A Real-Time Sines + Noise Timbre Transfer Model with Stretched Partials

Yutian Qin

yq2120@nyu.edu

Tandon School of Engineering, New York University

December 2021

Abstract

In this paper, a real-time Sines + Noise Timber Transfer Model with Stretched Partials is established to transfer the timbre from the piano into the timbre of a traditional Chinese musical instruments called *Liu-qin*. This model has an emphasis on the transfer of stretched partials. During the sound synthesis progress, amplitude of harmonic series are determined from the spectral envelop of *Liu-qin* and the frequencies are the sum of the theoretical frequency and the amount of stretching.

1 Introduction

The composition of harmonic series have a large influence on human's perception on musical timbres. While the theoretical calculation values of harmonics are all multiple integers of the fundamental frequencies, the stiffness of real strings lead to stretched partials in the frequencies. Research by Moore[5] has proved that the stretched partials have a non-negligible effect on human perception to musical timbres.

Previous research on timbre transfer tend to ignore the role of stretched partials. In this paper, a real-time sines+noise timbre transfer model with emphasis on stretched partials is introduced.

In this model, the object that the timbre transferred into is a traditional Chinese musical instrument called *Liu-qin*, whose image is shown as Figure 1 The timbre characteristic of *Liu-qin* is bright and warm, with relatively fewer harmonics in high frequencies.

2 Dataset

The dataset used for establishing the model are nineteen recordings of *Liu-qin* notes and nineteen recordings of the piano notes. The sound files of *Liu-qin* is recorded with sample rate 44.1 kHz. Nineteen notes in C major with abundant harmonics from four strings are chosen as the objective of

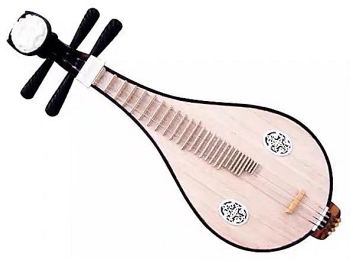

Figure 1: Image of A Liu-qin

Note	Standard Frequency (Hz)	MIDI Pitch	Note	Standard Frequency (Hz)	MIDI Pitch
G3	196.00	55	C5	523.25	72
A3	220.00	57	D5	587.33	74
B3	246.94	59	E5	659.26	76
C4	261.63	60	F5	698.46	77
D4	293.67	62	G5	783.99	79
E4	329.63	64	A5	880.00	81
F4	349.23	65	B5	987.77	83
G4	392.00	67	C6	1046.5	84
A4	440.00	69	D6	1174.7	86
B4	493.88	71			

Table 1: Notes and Frequencies Used for Analysis

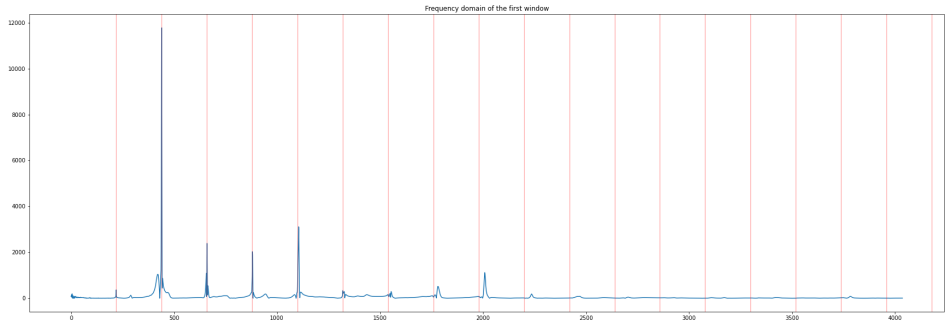

Figure 2: Stretched Partials of Liu-qin Sound

analysis. Accordingly, nineteen relative notes were recorded from a player piano with MIDI file. Each of the notes has a MIDI velocity of 127, which has the largest intensity and a MIDI duration of 1 second. The nineteen notes are shown in the following table:

3 Timbre Representation

3.1 Stretched Partials

In my research, stretched partials was first found when calculating the amplitudes of the harmonics. With long window length and high precision in frequency domain, the frequencies of the peaks was found not the multiple integers of the fundamental frequencies. After measurement, the stretched partials are different in amount between different musical instruments as shown in the following figures.

Figure2 shows the spectrum of *Liu-qin*, Figure3 shows the spectrum of the piano and Figure4 shows the spectrum of human voice.

To get the stretched partials, the fundamental frequencies and the frequencies of 17 levels of harmonics were calculated from the recordings of *Liu-qin* sounds. During the calculation, the window-length of FFT was selected as 32768 for a frequency accuracy of 1.1 Hz.

Under the selected window-length, the Nyquist frequency is 16382Hz, which determined the levels of harmonics. 17 is the highest level can be gotten for the note D6 under the Nyquist frequency. It can be seen from Figure44 in the Appendix.

Thus, a 19×18 two-dimensional list containing the stretching in Hz was calculated and stored in .npy for later synthesis.

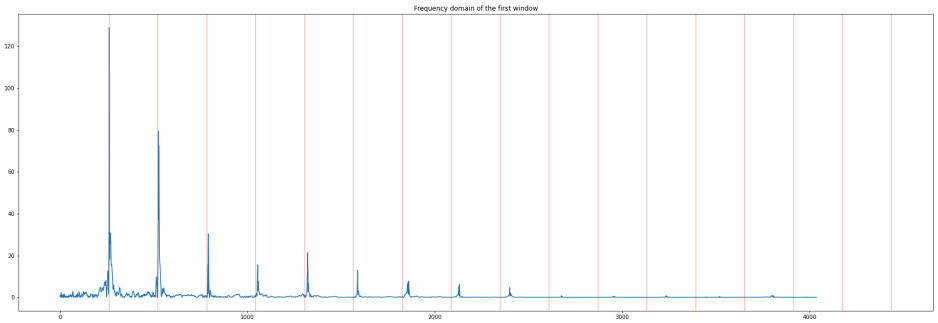

Figure 3: Stretched Partials of Piano Sound

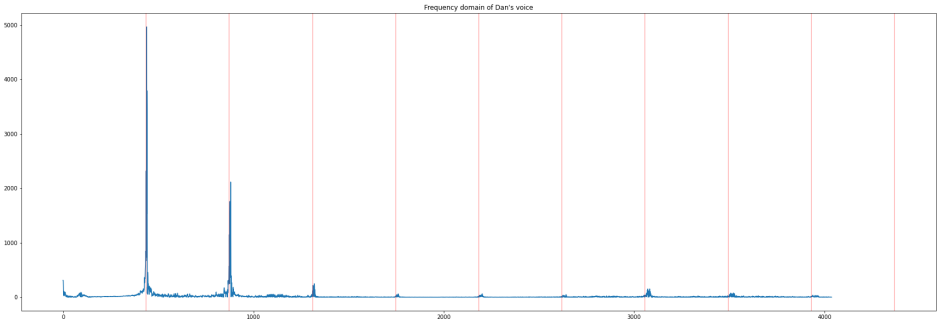

Figure 4: Stretched Partials of Human Voice

3.2 Spectral Envelope

There are several timbre models to describe the musical timbres. Jensen [3] developed a framework of sinusoidal modeling of musical timbre, Burred[1] introduced a dynamic spectral envelop modeling for timbre of musical instrument sounds. Spectral envelop was introduced by Wessel[9] as an effective modeling methof to describe musical timbre. In this model, spectral envelop is used to represent the timbre of *Liu-qin*. The spectral envelope calculated from *Liu-qin* recordings are as shown in Figure5.

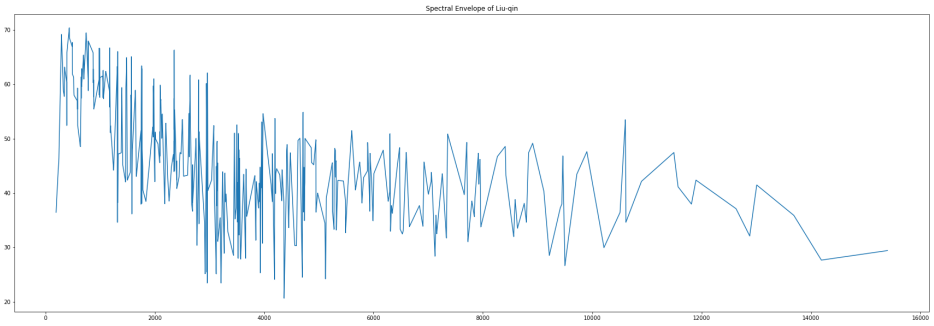

Figure 5: Spectral Envelope of Liu-qin

During the calculation, the amplitude of each fundamental frequencies and 17 levels harmonics are calculated and recorded. By sorting and plotting, Figure5 was obtained as the envelope. As the frequency intervals between different data points have different length, the linear interpolation was used to fill in the intervals and connect data points.

Also, because of the characteristics of *Liu-qin*, small peaks of amplitude appears on both sides of the majority amplitude peaks at the harmonics, which is shown in Figure6. As mapping makes it hard for human perception to distinguish the frequencies, the amplitudes at these minor peaks still makes a difference. To simplify the calculation, sum of square of the peak amplitude was calculated and then take the square root according to Parseval's Theorem. The range of amplitude peaks selection can be illustrated form Figure6. Spectrum is in blue in this figure. Red lines represents the lower bond of the selection range, which is the multiple integers of the fundamental frequencies. And the green lines are the upper bound of the selection range. The length of the range increases linearly with level considering the influence of stretched partials.

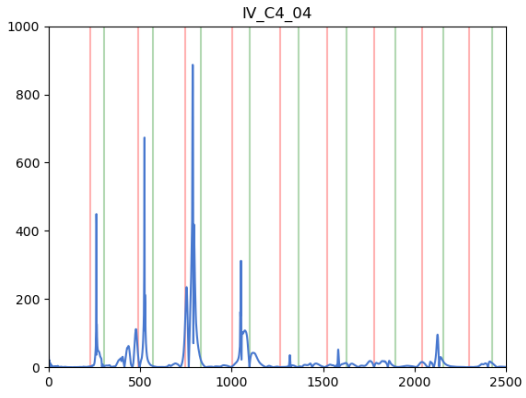

Figure 6: Amplitude Calculation Range for Harmonics

4 Real-Time Transfer Model

The transfer model designed in this project is based on the Sines+Transients+Noise Audio Representation [4]. Considering that the transfer object piano does not have an apparent attack as plunged string instruments, the Transients from was removed for simplification. Thus, the model I use is basically a Sines + Noise transfer model.

4.1 Sinusoidal Modeling

The algorithm used a Fast Fourier Transform with a window length of 4096 to detect the fundamental frequency of the input signal. Considering the limitation of FFT, the frequency precision is low at approximately 11 Hz. Thus, it is hard to get the accurate frequency. The detected frequency is obtained by comparing to the standard fundamental frequencies of the nineteen notes and the frequency with the smallest difference is the detected fundamental frequency.

4.2 Noise

The noise from the Sines+Transients+Noise model is gotten from a bark-band filtered noise. In this model, in order to get the noise form the input signal, a simple four-level Daubechies wavelet transform[7] is used to extract the noise from the input and add back to the output when synthesizing. The noise extraction can keep the noise the same as the environment in real-time and avoid doing Fourier transform to high-frequency part of the input.

The decomposition of signal is given by:

$$c(n) = h_0 x(2n) + h_1 x(2n+1) + h_2 x(2n+2) + h_3 x(2n+3) \tag{1}$$

$$d(n) = h_3 x(2n) - h_2 x(2n+1) + h_1 x(2n+2) - h_0 x(2n+3) \tag{2}$$

where x(n) is the input signal, c(n) and d(n) are signals intermediate obtained by transformation. After four levels of decomposition, four levels d(n) and one c(n) can be obtained. On every d(n) and c(n), a threshold was added to extract the noise. Signals under the threshold were maintained and signals exceed the threshold were filtered out. After filtering, the reversion of decomposition was operated by:

$$y(2n) = h_0 c(n) + h_2 c(n-1) + h_3 d(n) + h_1 d(n-1) \tag{3}$$

$$y(2n+1) = h_1 c(n) + h_3 c(n-1) - h_2 d(n) - h_0 d(n-1) \tag{4}$$

After 4 levels of inverse decomposition, the extracted noise can be obtained by Daubechies wavelet transforms.

4.3 Synthesis

During the synthesis, the sine waves of the seventeen harmonics and the fundamental frequencies as well as the noise extracted from the input is added together in time domain and compose the output. The phases were all defaulted to 0 as human ears are not sensitive to phase.

5 Discussion

The stretched partials calculated from the recordings are not larger than the multiple integers of the fundamental frequencies. A few in the data are equal or even smaller than the multiple integers of the fundamental frequencies. It may be attributed to the influence of the instrument and the recorder. More data should be collected to find the reason of this phenomenon.

The frequency detection in real-time has the limitation to get the accuracy frequency of the input piano sounds. So the assumption that the piano's tone is accurate is made. The deviation cannot be detected by the current algorithm.

6 Future Work

6.1 Stretched Partials

In current version, the stretched partials are calculated from recordings and saved in .npy files. As a fixed physical feature, the stretching of the partials can be calculated in the following ways[2]:

$$f_n = nf_0(1 + Bn^2)^{0.5} \tag{5}$$

$$B = \frac{pi^3 Qd^4}{64l^2 T} \tag{6}$$

where n is the partial numbers, A is Young's modulus, d is the diameter of the string, l is the length of the string and T is the tension of the string, and $f_0$is the fundamental frequency of the string without stiffness.

As shown from the equations, for a certain string, once the locations of the frets are fixed, the stretched partials would remain a constant. In the current model, the stretching was calculating from the recordings, which led into certain errors. In the future work, the precision of the stretching can be improved by the equations and more material.

6.2 Fundamental Frequency Detection

The current fundamental frequency detection relies on the Fast Fourier Transform. In order to keep the delay in real-time application small, the frequency accuracy is thus sacrificed, in this case is approximately 11. This means that it is almost impossible to recognize the accurate fundamental frequency of the input. In order to detect the accurate fundamental frequency, a more efficient detection method is needed. Real-time wavelet transform such as wavelet transform with tunable Q-factor[8] is going to be used in the next step as the replacement of Fast Fourier Transform for relatively precise temporal and spectral features of each frame.

6.3 Spectral Envelope

The spectral envelop is currently using linear functions to connect adjacent points, which means the current spectral envelop is not smooth enough. In the future work, a better curve fitting method is needed for smoother spectral envelope. Discrete Cepstrum[6] will be used to curve the spectral envelop in substitute of linear interpolation.

6.4 Amplitude Features of Different Musical Instruments

Besides the spectral envelope, amplitude envelop deserves optimization. The current transfer system didn't take the ADSR (Attack, Decay, Sustain, Release) of different musical instruments into consideration. Amplitude ADSR is also a significant feature to be considered in musical timbre. The attack of plucked string instrument is more apparent than that of piano and the decay rate of the vibration of different strings varies. Transients will be added into this model to simulate the noisy attack part for some plucked string instruments. Also, the decay will be collected and added.

6.5 Hardware Part

A pick-up designed according to the size of *Liu-qin* has been produced in the previous work which is shown in Figure 7.

A hardware system with single chip microcomputer will be designed in the future work for a new interface of *Liu-qin* as an electrification traditional Chinese musical instrument.

7 Conclusion

In conclusion, a real-time sine + Noise timbre transfer model which takes stretched partials of string instruments is established to transfer the timbre or the piano into the timbre of *Liu-qin* in real time. The timbre transfer is based on spectral envelope of *Liu-qin* and the fixed stretched frequencies collected.

Figure 7: Pickup Designed for Liu-qin

References

[1] J. J. Burred, A. Robel, and T. Sikora. Dynamic spectral envelope modeling for timbre analysis of musical instrument sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3):663–674, 2010.

[2] H. Järveläinen, V. Välimäki, and M. Karjalainen. Audibility of inharmonicity in string instrument sounds, and implications to digital sound synthesis. In *ICMC*, 1999.

[3] K. Jensen. The timbre model. *Journal of the Acoustical Society of America*, 112(5):2238–2238, 2002.

[4] S. N. Levine and J. O. Smith III. A sines+ transients+ noise audio representation for data compression and time/pitch scale modifications. In *Audio Engineering Society Convention 105*. Audio Engineering Society, 1998.

[5] B. C. Moore, R. W. Peters, and B. R. Glasberg. Thresholds for the detection of inharmonicity in complex tones. *The Journal of the Acoustical Society of America*, 77(5):1861–1867, 1985.

[6] D. Schwarz, X. Rodet, et al. Spectral envelope estimation and representation for sound analysis-synthesis. In *ICMC*. Citeseer, 1999.

[7] I. W. Selesnick. Wavelet transforms-a quick study. *Physics Today magazine*, 2007.

[8] I. W. Selesnick. Wavelet transform with tunable q-factor. *IEEE transactions on signal processing*, 59(8):3560–3575, 2011.

[9] D. L. Wessel. Timbre space as a musical control structure. *Computer music journal*, pages 45–52, 1979.

A Harmonic Series

The spectrum of both the piano and *Liu-qin* are included in the Appendix. In each figure, the blue line represents the spectrum and the red 'x' marker represents the peak value, which is the position of fundamental frequencies and the harmonic series.

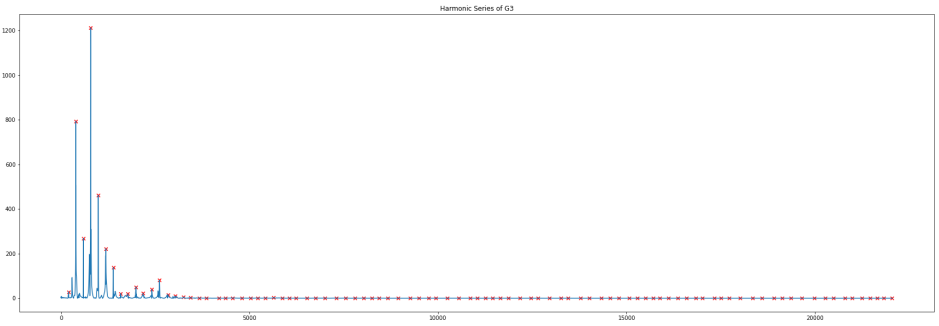

Figure 8: Harmonic Series of G3 on Liu-qin

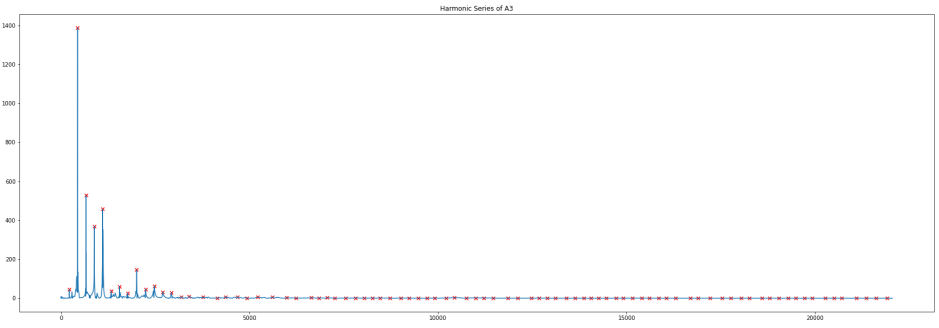

Figure 9: Harmonic Series of A3 on Liu-qin

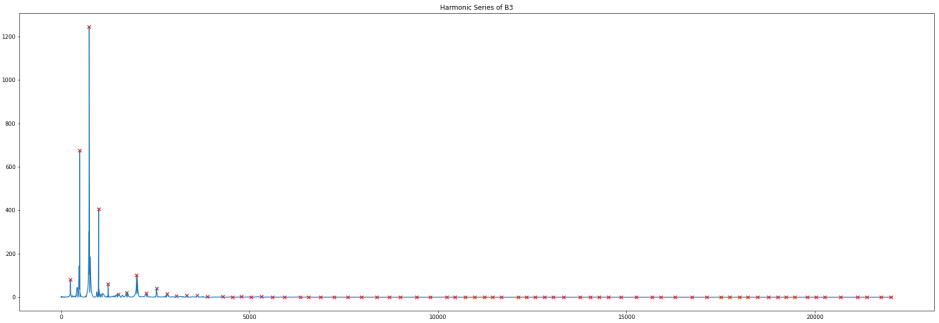

Figure 10: Harmonic Series of B3 on Liu-qin

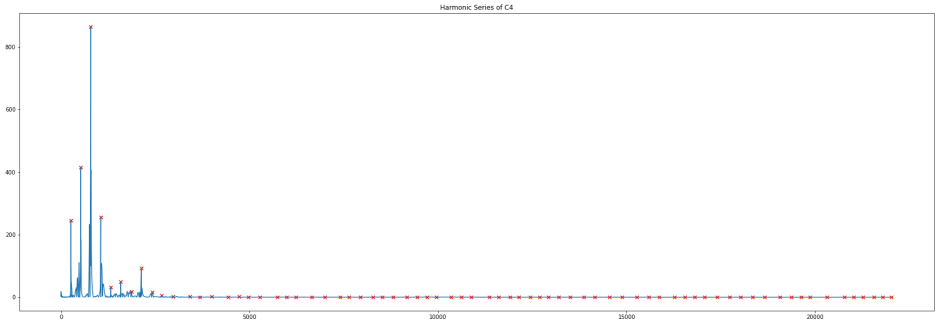

Figure 11: Harmonic Series of C4 on Liu-qin

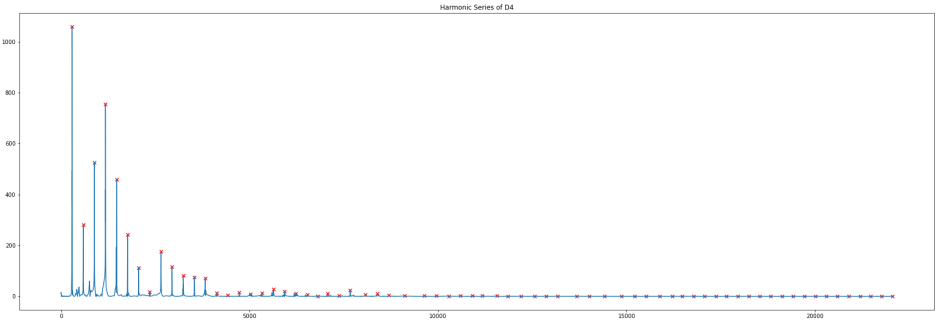

Figure 12: Harmonic Series of D4 on Liu-qin

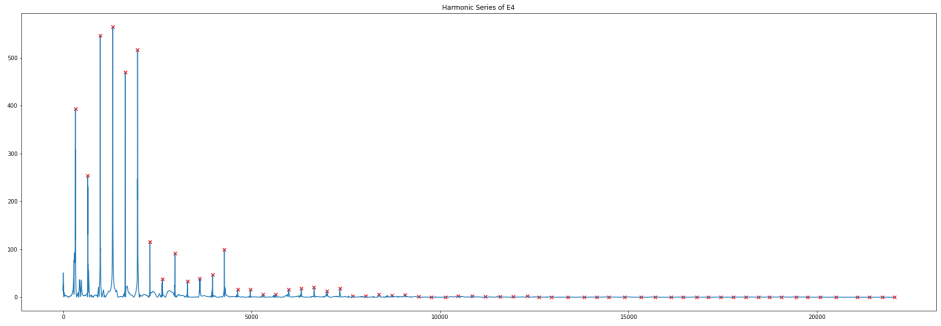

Figure 13: Harmonic Series of E4 on Liu-qin

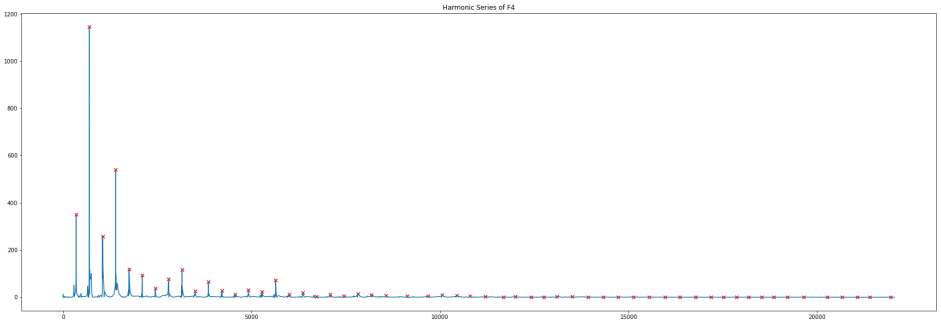

Figure 14: Harmonic Series of F4 on Liu-qin

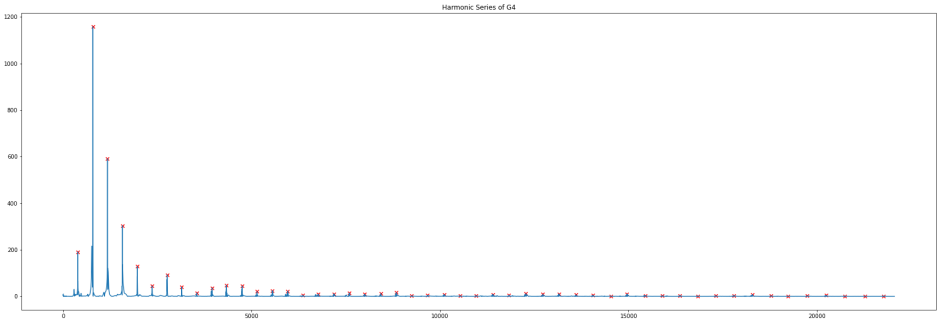

Figure 15: Harmonic Series of G4 on Liu-qin

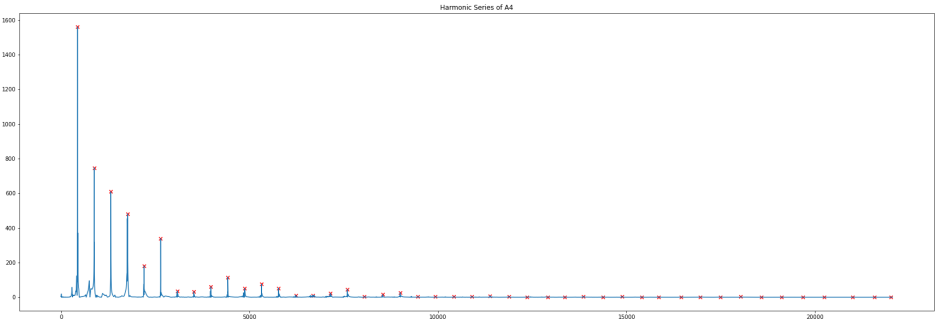

Figure 16: Harmonic Series of A4 on Liu-qin

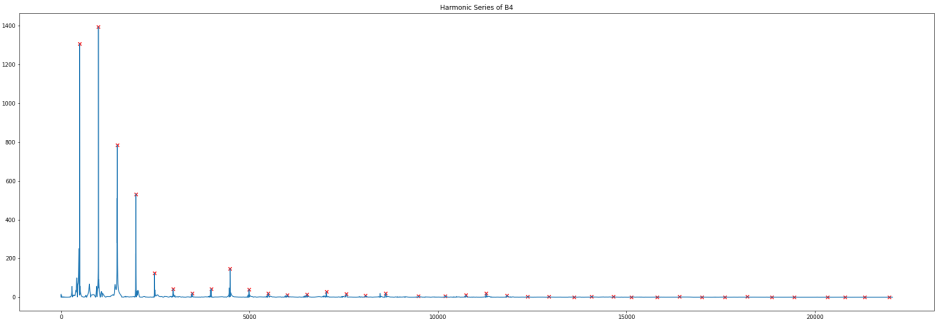

Figure 17: Harmonic Series of B4 on Liu-qin

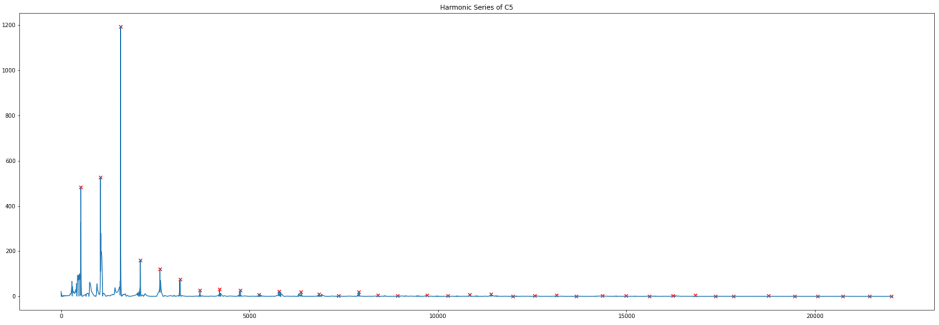

Figure 18: Harmonic Series of C5 on Liu-qin

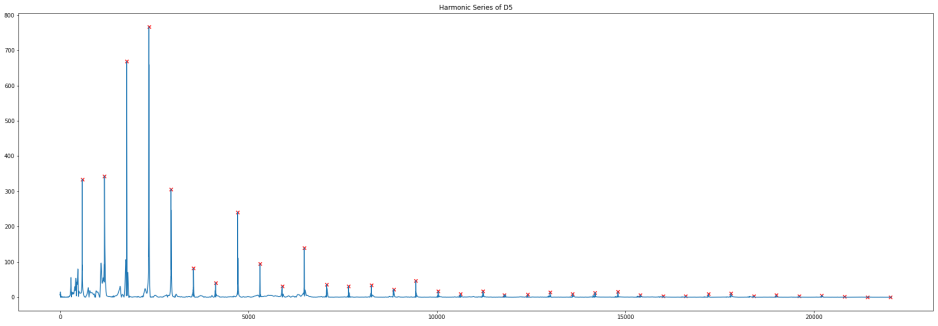

Figure 19: Harmonic Series of D5 on Liu-qin

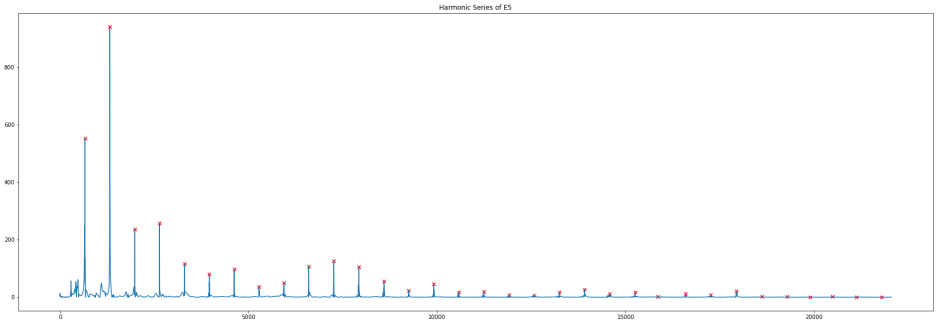

Figure 20: Harmonic Series of E5 on Liu-qin

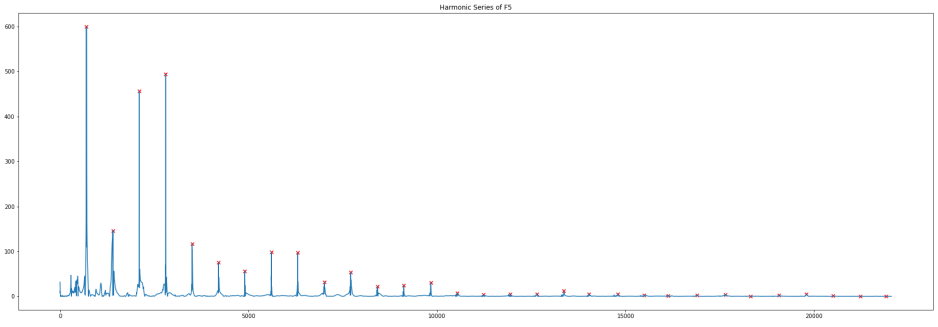

Figure 21: Harmonic Series of F5 on Liu-qin

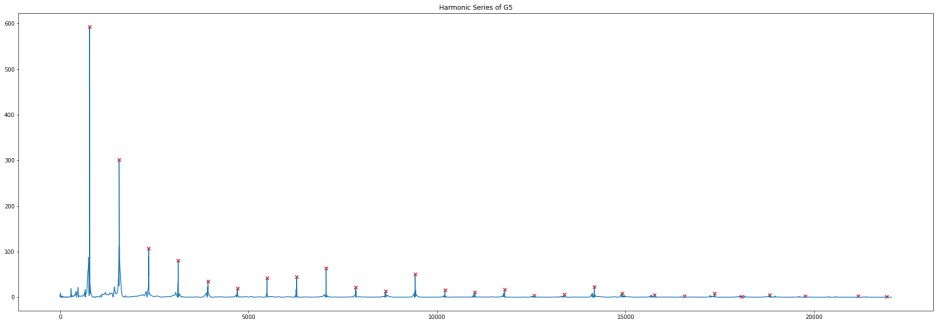

Figure 22: Harmonic Series of G5 on Liu-qin

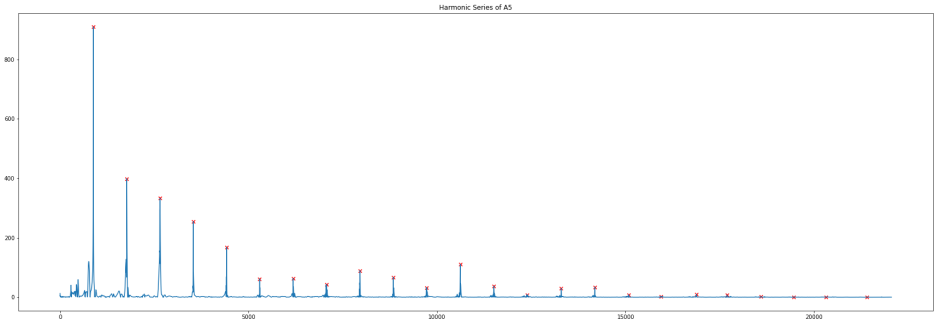

Figure 23: Harmonic Series of A5 on Liu-qin

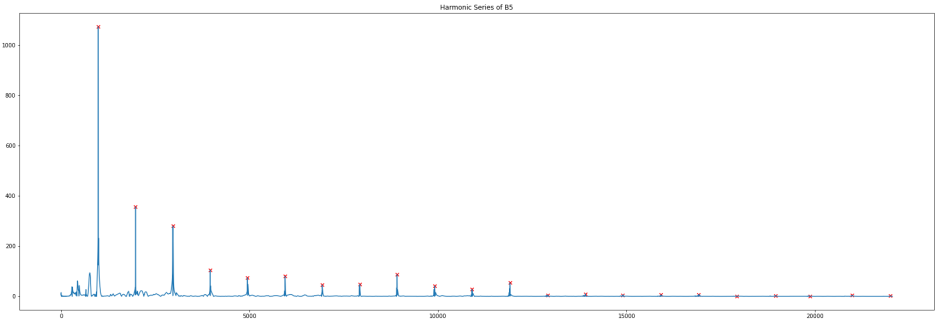

Figure 24: Harmonic Series of B5 on Liu-qin

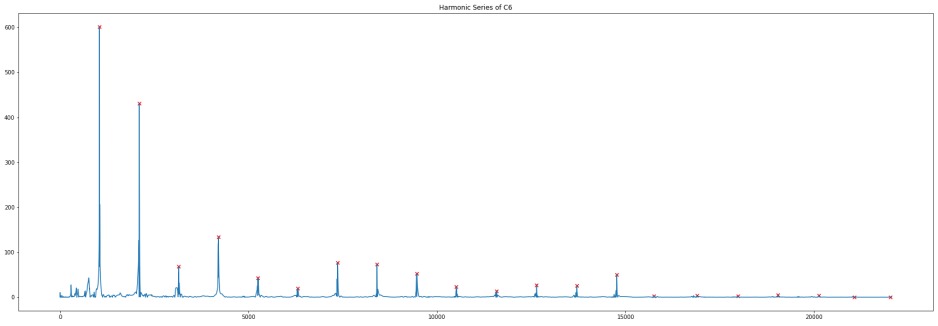

Figure 25: Harmonic Series of C6 on Liu-qin

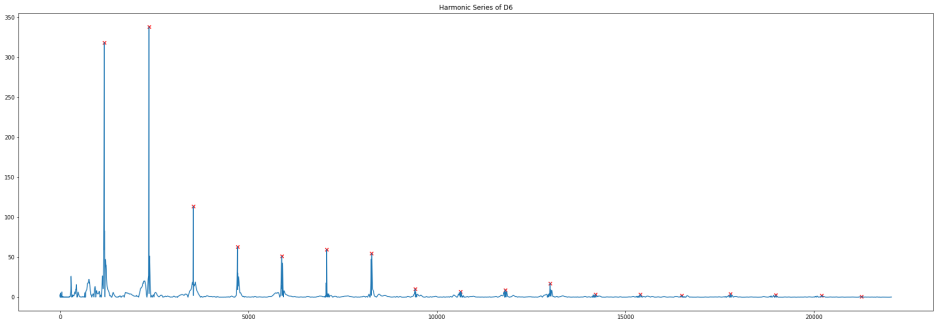

Figure 26: Harmonic Series of D6 on Liu-qin

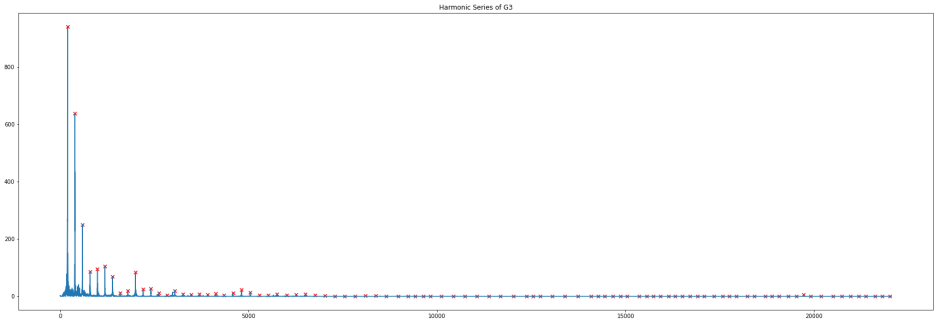

Figure 27: Harmonic Series of G3 on Piano

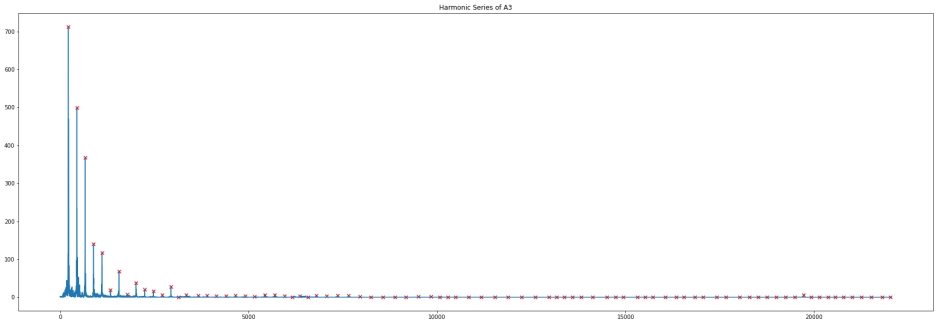

Figure 28: Harmonic Series of A3 on Piano

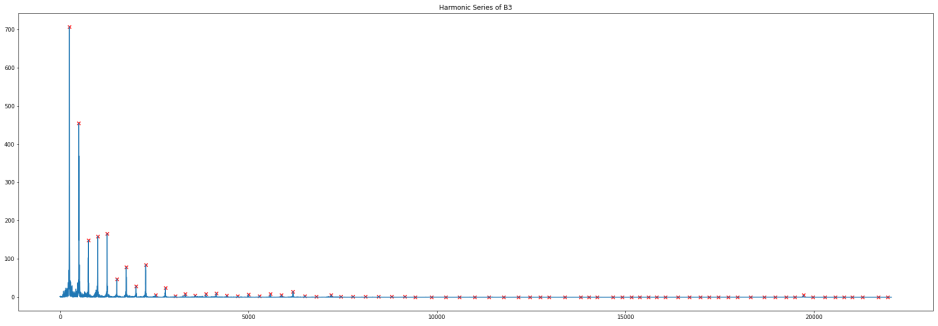

Figure 29: Harmonic Series of B3 on Piano

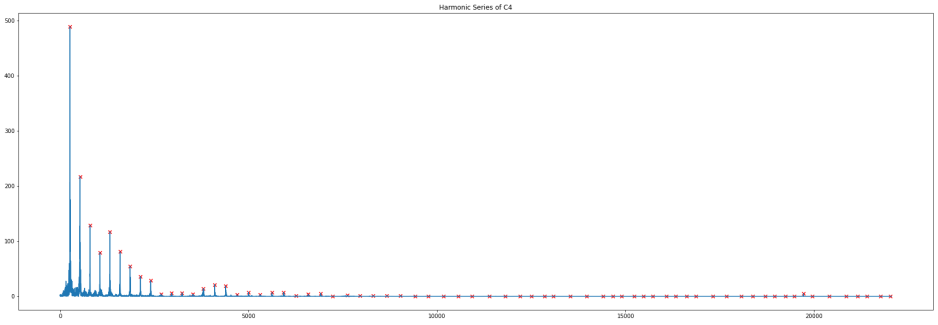

Figure 30: Harmonic Series of C4 on Piano

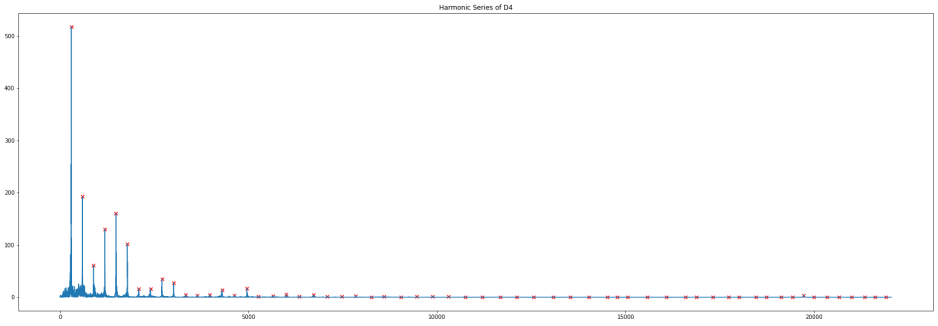

Figure 31: Harmonic Series of D4 on Piano

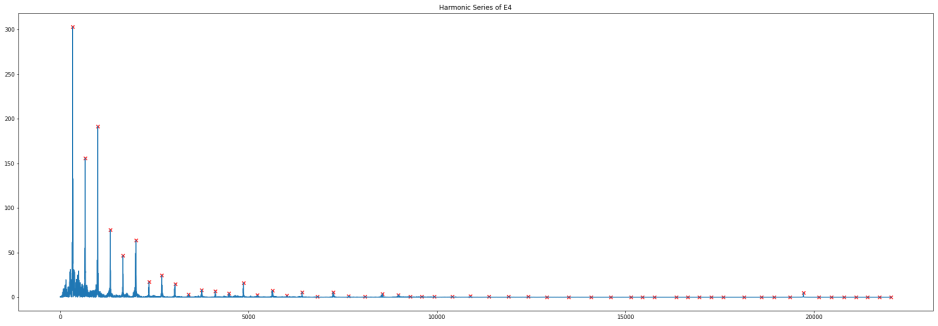

Figure 32: Harmonic Series of E4 on Piano

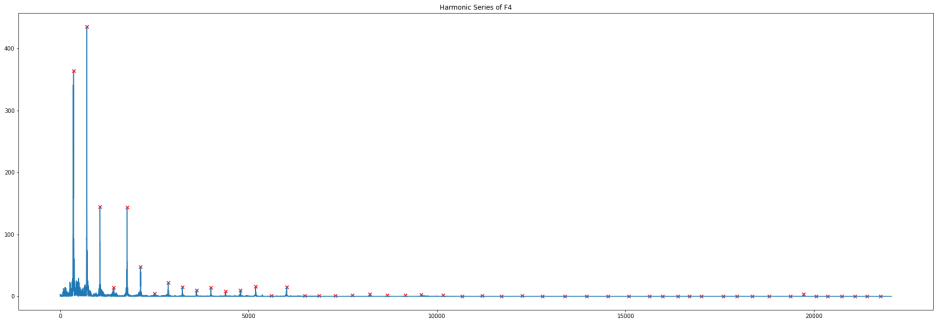

Figure 33: Harmonic Series of F4 on Piano

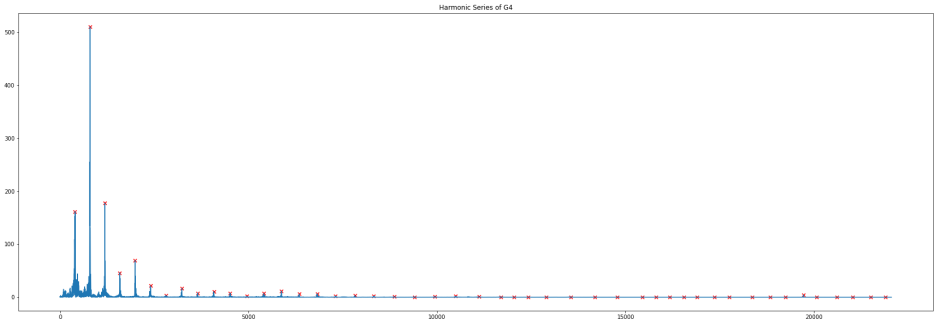

Figure 34: Harmonic Series of G4 on Piano

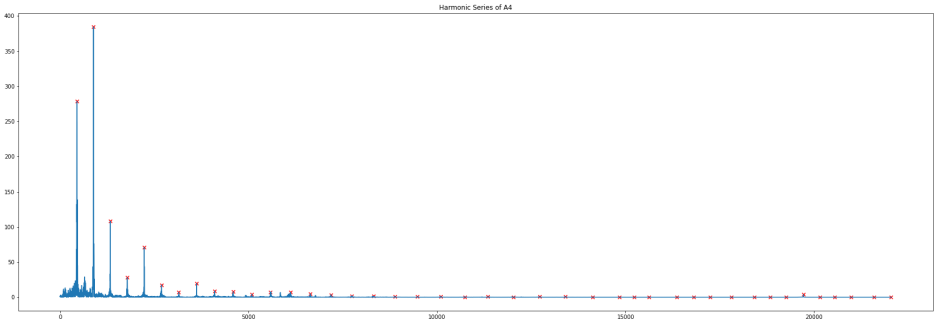

Figure 35: Harmonic Series of A4 on Piano

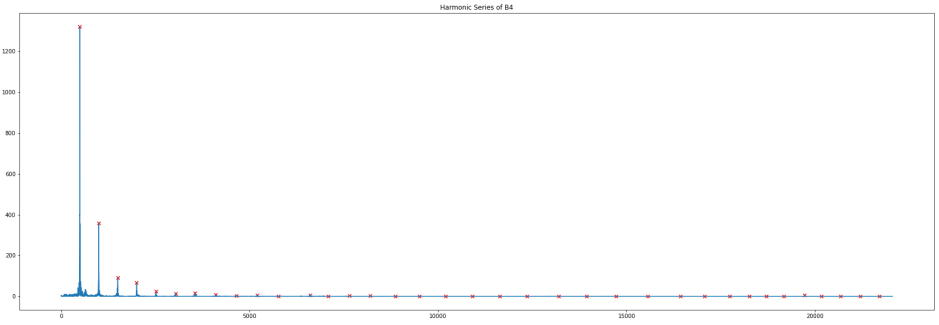

Figure 36: Harmonic Series of B4 on Piano

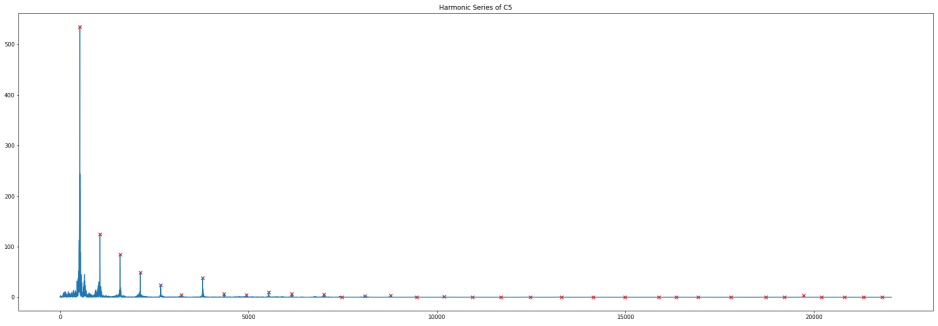

Figure 37: Harmonic Series of C5 on Piano

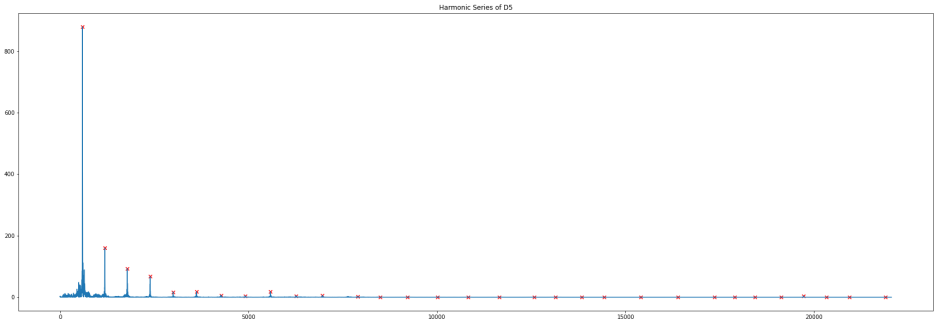

Figure 38: Harmonic Series of D5 on Piano

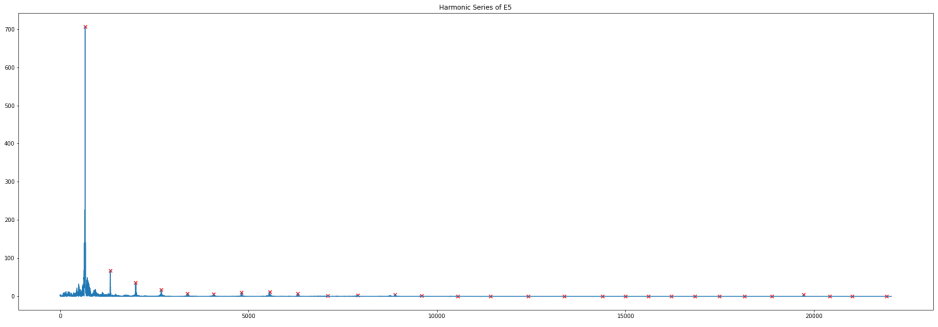

Figure 39: Harmonic Series of E5 on Piano

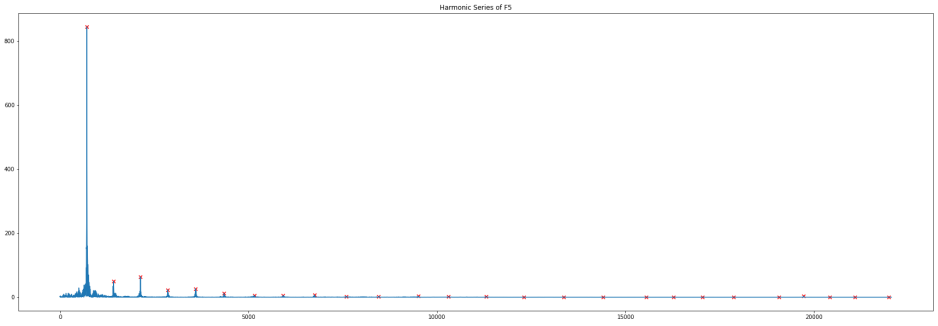

Figure 40: Harmonic Series of F5 on Piano

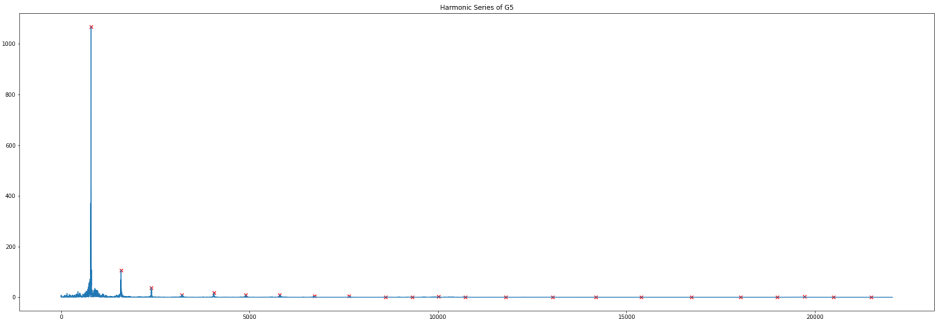

Figure 41: Harmonic Series of G5 on Piano

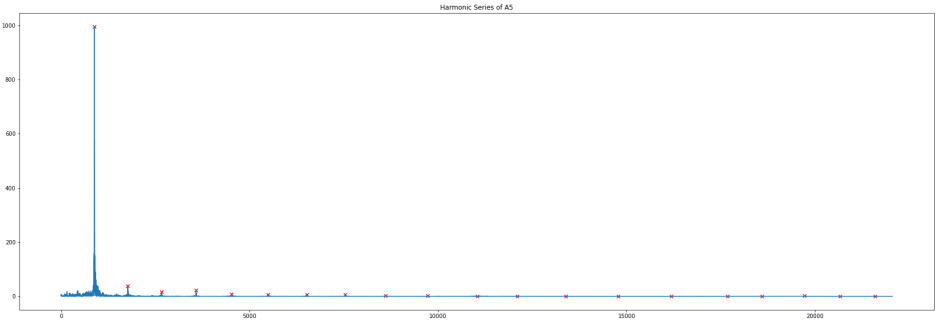

Figure 42: Harmonic Series of A5 on Piano

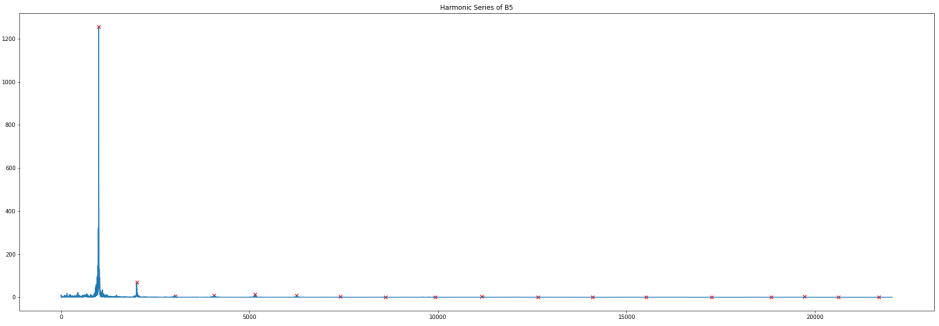

Figure 43: Harmonic Series of B5 on Piano

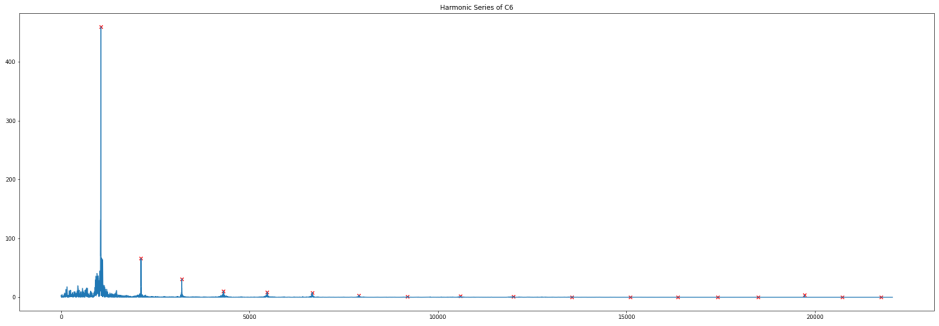

Figure 44: Harmonic Series of C6 on Piano

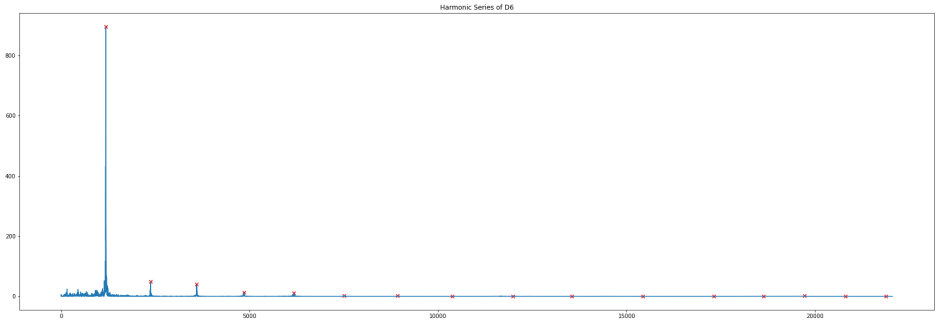

Figure 45: Harmonic Series of D6 on Piano