Software testing — Part 11: Guidelines on the testing of AI-based systems
ISO/IEC TR 24029-1:2021	Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview
ISO 9241-210:2019	Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems
ISO/IEC 11581-5:2004	Information technology — User system interfaces and symbols — Icon symbols and functions — Part 5: Tool icons
ISO/IEC 29138-1:2018	Information technology — User interface accessibility — Part 1: User accessibility needs
ISO/IEC 23007-2:2012	Information technology — Rich media user interfaces — Part 2: Advanced user interaction (AUI) interfaces

Table 6: Engineering Standards followed in our project

Schedule for ENSC 405

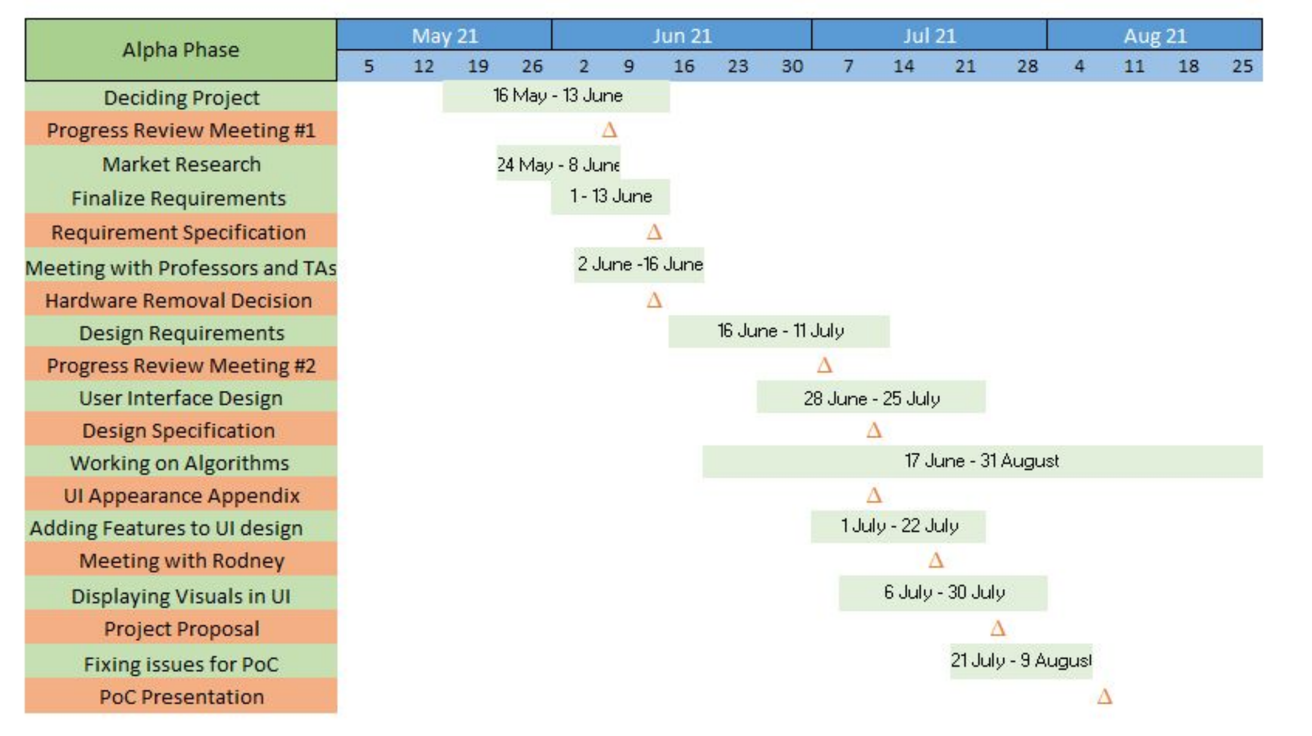


Figure 10 : Gantt Chart for schedule of ENSC 405

Schedule and Plan for 440

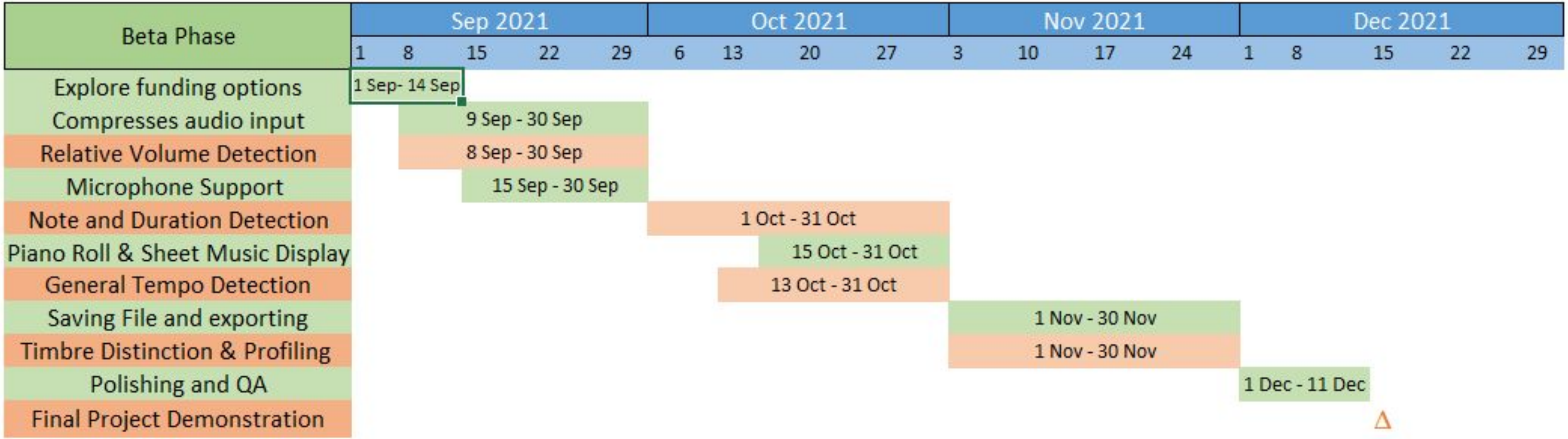


Figure 11: Gantt Chart for our ENSC 440 Plan

Technical Design and Research

C++/Python -> Python

- Most functions in qt are written in C++
- Python is easy to use for backend developing

Qt -> Qt + OpenGL

- Qt image display requires using objects that use too much memory and cause crashes
- As a result, we had an initial design that divided images into smaller pieces and did manual memory management in Python, but there were still problems modifying display images
- As a result, we are moving to a Qt + OpenGL solution that allows us to keep images in the same format that is output by the algorithms, and we can modify how they are displayed without modifying the image data itself

Risk Management

- Agile resources distribution, human resources can be quickly adjusted
- Programs are made translatable, so that we can always change languages or tools
- Have regular meetings every week, so that the information is updated and progress is being tracked
- Save progresses on GitLab, and create issues to track the progresses
- Have viewer to ensure the quality of the merging code

Team Reflection

- Documentation organisation
  - Delegating tasks and setting clear deadlines
  - Figuring out how each member works differently
- Coding organisation
  - Two different teams: UI and Algorithmic
  - Re-allocated human resources
  - Git branches and merge requests
- ENSC 440
  - Be dynamic in re-assigning people to different objectives
  - Use branches and merge requests more
  - Establishing internal deadlines

Conclusion

ENSC 405 has been as rewarding as it was challenging. We were able to improve our teamwork and communication skills.

Despite restricted time to develop the product due to the documentation component of the course our company was able to design the User Interface, input .wav files and display its spectrogram. Our algorithm can find points of interests for note detection in the spectrogram.

We will continue to work with focus and dedication towards completing our project in the coming term while overcoming any problems we face during its course.

HappyJam

Questions?

Thank you and have a Happy day!

References

[1] PoP voice, "Professional #1 Best Lavalier Lapel Microphone Omnidirectional Condenser Mic for Apple IPhone Android & Windows Smartphones,Youtube,Interview,Studio,Video Recording,Noise Cancelling Mic," amazon.ca, Oct. 17, 2016. [Online] Available: https://www.amazon.ca/Professional-Microphone-Omnidirectional-Smartphones-Cancelling/dp/B016C4ZG74/ref=sr_1_17?crid=3TNBHXX7OVOYS&dchild=1&keywords=omnidirectional+microphone&qid=1627001235&sprefix=omnidirectional+mi%2Caps%2C362&sr=8-17 [Accessed: July 22, 2021]

[2] Amazon Basics, "AmazonBasics Dynamic Vocal Microphone – Cardioid," amazon.ca, Dec. 8, 2018. [Online]Available :https://www.amazon.ca/AmazonBasics-Dynamic-Vocal-Microphone-Cardioid/dp/B07L6C5VRW/ref=sr_1_9?dchild=1&keywords=cardioid+microphone&qid=1627001339&sr=8-9 [Accessed: July 22, 2021]

[3] Akozon, "Recording Microphone, Akozon Super Hypercardioid Directional Shotgun Microphone for Live Photography Camera Interview," amazon.ca, Apr. 18, 2020. [Online]Available: https://www.amazon.ca/Microphone-Akozon-Hypercardioid-Directional-Photography/dp/B0877VNX83/ref=sr_1_46?dchild=1&keywords=hypercardioid+microphone&qid=1627001531&sr=8-46 [Accessed: July 22, 2021]

References

[4] Maono, “Camera Microphone, MAONO AU-CM10 Super-Cardioid Condenser Video Interview Mic for Canon, Nikon, Sony DSLR Cameras, Camcorder, PC, Phone, Android
,” amazon.ca, Oct. 7, 2019. [Online]Available: https://www.amazon.ca/Microphone-MAONO-Super-Cardioid-Condenser-Interview/dp/B07YTRQFRP/ref=sr_1_20?dchild=1&keywords=supercardioid+microphone&qid=1627001715&sr=8-20 [Accessed: July 22, 2021]

[5] PYLE, “Pyle Professional Dynamic Vocal Microphone - Moving Coil Dynamic Cardioid Unidirectional Handheld Microphone with ON/OFF Switch Includes 15ft XLR Audio Cable to 1/4" Audio Connection - PDMIC59,” amazon.ca, Feb. 11, 2016. [Online]Available: https://www.amazon.ca/Pyle-PDMIC59-Professional-Microphone-Unidirectional/dp/B01B1JHEX4/ref=sr_1_9?dchild=1&keywords=unidirectional+microphone&qid=1627001863&sr=8-9 [Accessed: July 22, 2021]

[6] Statista Research department, “Music Industry Revenue Worldwide from 2012 to 2023,” statista.com, June 10, 2021. [Online] Available: https://www.statista.com/statistics/259979/global-music-industry-revenue/ [Accessed: July 23, 2021]

References

[7] “The job market for music teacher market in the United States,” careerexplorer.com
[Online] Available: https://www.careerexplorer.com/careers/music-teacher/job-market/ [Accessed: July 23, 2021]