
anna-T
Automated Non-Neural Audio Transcription
Keshav Nath

01
Problem & Motivation

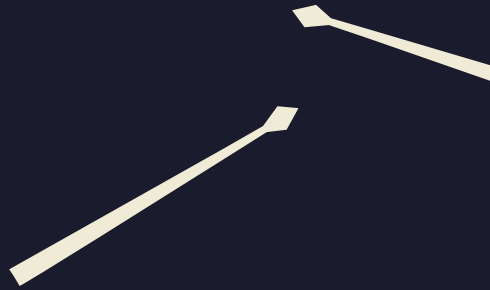
02
Demonstration

03
Approach

04
Results

05
Future Developments

06
Wrap Up

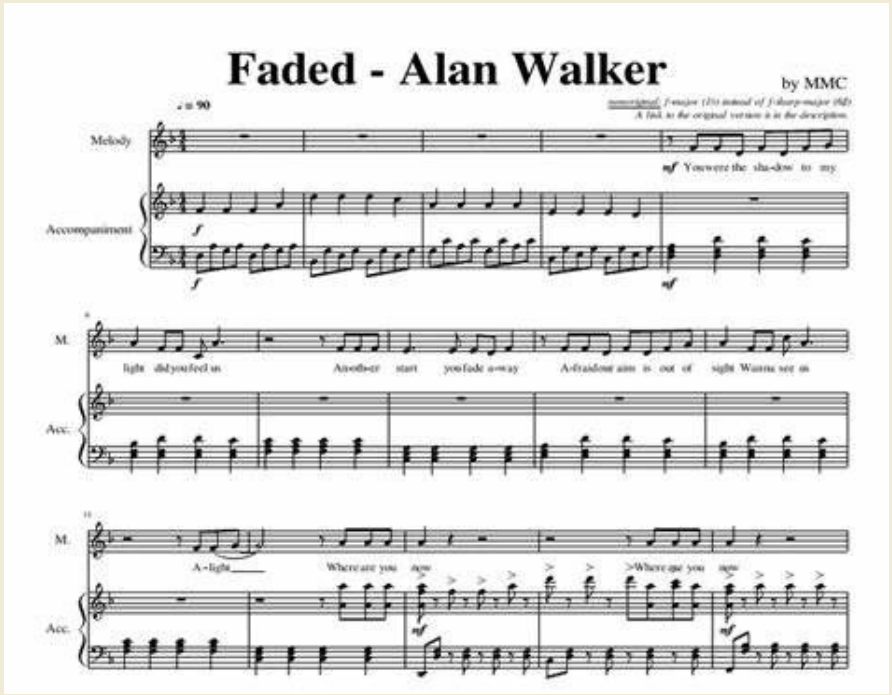


Problem

Current Transcription Services

- Largely manual
- Limited libraries
- Focus on classical music

Piano Sheet sample for Alan Walker's Faded



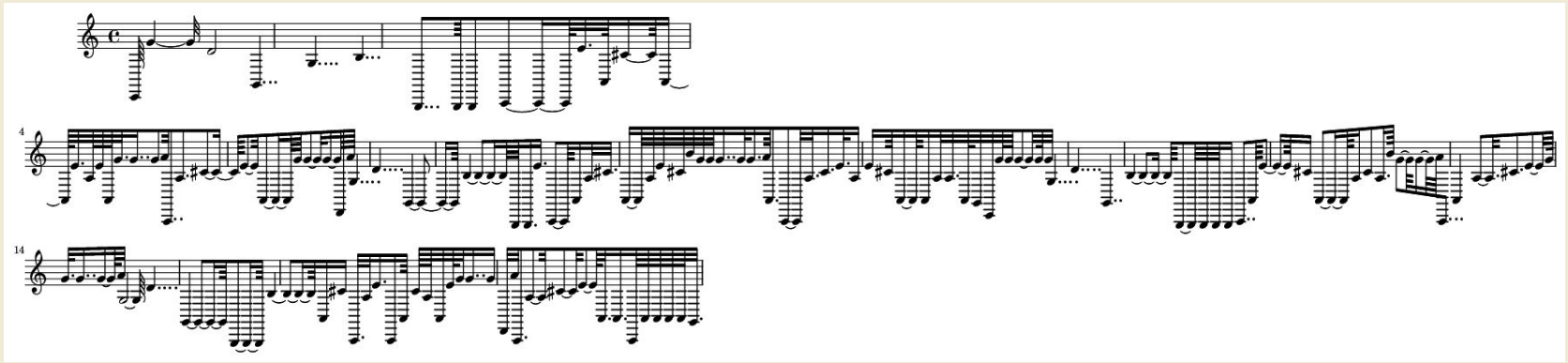
Motivation

People should be able to take any song they like and obtain its Sheet Music. This ensures that they don't have to wait for a song to get popular and get transcribed by someone on the internet.

Key Product Features

- Handles standard issue .mp3 files (having .wav is not necessary)
- Custom Encoder that allows model to be retrained with any dataset
- Provides Sheet Music in simple PNG Format
- Trained on pop and rock music, so it can transcribe modern music well

Demo



Sheet generated from Calvin Harris - My Way

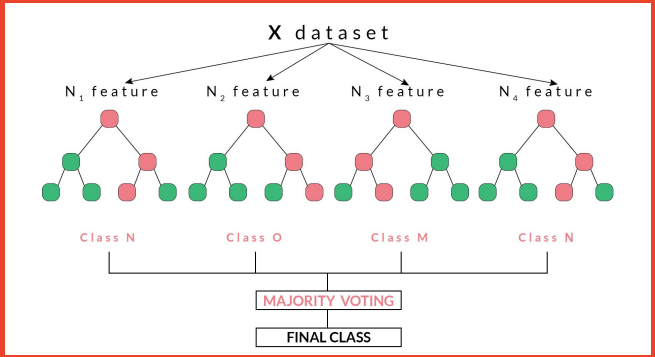
Technical Approach Summary

Model Input

- MIDIs and their respective mp3s from Cprato.com
- Split data into Train and Validation
- Transform Audio to CQT Spectrum
- Encode MIDIs as framewise One-Hot-Encoded Notes

Model

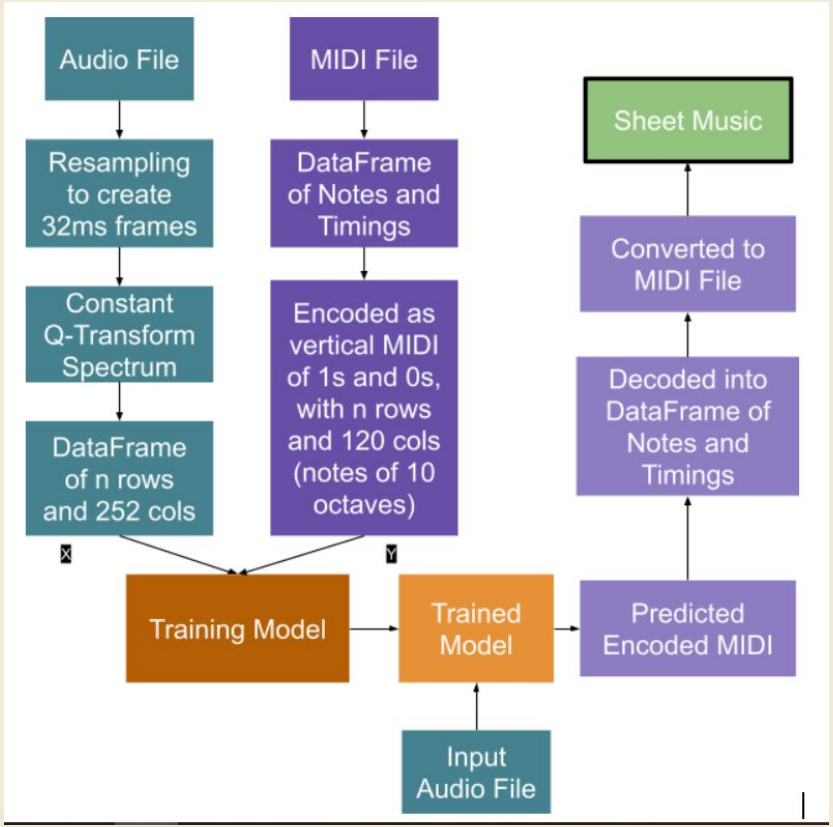
- Multi-Out Random Forest



Model Output

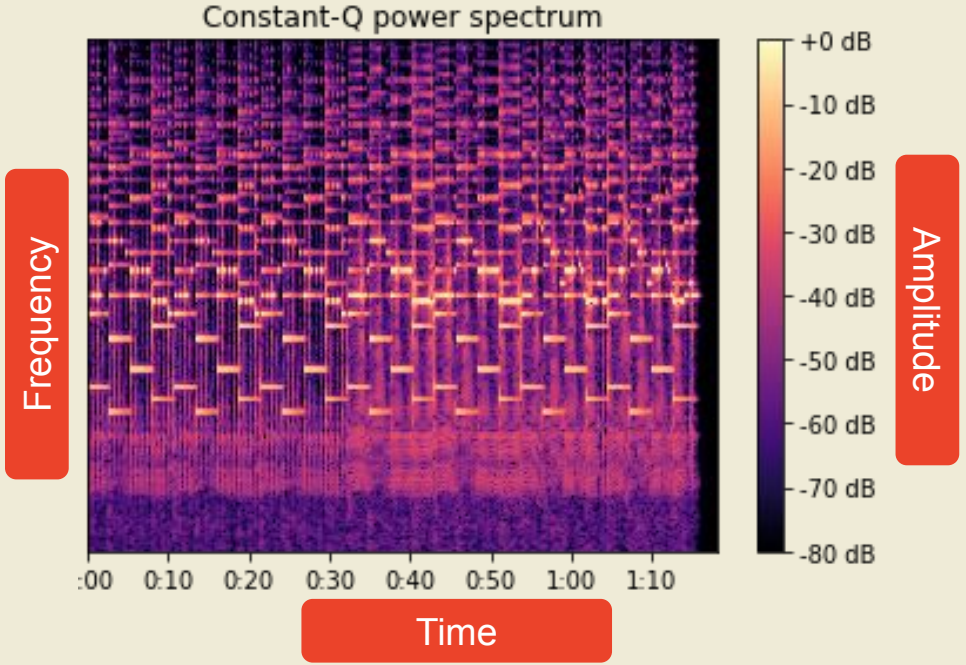
- Encoded Midi -> Decoder -> MIDI Stream -> Sheet Music (PNG)

Technical Approach Summary



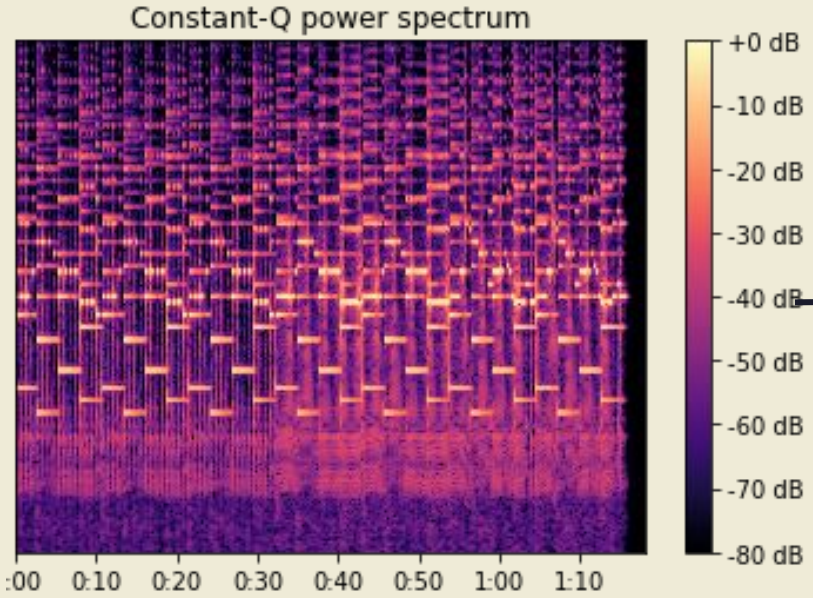
Audio Transformation and Model Input

Audio Sample



CQT (Constant Q-Transform) Spectrogram

Audio Transformation and Model Input



DataFrame of **252 features** and (1*60 + 18)/(0.032) = **2443 approx rows**

	0	1	2	3	4	5	6	
0	0.000733	0.003213	0.005608	0.005022	0.003947	0.002346	0.000042	0.0023
1	0.000765	0.003195	0.005601	0.004983	0.003936	0.002343	0.000182	0.0023
2	0.000844	0.003146	0.005584	0.004881	0.003906	0.002331	0.000332	0.0023
3	0.000948	0.003069	0.005557	0.004721	0.003855	0.002312	0.000460	0.0023
4	0.001068	0.002968	0.005516	0.004503	0.003784	0.002288	0.000556	0.0023

MIDI Encoding and Decoding

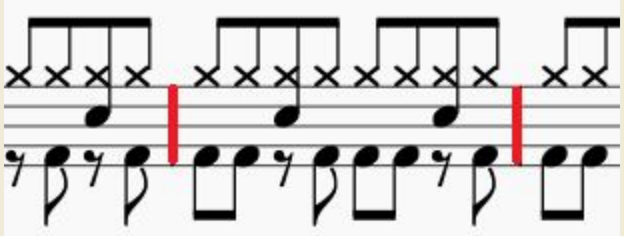
[	note_name	start_time	duration	velocity	tempo
0	F#4	0.0	0.500	100	90.0
1	E-3	0.0	4.000	100	90.0
2	F#4	1.0	0.500	100	90.0
3	F#4	2.0	0.500	100	90.0
4	B-4	3.0	0.500	100	90.0
..	...	...	...	...	...
196	F4	109.0	0.417	100	90.0
197	F4	109.5	0.417	100	90.0
198	F#4	110.0	1.000	100	90.0
199	G#4	111.0	0.500	100	90.0
200	F#4	111.5	1.000	100	90.0

	C1	C#1	D1	E-1	E1	F1	F#1	G1	G#1	A1	B-1	B1	C2	C#2	D2	E-2	E2	F2	F#2	G2	G#2	A2	B-2	B2	C3	C#3	D3	E-3
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Model Output and Postprocessing

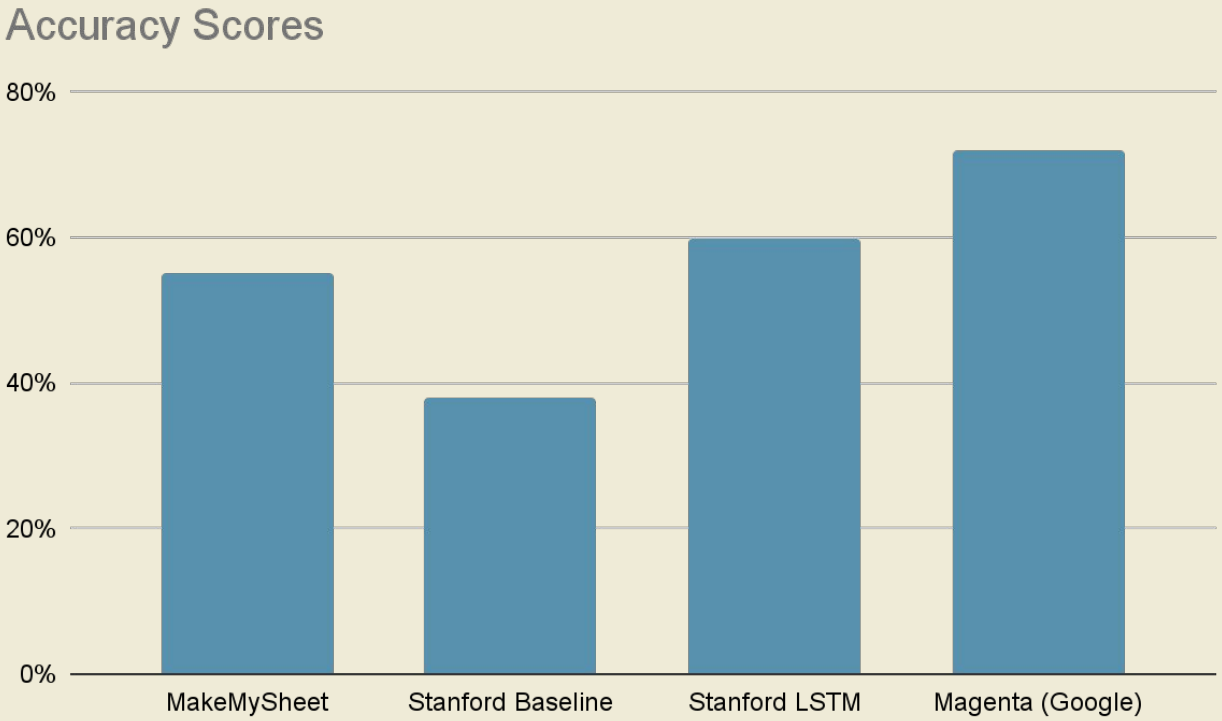
3- 1	B1	C2	C#2	D2	E-2	E2	F2	F#2	G2	G#2	A2	B-2	B2	C3	C#3	D3	E-3
.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

```
[     note_name  start_time  duration  velocity  tempo
0          F#4         0.0     0.500       100   90.0
1          E-3         0.0     4.000       100   90.0
2          F#4         1.0     0.500       100   90.0
3          F#4         2.0     0.500       100   90.0
4          B-4         3.0     0.500       100   90.0
..         ...         ...       ...       ...    ...
196         F4       109.0     0.417       100   90.0
197         F4       109.5     0.417       100   90.0
198        F#4       110.0     1.000       100   90.0
199        G#4       111.0     0.500       100   90.0
200        F#4       111.5     1.000       100   90.0
```



1. Multilabel Binary Classification output
2. Decoding of predictions back to MIDI Stream
3. Conversion of MIDI Stream into Sheet Music

Evaluation Comparison



Although existing methods do perform better, clearly our model belongs in the same ballpark and has the ability to one day be competitive.

Key Takeaways

Diversity

Instead of the standard piano (MAPS and MAESTRO) datasets, we created our own Dataset from Cprato.com to train using pop, EDM and rock music.

Decision Trees

Using Random Forest models instead of LSTM makes for a much speedier prediction and training time, compensating for the long preprocessing times.

Custom Encoder

Using a “home-made” encoder and decoder, our model can be trained on any dataset, including ones without premade labels

Future Work

Reinforcement

Feedback loop from user input to improve sheet music generation

Drums

Separating drum tracks (if present) and creating separate sheet(s) for them, as they muddle the normal sheets with a lot of low notes

Note Velocity

Additional data points give more insight into the ‘style’ of how a note is played

Ensembles

Transcribing music requires many different tasks. With more time and resources we hope to create additional models which can infer tempo, measures, genres, note durations, etc.

Refining

Training on a larger and even more diverse set of data for better accuracy scores with varied music

THANKS!

ANNA-T

By Keshav Nath