The Iby and Aladar Fleischman
Faculty of Engineering
Tel Aviv University

Auto-Transcriber for Electric Guitar

Project Number: 20-1-1-1980
Names: Gersh Yagudaev and Ethan Daniel
Advisor: Omer Tzidki

Introduction

Goal: Create a software system, which receives an amplified output of an Electric Guitar as its input, and uses Music Information Retrieval (MIR) techniques to attempt transcription of the sound.

Overview: Automatic Music Transcription is an active, and yet unsolved field of international research. There exist multiple bodies of work that attempt to implement systems necessary for transcription, such as pitch detection, onset detection, etc.

Motivation: We wish to draw upon the existing body of research to create a functional all-in-one real time transcription algorithm.

Desired Feature Set

- (Pseudo) Real Time Operation
- Note Segmentation – t_{start} , t_{stop}
- Pitch Detection - f_{note}
- Expression Detection: Vibrato, Bend, Slide
- Tablature Generation

Conclusions:

The industry is still far away from a human-competitive transcription system. The designed software achieves good results, with some limitations. It currently works with monophonic signals, and its detection abilities are focused to segmentation and pitch detection in real time.

Further Potential:

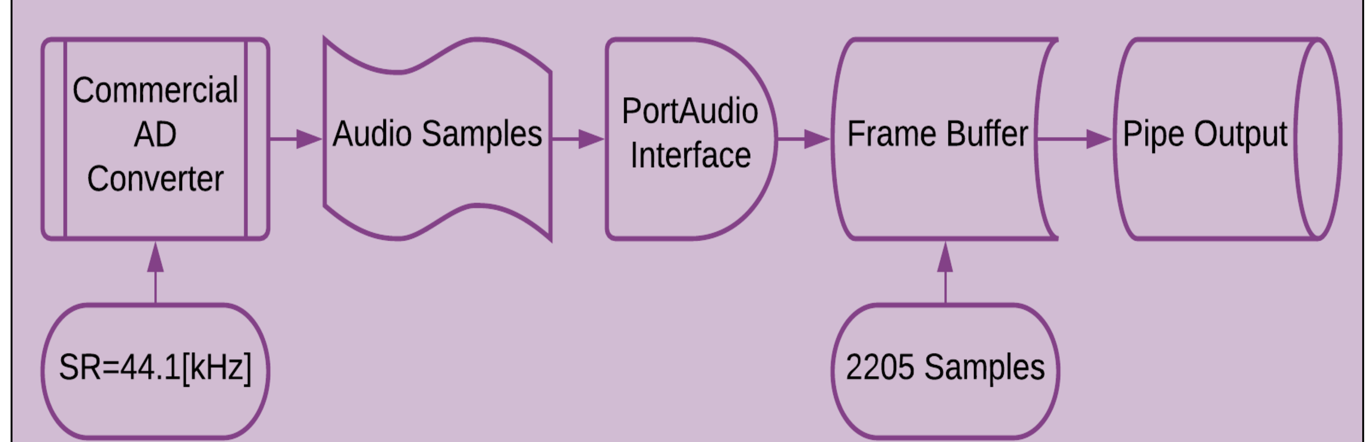
- Extend to Polyphonic Detection (eg by using Klapuri's algorithm, NMF approaches, or Neural Network)
- Complete Expression Detection to a reasonable fidelity
- Implement Heuristic Tab generation to account for "ease of playing"
- Implement Real-Time Optimizations

Reference:

Aasha, Diana Emerald, et al. "COMPARATIVE STUDY OF PITCH ESTIMATION USING HARMONIC PRODUCT SPECTRUM DERIVED FROM DFT, DCT, HAAR AND KL TRANSFORMS." *International Journal of Pure and Applied Mathematics*, 2017, acadpubl.eu/jsi/2017-115-6-7/articles/6/55.pdf.

Audio Input

Sampling from a commercial Audio Interface using PortAudio, grouping data into frames of 2205 samples each, corresponding to $50[ms]$ of audio. Send frames to segmentation process via pipe.



Tablature Generation

We generate a pure-text tab (ASCII Format), as this can be both read as-is and also converted to more sophisticated formats (eg *.gpx).

Tablature Generation Result

e	-	-	-	-	-	**0**	-	-	-	-	-	**0**	-	**5**	-	**7**	-	**0**
B	-	-	-	-	-	5	-	-	-	-	-	5	-	-	-	-	-	5
G	-	**6**	-	**4**	-	-	-	**7**	-	**6**	-	-	-	-	-	-	-	-
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

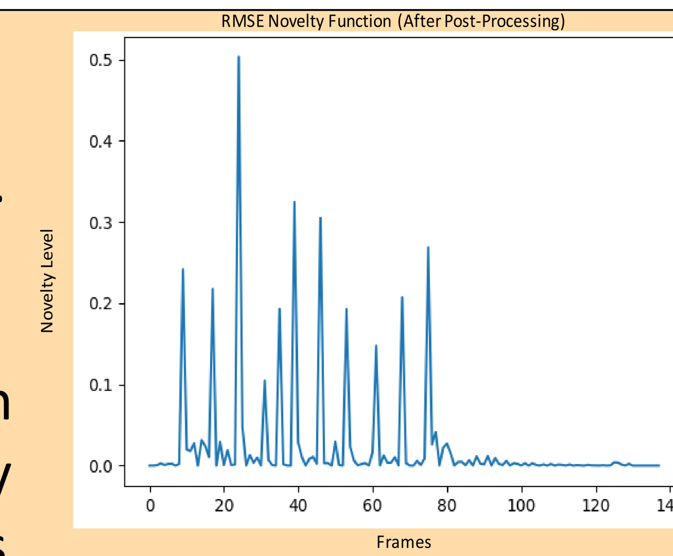
The numbers in bold are the actual output, with green cells indicating correct response. In the case of errors – the mistake is marked in red and the correct note is underlined. Note that in the case above the note e-0 is equivalent to B-5, so the mistake does not result in loss of musical information.

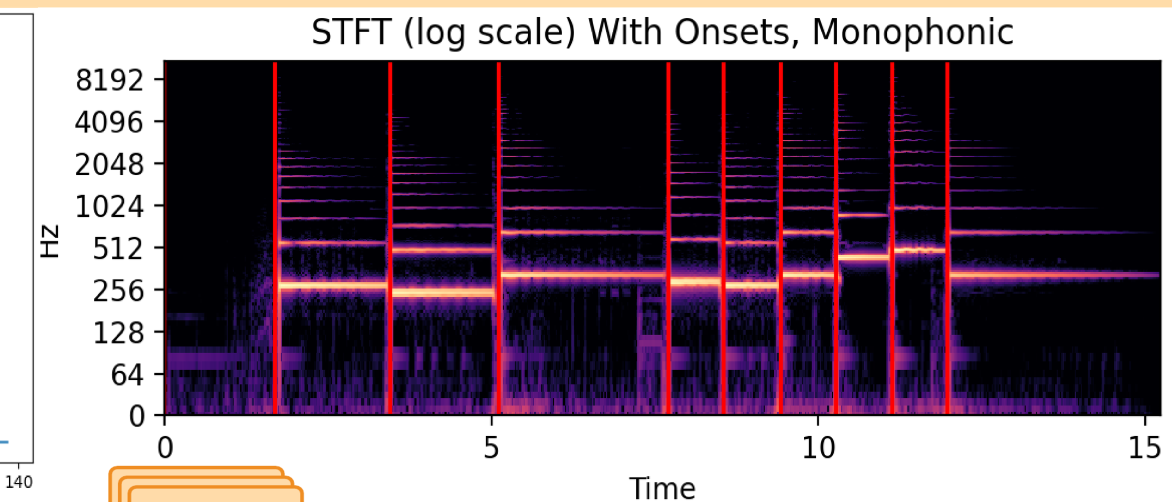
Note Segmentation:

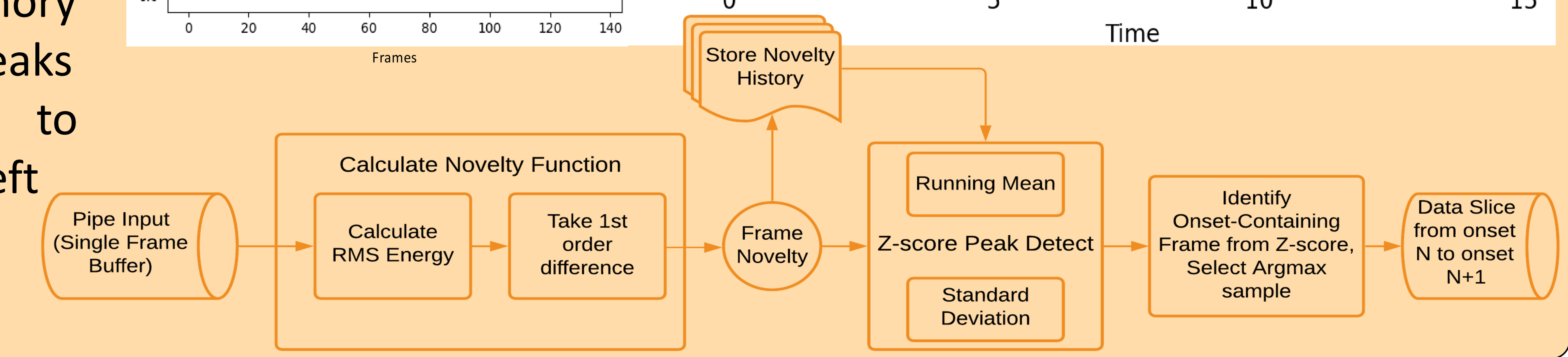
Root Mean Square Energy approach.

The Algorithm:

1. Receive input frame
2. Compute RMSE Novelty Function for the frame, append to memory
3. Suppress noise, Strengthen peaks
4. Detect peaks, backtrack to nearest local minima on the left
5. Convert peak location to time







Monophonic Pitch Detection:

We only hear the fundamental frequency as "pitch", while the harmonics define "timbre". Use 2 algorithms:

Autocorrelation:

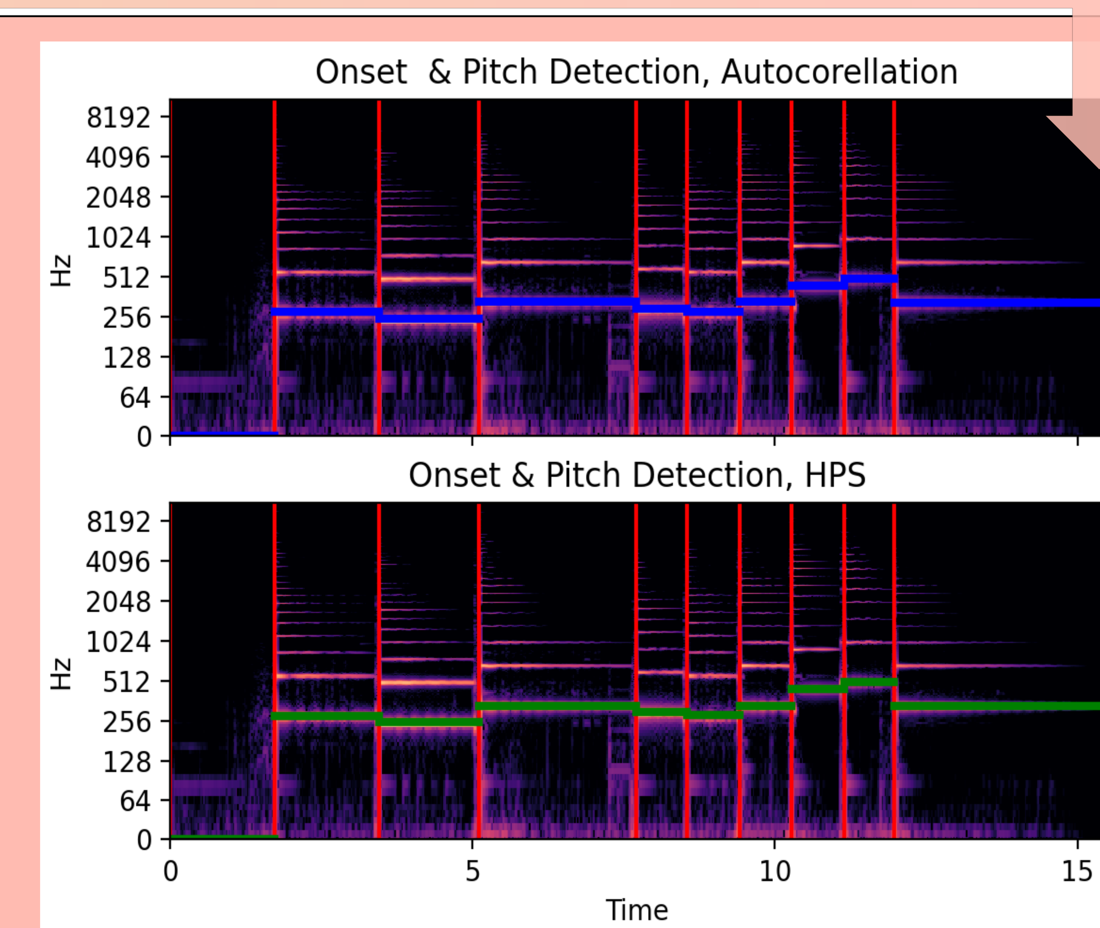
Time domain algorithm to estimate the fundamental period of a signal by finding the minimal value of m that maximizes:

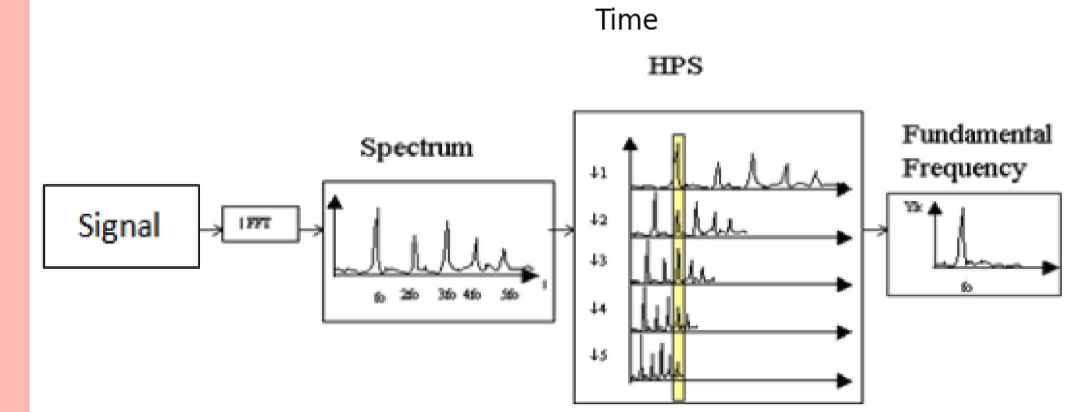
$$R[m] = \frac{1}{N-m+1}\sum_{n=m}^{N} x[n]x[n+m] \quad ; \quad m \neq 0$$

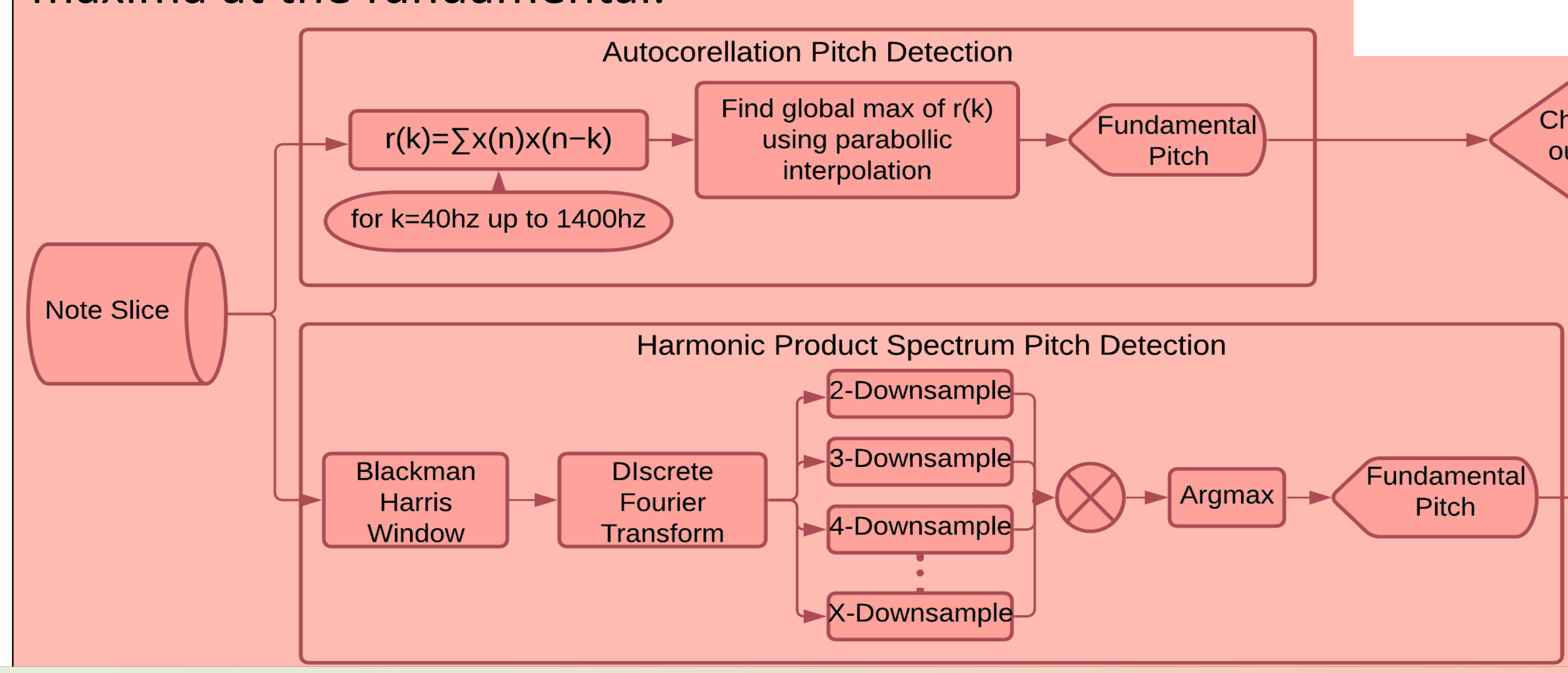
Harmonic Product Spectrum (HPS):

Note Spectrum: $f_{note} = \{f_0, 2f_0, 3f_0, \dots, Nf_0\}$

If we down-sample the FFT, our new spectrum will contain a peak at the fundamental. A product of multiple downsampled instances will yield a global maxima at the fundamental.







Truth (MIDI)	Truth (Hz)	Autocorrelation(Hz)	HPS (Hz)
61	277.18	279.04	279.4
59	246.94	249.08	249.3
64	329.63	331.8	331.9
62	293.66	296.4	297
61	277.18	279.8	280.3
64	329.63	331.9	331.8
69	440	443.4	444
71	493.88	499.9	501
64	329.63	331.2	331.5
Av. Auto Error (Hz,%)		2.754	0.80
Av. HPS Error (Hz, %)		3.163	0.92