Spectral Clustering Analysis of High-speed Train Seismic Events and 4D Ground Frequency Map

Yiran Jiang1,2, Tiezhao Bao1,2, Jieyuan Ning1,2, and Xianbing Zhang1
1 School of Earth and Space Sciences, Peking University, Beijing, 100871, PRC
2 The Joint Research Group of High-Speed Rail Seismology, Beijing 100029, PRC
Email: yiranj@pku.edu.cn

Abstract. Employing a temporary array under a high-speed railway viaduct deployed by Peking University, we observe how the spectral characteristics vary with changes of the speed and the model of the train as well as the rail and groundsill by using the clustering algorithm K-Means. For a train in uniform motion, the spectrum of high-speed rail seismic wave is mainly composed of nearly equally spaced peaks and its fundamental frequency is equal to the ratio of the train speed to the carriage length. We reduce the influence from the train speed by aligning the fundamental frequency to make the spectrum pattern clear and easy for comparison. Clustering results show that the spectra of the high-speed rail seismic events have stable patterns under the same train model, rail and groundsill conditions; the stable spectrum patterns change significantly with the changes of the train model, rail and groundsill conditions; monitoring the stable spectral characteristics might be used in safety control of high-speed rail. We apply the clustering method on all the stations of the array. In order to obtain spectra with higher signal to noise ratio at farther stations, we consider the variation of train type and stack the spectrums of three components of high-speed rail seismic signal produced by trains of the same type on the same station. Using the clustering algorithm, we get the regular pattern of how the three component spectra vary with the train type and station position. Based on the above research on the characteristics of high-speed rail seismic spectra and their variation, we propose the concept of 4D ground-frequency map, and discuss its practicability in monitoring the status of high-speed rail and its surrounding media.

1. Introduction
High-speed trains make much noise. We try to turn it into useful signal. Studies on high-speed railway earthquakes have been available since 2001, and there have been more and more related studies in recent years [1,2,4,5]. Liu and Jiang [3] developed a method to cluster high-speed rail events’ waveform spectra. We improved this method and applied it on our research, as we focus on understanding how the wavefield varies with the train model, moving velocity and the rail and groundsill conditions in frequency domain. Map in Figure 1 shows the locations of the high-speed rail (HSR) and the stations under viaduct. Stations are numbered from north to south (1~11).
2. High-speed rail spectrum

High speed rail seismic events have equally partitioned spectra (Figure 2). High-speed rail consists of 8 or 16 carriages. We simplify the situation that the wavefield from different carriages are the same with different time delay according to the carriage location (Figure 3). So, the total waveform $r(t)$ can be expressed as:

$$r(t) = \sum_{k=0}^{n-1} r_0(t - k \cdot \frac{L}{v})$$

(1)

where, $r_0$ is the waveform induced by a single carriage; $n$ is the number of carriages; $L$ is the length of carriage; $v$ is the velocity of train. And $r(t)$’s spectrum is:

$$R(f) = R_0(f) \sum_{k=0}^{n-1} e^{k \frac{L}{v} \cdot 2\pi f i}$$

(2)

where, $R$ is the spectrum of the total waveform; $R_0$ is the spectrum of a single carriage. $\sum_{k=0}^{n-1} e^{k \frac{L}{v} \cdot 2\pi f i}$ make the $R_0$ equally partitioned and its fundamental frequency $f_0$ is:

$$f_0 = \frac{v}{L}$$

(3)

We sort the trains by their velocity and can see the fundamental frequency vary with the velocity (Figure 4).

Figure 1. Distribution of stations and the high-speed rail.

Figure 2. Waveforms and their spectra of a high-speed rail seismic event.
3. **Align Spectra by fundamental frequency**

To reduce the influence of train’s velocity, we align the spectrum by its fundamental frequency. We developed a method to find the accurate value of the fundamental frequency. First, we calculate the ‘spectrum’ (Figure 5(b)) of the train’s spectrum (Figure 5(a)) and find its max peak in the ‘spectrum’. We choose the max peak’s reciprocal as a rough estimate of fundamental frequency called $f_0'$. Then, we search around $f_0'$ to find a more accurate frequency $f_0$ by maximizing the following formula:

$$\max \sum_{k=0}^{f_{\text{max}}-f_0 \leq 0.2, f_0 + 0.2} |R(f_0 * k)|, \ f_0 \in [f_0' - 0.2, f_0' + 0.2] \quad (4)$$

where $f_{\text{max}}$ is the Nyquist frequency of records. We align different the spectra according to the $f_0$ (Figure 5(c)).

4. **Stack Spectra and Cluster**

We stack all the aligned spectra (Figure 6) and select 22 frequency with high energy as characters. As the trains from north to south are different from that from south to north, we consider these two kinds of trains separately. In Figure 6, we stack all the spectra on the same station from the same direction (b and c) and stack all the spectra of the same train (d and e).
5. Station Clustering

Figure 7 and Table 1 show the clustering results on station. Different kinds of stations have different spectrum patterns. As Table 1 shows, sometimes the spectra on the same station with different directions are different. On the same station, the groundsills are the same, but its spectra with different directions are different. The difference may be related to the difference between the rails with opposite directions.

![Figure 7. Clustering results on stations. The spectra of each class (a-d) and their stacked spectra (e-h)](image)

| Station Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---------------|---|---|---|---|---|---|---|---|---|----|----|
| From North to South | 1 | 1 | 4 | 1 | 2 | 2 | 4 | 1 | 4 | 2 | 3 |
| From South to North | 1 | 4 | 4 | 1 | 3 | 2 | 1 | 1 | 4 | 3 | 3 |

Note: The first line is the station number, the second line is the clustering result of the train records from north to south at the station, and the third line shows the train records from south to north at the station. Clustering results

6. Train Clustering

We stack spectra of the same train on every station and cluster them. Figure 8, Figure 9 and Table 2 show the clustering results of trains with different directions. Trains form north to south can be classified into two classes: 1R and 2R; trains form south to north can be classified into two classes: 1L, 2L and 3L.

To find whether the spectrum pattern is stable or not, we associate each event from north to south with specific trips and their models as Table 2 shows. The spectra of the trains belonging to the same trip in different days have negligible difference. The trains of the same model also have much similar frequency pattern (trains of models like CRH380ATXCL, CR400AFCL, CRH380A, NCRH380AL belongs to class 1 and trains of models like NCRH380AL belongs to class 2). All these have shown that the spectra of high-speed rail seismic events have stable and distinguishable patterns.

Totally, the spectra can be classified into three classes. Two classes are similar and are far different
from the third class. According to the arrival time, the third class are trains called ‘high-speed overnight sleeper’ which have a different model from the norm high-speed rail.

![Figure 8](image)

**Figure 8.** Clustering results on trains from N to S. The spectra of each class (a-b) and their stacked spectra(c-d).

![Figure 9](image)

**Figure 9.** Clustering results on train from S to N. The spectra of each class (a-c) and their stacked spectra (d-f).

### 7. 4D Frequency Map

We apply the previous method on all stations (Figure 10(a)) to see the spectrum pattern distribution in space domain. We sort them by distance from high-speed rail (Figure 10(b)). As we find different model have different spectrum pattern, we stack spectra of the trains which have been classified into the same class (1R or 2R). For trains belonging to 1R, the result is showed in Figure 11. After stacking, the low-frequency energy can be seen at stations very far from high-speed rail.

Then we cluster different components’ spectra (R, T, Z) separately. The results (Figure 12) show that at most stations, the classes are the same. But we can see that there are still some differences among different components. It indicates that different components may have different information. And close stations often have similar classed. It indicates that the spectra patter may be up to the propagation path and the conditions of the medium around them.

Based on the existing theory, it is difficult to directly calculate the impact of changes in EMUs, railroad tracks, subgrades, and media on the high-speed rail’s frequency spectrum. That is, considering the change of the frequency spectrum in the time dimension of the train passing, the original three-dimensional spectrum (R, T, Z) is expanded to four dimensions. The change of spectrum mode on a single station distinguishes the effects of EMUs, railroad tracks, subgrades, and media. If the characteristics of the spatial distribution of this spectrum mode change are considered, it is possible to identify whether these factors change or not. We proposed that as a monitoring concept called 4-D Frequency Map which may help monitoring the condition change around high-speed rail. The specific verification and practicality need to be confirmed in the longer and denser observation in the future.
Table 2. Clustering results, train number and train type of trains from north to south

| hour | minute | Class | Trips | model       |
|------|--------|-------|-------|-------------|
| 9    | 5      | 8     | 0     | G309        |
| 11   | 10     | 9     | 0     | G605        |
| 12   | 15     | 7     | 0     | G609        |
| 13   | 20     | 9     | 0     | G611        |
| 18   | 31     | 10    | 0     | G1571       |
| 11   | 15     | 8     | 0     | G65         |
| 11   | 0      | 10    | 0     | G6741       |
| 17   | 11     | 7     | 0     | G523        |
| 14   | 32     | 0     | 4     | G661        |
| 15   | 16     | 0     | 4     | G673        |
| 16   | 58     | 0     | 6     | G573        |
| 17   | 33     | 0     | 5     | G561        |
| 18   | 9      | 0     | 7     | G563        |
| 19   | 14     | 0     | 9     | G6705       |
| 21   | 53     | 0     | 10    | G6745       |

Note: The first and second columns are the time to arrive in Baoding East from the time of day and speed estimation; the third and fourth columns indicate the number of high-speed rail records corresponding to the trips are divided into two types; the fifth column is the number of trips; the sixth column is the model of the query.

Figure 10. Locational relationship between stations and high-speed rail. (a) Station distribution map. The black line indicates the high-speed rail line, the red triangle indicates the station, the horizontal axis indicates the east longitude, and the vertical axis indicates the north latitude. (b) The distance between the station and the high-speed rail: The stations are numbered from near to far by the distance of the rails, denoted by s; the vertical axis is the distance from the station to the high-speed rail, denoted by d.

Figure 11. Stacked three-component (R, T, Z) amplitude spectrums of 1R high-speed rail seismic events on stations from the near to the distant.
8. Conclusions

1) The spectra of HSR seismic signals are mainly composed of equally-partitioned peaks which mainly depend on the train speed and the length of the carriage.
2) Clustering analysis shows that the spectral modes have stable patterns varying with train model, as well as rail and groundsill conditions.
3) The stable frequency pattern can be used for HSR operation safety monitoring.
4) After stacking, stable discrete peaks can be observed at stations farther away from the high-speed rail.
5) The three components of high-speed rail seismic signals have different spectral characteristics; different types of EMU models also have different spectral characteristics.
6) The high-speed rail seismic signals are affected by the propagation path and the conditions of the medium around them. The frequency spectrum of high-speed rail seismic signals at stations with similar environments is also similar.
7) Taking the vehicle type into consideration, studying the superimposed 3D spectrum of the three-component frequency spectrum with time and space (station) may be used to monitor the safety of the high-speed rail and its surrounding environment changes.

References

[1] Cai Y, Sun H and Xu C 2008 Three-dimensional analyses of dynamic responses of track–ground system subjected to a moving train load Computers & Structures 86 816-824.
[2] Ditzel A, Herman G C and Holscher P et al. 2001 Elastic waves generated by high-speed trains Journal of Computational Acoustics 09 833–840.
[3] Liu L and Jiang Y 2019 Attribution extraction and feature analysis for large amount of high-speed-train seismic events Chinese Journal of Geophysics-Chinese Edition 62 2313-2320.
[4] Quiros D A, Brown L D and Kim D 2016 Seismic interferometry of railroad induced ground motions body and surface wave imaging: Geophysical Journal International 205 301-313.
[5] Sussmann T R, Thompson H B and Stark T D et al. 2017 Use of seismic surface wave testing to assess track substructure condition Construction and Building Materials 155 1250-1255.