Time-Frequency Analysis and Wavelet Transform

Term Paper (書面報告)

R06942065 電信二 黃宇平

Topic:

When InfluxDB Meets Time-Frequency Analysis

Github Link:

https://github.com/tom6311tom6311/PyInfluxTFA

Abstract

1. Introduction

InfluxDB is the state-of-the-art database for storing time series data. This time series database provides support for concurrent read/write and real-time analytics of massive amounts of time-stamped information with SQL-like querying language. However, the analytics of InfluxDB is currently constrained in time-domain analysis for real-time purposes. With knowledge gained from this course, I propose a practical way to integrate time-frequency analysis with InfluxDB, which expands horizons of data analytics on time series databases and enables users to have a more clear view on both time and frequency aspects of real-time data. To get real-time result of Time-Frequency Analysis, there is a trade-off between segmentation period and precision of the spectrum. I also investigate this trade-off by experiments.

2. Related Work

2.1 InfluxDB

InfluxDB is an open-sourced database developed by InfluxData that is optimized for fast, high-availability storage and retrieval of time series data. With the rise of Internet of Things and Big Data, real-time monitoring and analytics of sensor data becomes a top concern in related field. Due to the nature of handling time-stamped data, InfluxDB stands out from all other popular databases and dominates the fields of time series data analytics.

2.2 Time-Frequency Analysis

Time-Frequency Analysis is a set of transformation techniques that analyzes given input signal in both time and frequency domain simultaneously. Such techniques includes: Short-time Fourier transform (STFT), Wigner distribution function (WDF), Wavelet transform, Time-Variant Basis Expansion, Hilbert-Huang Transform(HHT), ...etc. With the great improvement of hardware and computational power nowadays, time-frequency analysis is now considered as a practical way of analyzing data, which provides a more sophisticated view than ever.

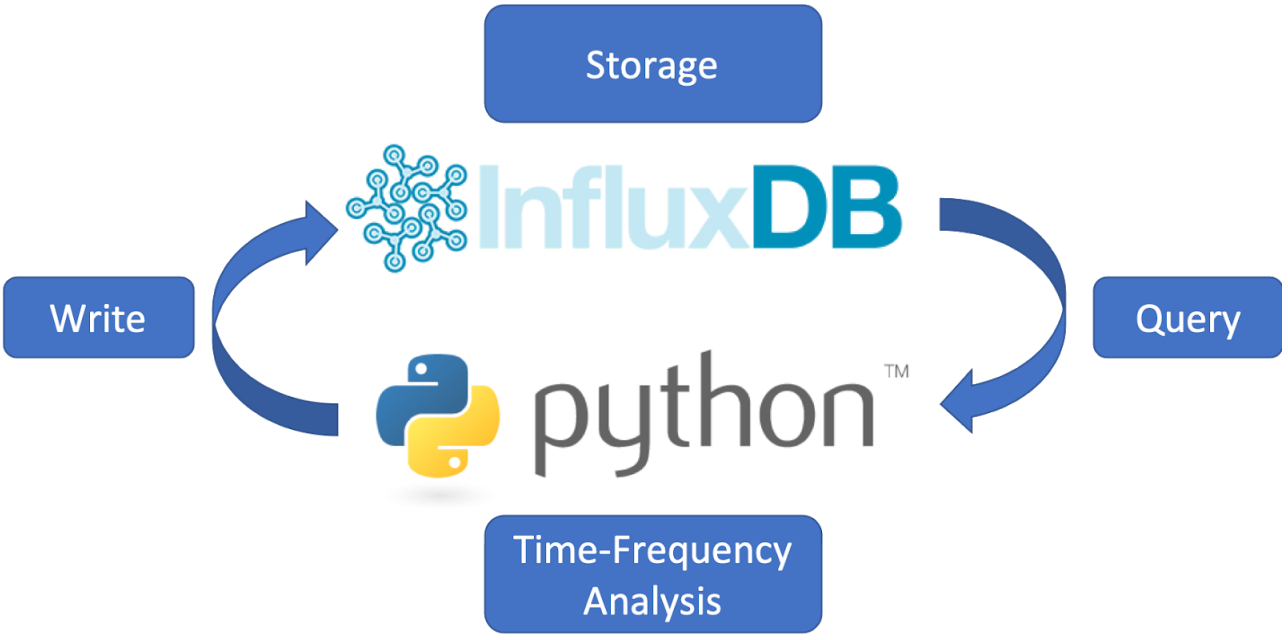
3. Problem Statement

In spite of the high performance of storing and monitoring time-series data in real-time, InfluxDB is very limited in analyzing data from perspective of frequency domain. Frankly speaking, this is due to the fact that Fourier-Transform-based analysis requires to compute integral of the signal from start to end, rather than a segment of time. This nature stops FT from being applied to real-time analysis of time-series data. However, many characteristics of signal can only be observed from frequency domain. Besides, many digital signal processing applications are related to the spectrum or the bandwidth of a signal. Thus, there is an urgent need of integration of InfluxDB with techniques for analyzing data in both time and frequency domains simultaneously in nearly real-time.

4. Solution

To match the need of performing analysis in both time and frequency domains on InfluxDB simultaneously, I propose a way to integrate Time-Frequency Analysis with InfluxDB.

4.1 System Model



My implementation is composed of 2 parts:

(a) An InfluxDB instance:

This is the database from which time series data are stored and queried.

(b) A Python Client Module

This client module queries time-series data from the InfluxDB instance and implements Time-Frequency Analysis such as Short-time Fourier Transform and Hilbert-Huang Transform on the data, then visualize the processed data.

5. Experiments

5.1 STFT Spectrums with Different Period and Sampling Rates

In this experiment, I use a piano audio signal as input and draw STFT spectrums as output. There are 2 variables in this experiment:

(a) Segmentation Period:
The time duration each STFT performs on, which is also the period from one InfluxDB query to another. This determines the latency of real-time spectrum results.

(b) Frames/Second (fs)
The sampling rate of signal used to plot STFT spectrums. This is closely related to the computational complexity and spectrum quality.

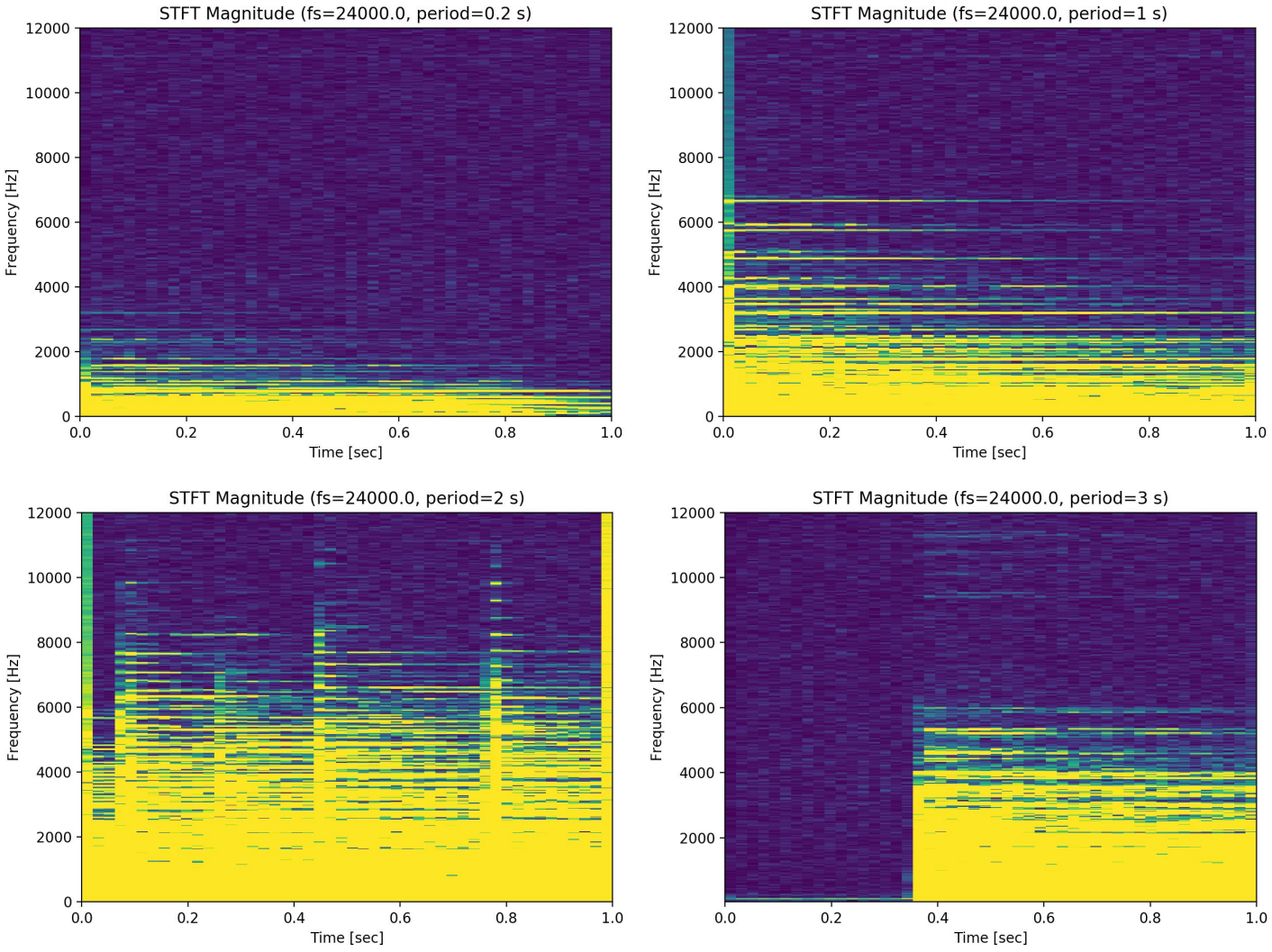
I perform this experiment with process below:

(a) Store the audio data to InfluxDB
(b) Query a chunk of data from InfluxDB, filtered with segmentation period and fs.
(c) Compute STFT of the chunk.
(d) Repeat (b)~(c) until reaching end of the data.
(e) Concatenate all STFT results and plot it out.

(Note that in real-time applications, we should plot the result directly after each STFT, rather than collect all results and plot it out in the end.)

The result of experiment 5.1 is below:

(a) With different Segmentation Period:



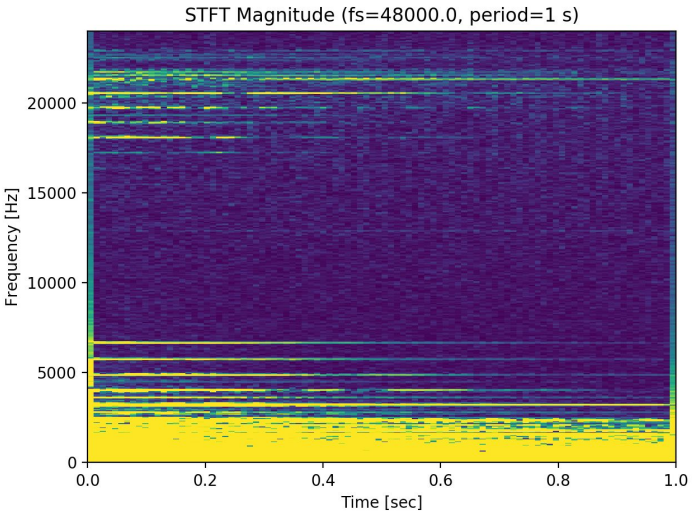
Observation:

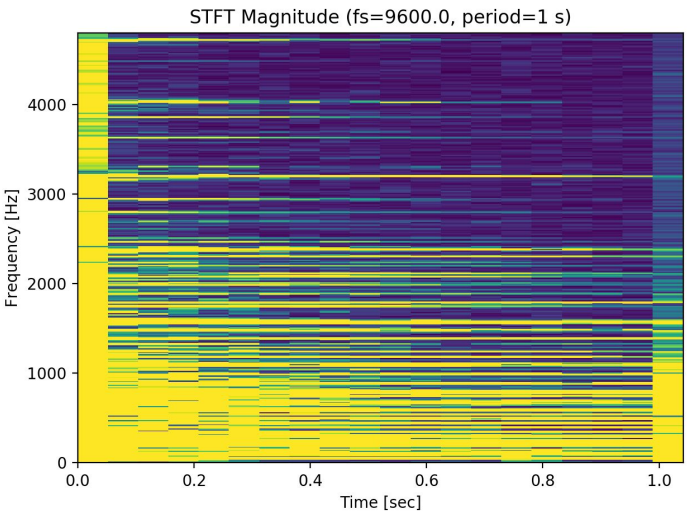
(i) Shorter segmentation period may cause frequency shift and force frequency components to be squeezed together.

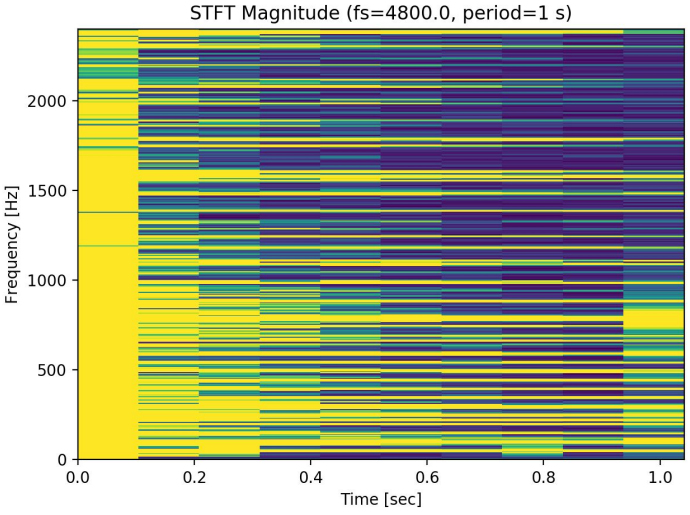
(ii) We can observed some repeated patterns in time domain, these are caused by segmentation.

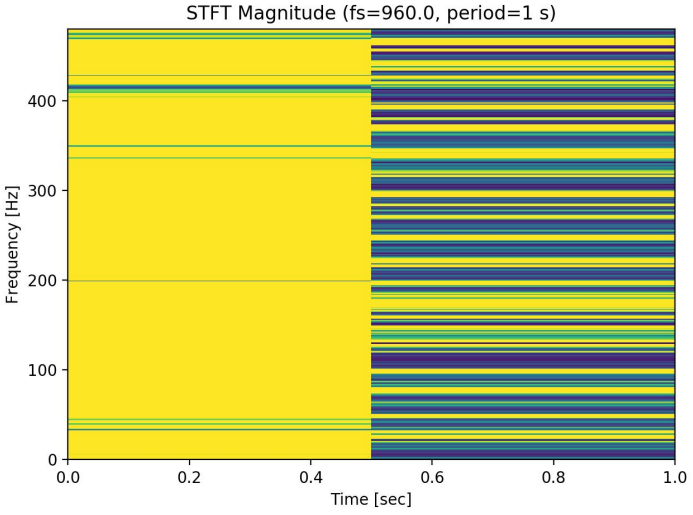
(iii) STFT spectrum is affected by segmentation significantly. Thus it does not fit to real-time applications considering precision.

(b) With Different Frames/Second (fs)









Observation:

(i) Sampling rate signifcantly affects the time resolution of STFT spectrum.

(ii) However, relative position of frequency components are still kept in spectrum of lower sampling rates. In other words, we can reduce the complexity while preserving some rough frequency characteristics.

5.2 HHT Results Spectrums with Different Period and Sampling Rates

In this experiment, I use a chirp-like signal as input and draw envelop and instantaneous frequency of HHT as output. There are 2 variables in this experiment:

(a) Segmentation Period:
The time duration each HHT performs on, which is also the period from one InfluxDB query to another. This determines the latency of real-time spectrum results.

(b) Frames/Second (fs)
The sampling rate of signal used to plot HHT spectrums. This is closely related to the computational complexity and spectrum quality.

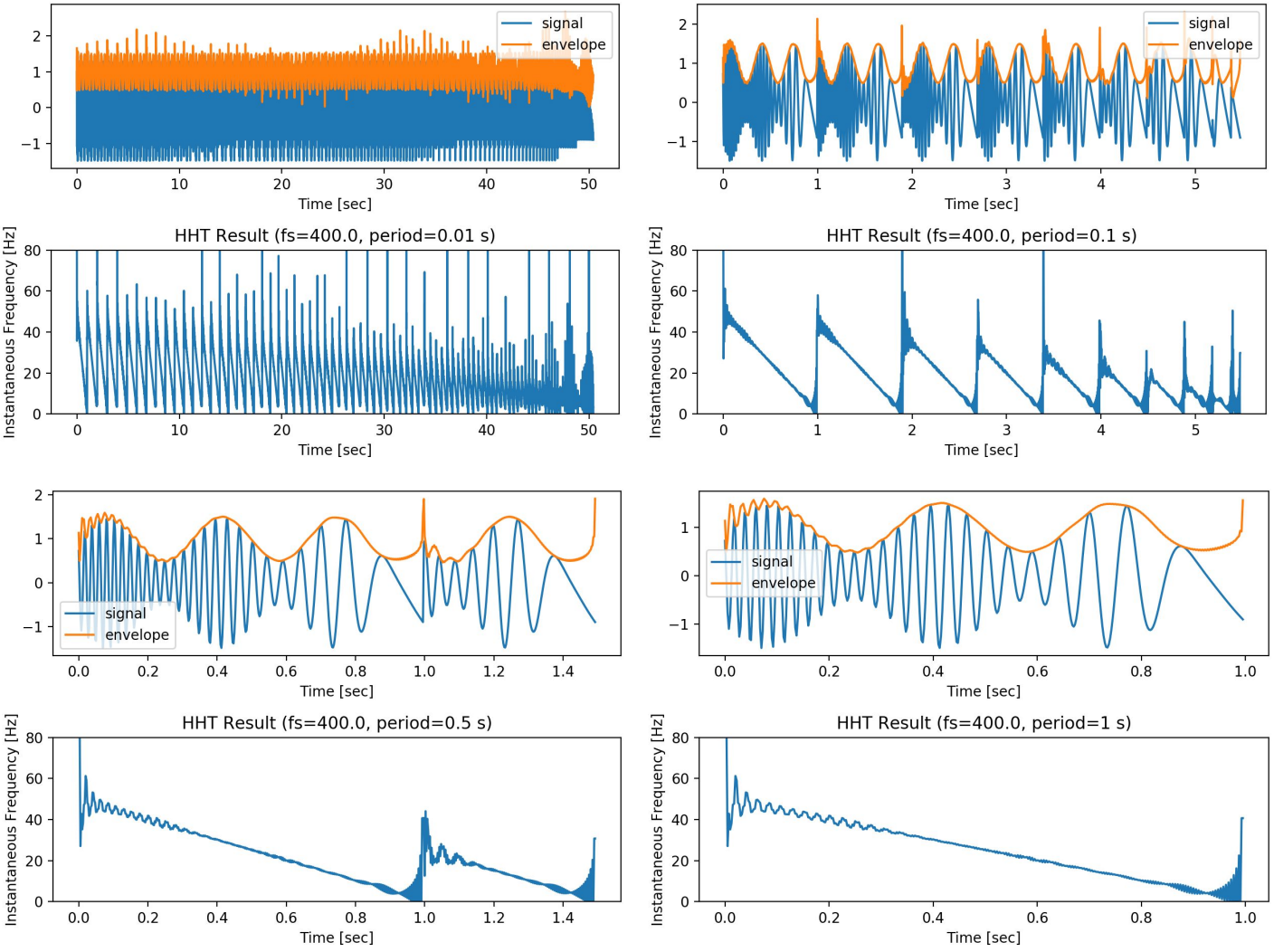
I perform this experiment with process below:

(a) Store the data to InfluxDB
(b) Query a chunk of data from InfluxDB, filtered with segmentation period and fs.
(c) Compute HHT of the chunk.
(d) Repeat (b)~(c) until reaching end of the data.
(e) Concatenate all HHT results and plot it out.

(Note that in real-time applications, we should plot the result directly after each HHT, rather than collect all results and plot it out in the end.)

The result of experiment 5.2 is below:

(a) With different Segmentation Period:

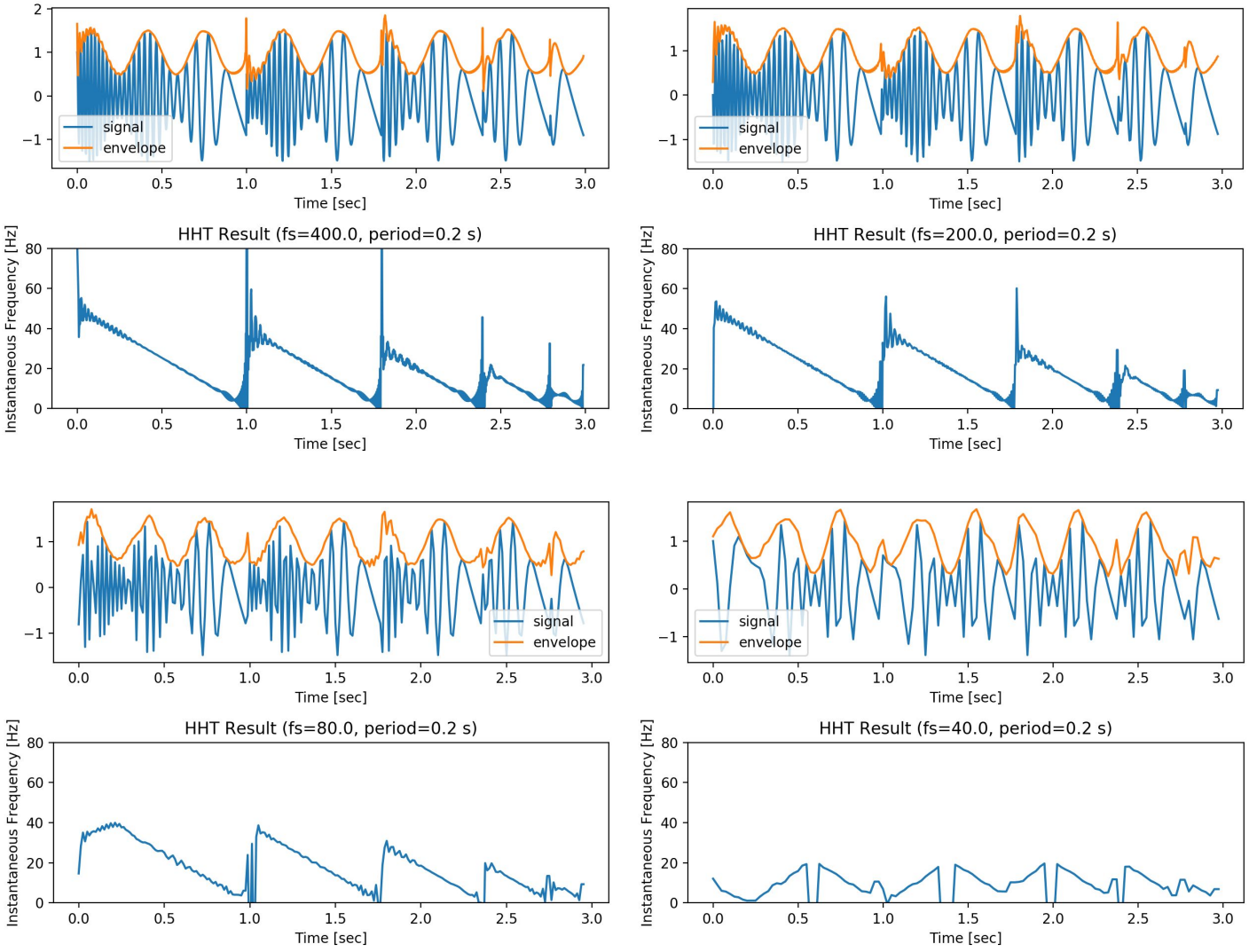


Observation:

(i) Comparing with STFT, HHT is more robust to STFT, which means it is better for real-time analysis of frequency behavior.

(ii) There are some peaks on the envelope curves and are more serious for shorter periods. These can be reduced with further low-pass filtering.

(b) With Different Frames/Second (fs)



Observation:

(i) Both envelope curve and instantaneous frequency are robust to sampling rates, which means we can perform real-time analysis while preserving both data rate and the trend of the data.

(ii) The peaks are reduced in lower-fs settings.

6. Conclusion

By implementation, I proposed a practical way to integrate time-frequency analysis with InfluxDB, which expands horizons of data analytics on time series databases and enables users to have a more clear view on both time and frequency aspects of real-time data. Besides, I also conduct some experiments to investigate the trade-off between segmentation period and precision of the spectrum, which is a serious problem for all FT-based TFA approaches but less severe for non-FT approaches such as HHT. Thus, HHT is a better choice for real-time analysis of both time and frequency domains on time-series databases such as InfluxDB.

7. References

[1] NTU GICE *Time-Frequency Analysis and Wavelet Transform* course material. http://djj.ee.ntu.edu.tw/TFW.htm, Jan 2018.

[2] S. H. Nawab and T. F. Quatieri, “Short time Fourier transform,” in Advanced Topics in Signal Processing, pp. 289-337, Prentice Hall, 1987.

[3] S. Qian and D. Chen, Joint Time-Frequency Analysis: Methods and Applications, Chap. 5, Prentice Hall, N.J., 1996.

[4] E. P. Wigner, “On the quantum correlation for thermodynamic equilibrium,” Phys. Rev., vol. 40, pp. 749-759, 1932.

[5] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” Proc. R. Soc. Lond. A, vol. 454, pp. 903-995, 1998.

[6] N. E. Huang and S. Shen, Hilbert-Huang Transform and Its Applications, World Scientific, Singapore, 2005.

[7] InfluxDB. https://github.com/influxdata/influxdb, Jan 2018.

[8] influxdb-python. https://github.com/influxdata/influxdb-python, Jan 2018.

[9] SciPy Signal Processing. https://docs.scipy.org/doc/scipy/reference/signal.html, Jan 2018.