The Broadband Virtual Shot Gathers Construction Based on High-Speed Train-Induced Seismic Wave

Xiaokai Wang, *Member*, *IEEE*, Shengpei Xia, Xinyue Pan, Baoli Wang, Dawei Liu, and Wenchao Chen, *Member*, *IEEE*

***Abstract*— The moving high-speed train (HST) generates strong and repeatable vibrations in the railway roadbed, resulting in the propagation of complicated seismic waves into the subsurface medium. Therefore, the moving HST could be considered as a novel seismic source to detect the subsurface structure near high-speed railways (HSRs). An HST consists of several carriages. Therefore, the moving HST is one typical combined moving source that stimulates complex interference wavefields. Seismic interferometry (SI) is one commonly used method to generate virtual shot gather. However, the HST-induced seismic signal has one typical equidistant narrowband discrete (ENBD) spectrum feature, which means many frequency components are missing. Thus, the virtual shot gather based on an HST-induced seismic signal does not have sufficient bandwidth. In this study, we propose a processing scheme to construct broadband virtual shot gathers by distinguishing the moving direction of the HST and the propagation direction of seismic waves and stacking the HST events with different speeds. The feasibility of our proposed method is validated using both synthetic and real HST data, providing a broadband virtual shot gathers from HST-induced seismic signals.**

***Index Terms*— Broadband, high-speed train (HST), seismic interferometry (SI), seismic waves, virtual shot gather.**

I. Introduction

CHINA has constructed the largest high-speed railway (HSR) network in the world. By the end of 2023, China HSR's mileage had exceeded 45 000 km, which accounts for more than two-thirds of the world's total HSR mileage. Thousands of heavy high-speed trains (HSTs) run fast across this network every day, generating very complicated seismic waves. Therefore, the moving HST is recognized as one novel seismic source.

Conventional seismic sources, such as explosive sources and vibroseis sources, are strictly prohibited to use near HSR. Therefore, it lacks effective seismic sources to detect the underground medium near HST. The HST can be considered as one potential effective seismic source near HSR due to the following advantages. HSTs have similar carriage structures. Therefore, the HSTs can stimulate repeatable seismic waves, which provide ideal datasets for multiple stacking. HST has a very large mass. Thus, it can stimulate stronger HST seismic waves than other traffic vehicles. HST is environmentally friendly because it does not add extra source costs. Therefore, HST can be considered as one unique seismic source due to the abovementioned advantages. The geophones near HSR can robustly record the seismic waves induced by moving HST, and the received datasets have been used for underground structure imaging, urban underground space detection, and HST moving status monitoring [1], [2], [3].

In China's intensive HSR network, viaducts make up a significant proportion [4]. For example, Beijing–Shanghai HSR's viaduct ratio is 86%. HST generates vibrations on the bridge piers during their operation, which are transmitted to the ground through the bridge-pier structure, and various seismic waves are stimulated and propagated along the medium [5], [6]. Although the viaduct roadbed construction cost is more expensive than the flat ground roadbed construction cost, the moving-HST-induced seismic waves under the viaduct situation can propagate longer distances than those under the flat ground roadbed situation [7]. Therefore, the moving-HST-induced seismic waves under the viaduct situation carry rich underground media information.

An HST typically consists of several carriages with the same structure, making the moving HST a complex seismic source, often referred to as a combined moving source. The complex HST source and viaduct roadbed make the stimulated seismic wavefield extremely complex. Thus, the common seismic imaging methods cannot be easily used for the HST-induced seismic dataset. Seismic interferometry (SI) can simplify the complex wavefield by constructing a virtual shot gather, which is frequently used in shallow structure imaging. SI computes the cross correlation function between two receivers, where one receiver serves as the seismic source and the other as a receiver.

The virtual shot gather generated by SI is similar to the commonly used shot gather, which provides researchers convenience to image structures. Shapiro et al. and Behm et al. applied SI to underground structure imaging [8], [9]. Nakata et al. [10] and Quiros et al. [11] treated freight trains

Received 21 May 2024; revised 3 September 2024; accepted 7 October 2024. Date of publication 22 October 2024; date of current version 29 October 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFA0716904 and in part by the National Natural Science Foundation of China under Grant 41974131. *(Corresponding author: Wenchao Chen.)*

Xiaokai Wang, Shengpei Xia, Xinyue Pan, Dawei Liu, and Wenchao Chen are with Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China (e-mail: xkwang@xjtu.edu.cn; 3120305117@stu.xjtu.edu.cn; pxy0022@stu.xjtu.edu.cn; liudawei2015@stu.xjtu.edu.cn; wencchen@xjtu.edu.cn).

Baoli Wang is with Xi'an Research Institute of China Coal Research Institute, Xi'an, Shaanxi 710077, China (e-mail: pooly1981@163.com).

Digital Object Identifier 10.1109/TGRS.2024.3481272

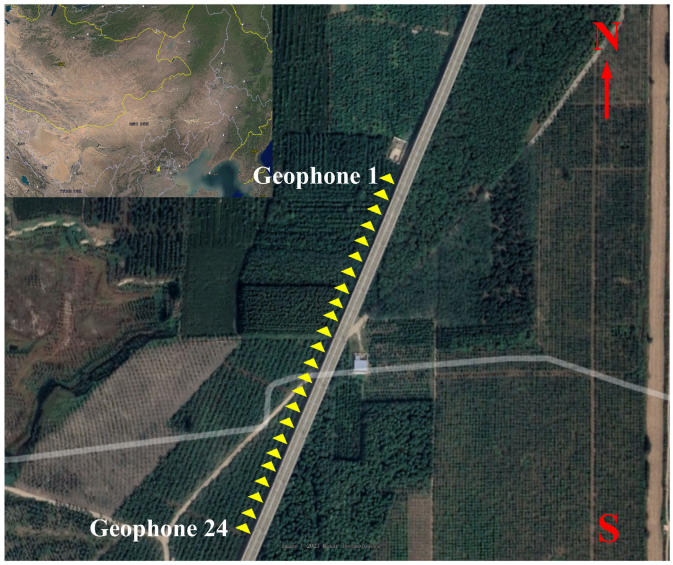

Fig. 1. Satellite photograph of the geophone array in Baoding, China.

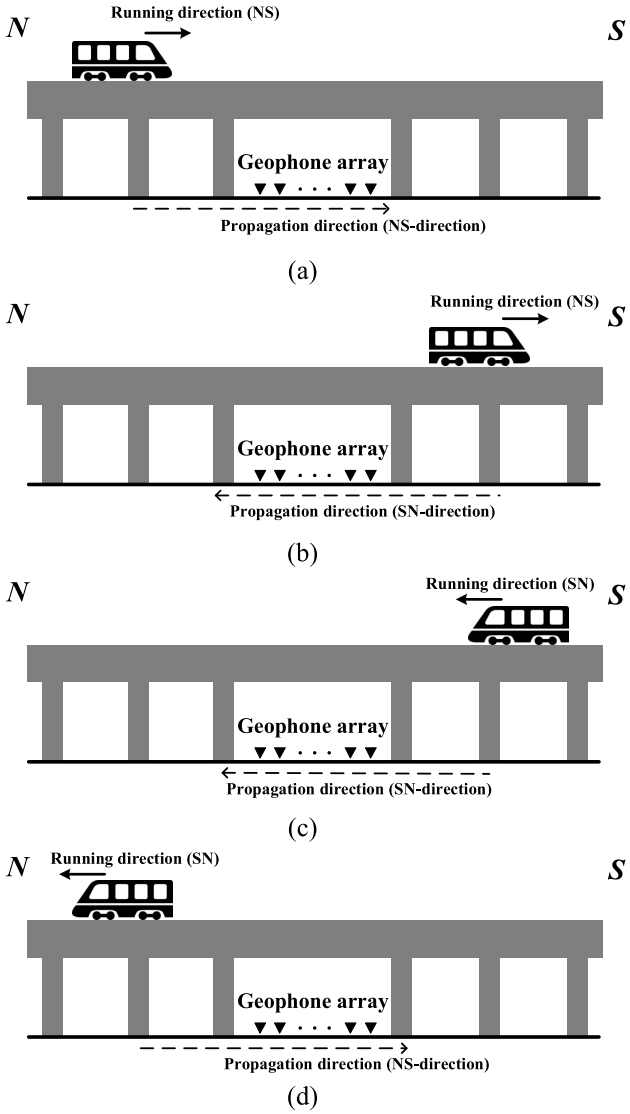

Fig. 2. HST moving direction and wave propagating direction. (a) HST approaching the array with NS direction, (b) HST leaving the array with NS direction, (c) HST approaching the array with SN direction, and (d) HST leaving the array with SN direction.

as seismic sources and successfully extracted both surface waves and body waves from traffic. Zhang et al. studied SI for HST-induced seismic data. Liu et al. [13] analyzed the effects of HST moving direction and speed on SI. Furthermore, Liu et al. [14] successfully extracted scattered surface waves using HST-induced seismic data. You et al. [15] proposed a new HST signal selection method, which automatically selects the HST-induced seismic data when the HST passes through a stationary phase zone. Shao et al. applied SI to the HST-induced seismic dataset recorded by the distributed acoustic sensing (DAS) observation system, which effectively improved the signal-to-noise ratio (SNR) of the extracted surface waves [7].

When the HST passes by the geophone array, the received HST-induced seismic dataset suffers from serious crosstalk interference caused by multiple piers simultaneously stimulating seismic waves. Therefore, the conventional methods for SI involve removing the portion where the HST passes by and then performing cross correlation on the remaining HST data. However, some influencing factors, such as the HST moving direction, the wave propagating direction, and the number of HSTs with different speeds (different speeds of HSTs induce waves with varying wavelength frequency bands), are not fully considered. Therefore, we pay attention to these factors and propose one robust broadband virtual shot gather construction method.

In 2019, we installed a geophone array with 24 three-component geophones under the Beijing–Guangzhou HSR viaduct in Baoding, China. The geophone array is parallel to the railway track (as shown in Fig. 1). The length of the geophone array is 368 m with a spacing of 16 m. 49 HST passing-by HST events are recorded within 2 h of observing time.

The HST passes through this array in the north–south direction (NS direction) or south–north direction (SN direction). Therefore, the seismic waves stimulated by the HSTs also propagate along the NS direction or SN direction. When the HST approaches (or leaves) this array with NS direction, the seismic wave propagates in the NS direction (or SN direction) within the geophone array area [as shown in Fig. 2(a) and (b)]. Similarly, when the HST approaches (or leaves) this array with SN direction, the seismic wave propagates in the SN direction (or NS direction) within the geophone array area [as shown in Fig. 2(c) and (d)]. Therefore, the same direction waves can be stacked to improve the quality of the virtual shot gather. Furthermore, the seismic waves stimulated by the moving HST have a typical equidistant narrowband discrete (ENBD) spectrum, which cannot provide enough frequency components to form a wide frequency band virtual shot gather. The seismic waves stimulated by the HST with different moving speeds have different peak intervals in the frequency domain. Thus, they can be stacked to enrich the frequency components and further generate a broadband virtual shot gather.

In this work, we propose a method for constructing the broadband virtual shot gathers by fully considering the HST moving direction, wave propagating direction, and the HST events with different speeds. The structure of this article is organized as follows. Section II introduces the proposed method for constructing broadband virtual shot gather from HST-induced seismic data. Section III validates the proposed method's effectiveness by constructing virtual shot gathers on synthetic HST data and real HST data. The frequency velocity spectra are also used to further show the proposed method's performance. Some conclusion is drawn in Section IV.

II. Methods

A. HST Source Function

When the HST runs along the tracks, the HST load is acted to the ground through the wheels and tracks, inducing vibrations [13], [16], [17]. In general, an HST is usually composed of a series of carriages with the same structure. Fig. 3 shows an HST's schematic, each carriage containing two bogies on each side and each bogie has two wheelsets. An HST with several carriages runs at a stable speed of v_s. The length of each carriage is L. d_1 represents the distance between two wheelsets, d_2 represents the distance between two bogies within one carriage, and $d_3 = d_1 + d_2$. For the consistency of subsequent formulas, d_0 is equal to 0. The Dirac delta function $\delta(t)$ is used to model one wheelset's effect, which can be treated as the input function. $r(t)$ is used to represent one wheelset's load function, which can be treated as the track system's response function of one wheelset. Therefore, when the HST passed x_s in Fig. 3, the source function $s(t)$ can be written as the convolution of the HST input function $h(t)$ and the response function $r(t)$

$$\begin{aligned} s(t) &= \sum_{n_c=0}^{N_c-1}\sum_{w=0}^{3} r\left(t-\frac{d_w}{v_s}-\frac{n_cL}{v_s}\right) \\ &= r(t) * \sum_{n_c=0}^{N_c-1}\sum_{w=0}^{3}\delta\left(t-\frac{d_w}{v_s}-\frac{n_cL}{v_s}\right) = r(t)*h(t) \end{aligned} \quad (1)$$

where $\delta(t)$ is the Dirac delta function, N_c represents the carriage number of one HST, and * represents the convolution. The HST input function $h(t)$ is the summation of a series of Dirac delta functions

$$h(t) = \sum_{n_c=0}^{N_c-1}\sum_{w=0}^{3}\delta\left(t-\frac{d_w}{v_s}-\frac{n_cL}{v_s}\right). \quad (2)$$

The structural parameters of each carriage within an HST are fundamentally similar. Therefore, the HST input function $h(t)$ can be decomposed into a series of carriage input functions

$$\begin{aligned} h(t) &= \sum_{n_c=0}^{N_c-1} c\left(t-\frac{n_cL}{v_s}\right) = c(t) * \sum_{n_c=0}^{N_c-1}\delta\left(t-\frac{n_cL}{v_s}\right) \\ &= c(t) * p(t) \end{aligned} \quad (3)$$

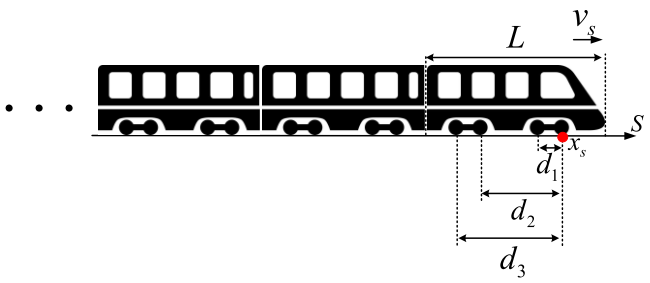

Fig. 3. Schematic of an HST.

where $p(t)$ is one simple function to represent the periodic input of N_c carriages

$$p(t) = \sum_{n_c=0}^{N_c-1}\delta\left(t-\frac{n_cL}{v_s}\right) \quad (4)$$

and $c(t)$ is a single carriage input function

$$\begin{aligned} c(t) &= \delta(t) + \delta\left(t-\frac{d_1}{v_s}\right) + \delta\left(t-\frac{d_2}{v_s}\right) + \delta\left(t-\frac{d_1+d_2}{v_s}\right) \\ &= c_1(t) * c_2(t). \end{aligned} \quad (5)$$

Because one carriage has two bogies with the same structure, single carriage input function $c(t)$ can be decomposed to the convolution of two simple functions

$$c_1(t) = \delta(t) + \delta\left(t-\frac{d_1}{v_s}\right) \quad (6)$$

$$c_2(t) = \delta(t) + \delta\left(t-\frac{d_2}{v_s}\right) \quad (7)$$

where $c_1(t)$ is one bogie input function and $c_2(t)$ is the simple function related to two bogies.

Therefore, the whole HST input function $h(t)$ can be simplified to three simple function's convolutions

$$h(t) = c(t) * p(t) = c_1(t) * c_2(t) * p(t). \quad (8)$$

To further show the amplitude spectrum features of the HST input function $h(t)$, the Fourier transform is applied to the whole HST input function h(t) in (8)

$$H(f) = C(f)P(f) = C_1(f)C_2(f)P(f) \quad (9)$$

where

$$C_1(f) = 2e^{-j\pi f\frac{d_1}{v_s}}\cos\left(\pi f\frac{d_1}{v_s}\right) \quad (10)$$

$$C_2(f) = 2e^{-j\pi f\frac{d_2}{v_s}}\cos\left(\pi f\frac{d_2}{v_s}\right) \quad (11)$$

$$P(f) = e^{-j\pi f\frac{(N_c-1)L}{v_s}}\frac{\sin\left(\pi f\frac{N_cL}{v_s}\right)}{\sin\left(\pi f\frac{N_c}{v_s}\right)}. \quad (12)$$

Therefore, the whole HST input function's amplitude spectrum can be obtained

$$\begin{aligned} |H(f)| &= |C(f)||P(f)| \\ &= 4\left|\cos\left(\pi f\frac{d_1}{v_s}\right)\cos\left(\pi f\frac{d_2}{v_s}\right)\frac{\sin\left(\pi f\frac{N_cL}{v_s}\right)}{\sin\left(\pi f\frac{N_c}{v_s}\right)}\right|. \end{aligned} \quad (13)$$

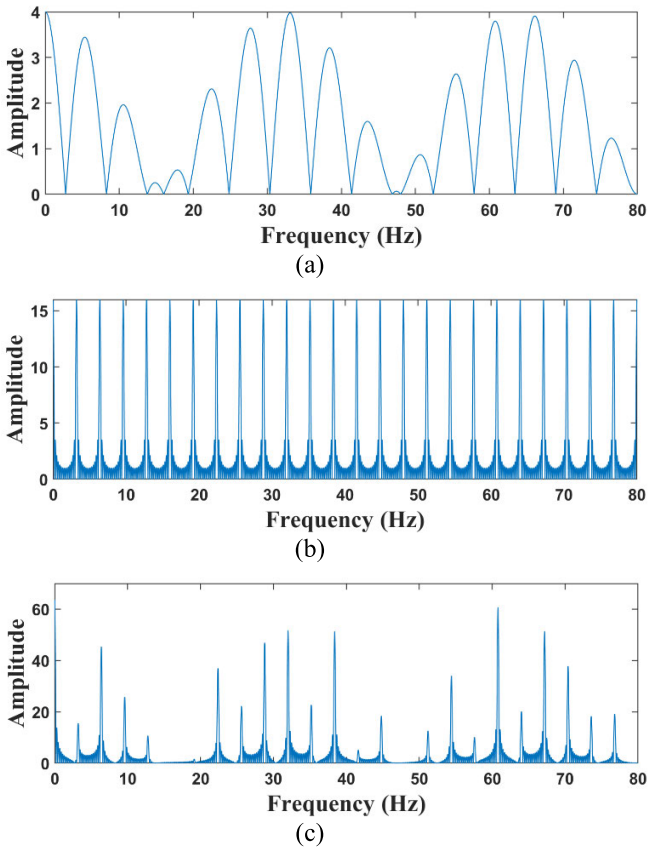

Fig. 4. Amplitude spectra of (a) single carriage input function ($|C(f)|$), (b) periodic input of HST carriages ($|P(f)|$), and (c) whole HST input function ($|H(f)|$).

A typical carriage length of China's HST is 25 m, while wheelset distance d_1 and bogie distance d_2 are 2.5 and 14.5 m, respectively. One typical commercial HST operating speed is 83.33 m/s (300 km/h). The corresponding amplitude spectra $|C(f)|$, $|P(f)|$, and $|H(f)|$ are shown in Fig. 4(a)–(c), respectively. It can be observed that the frequency peaks $|C(f)|$ are significantly wider than the frequency peaks of $|P(f)|$. Equation (12) also shows that $|P(f)|$ is one typical equal-peak-interval narrowband spectrum. $|P(f)|$'s peak interval ΔF is determined by the ratio of an HST speed to the carriage length

$$\Delta F = \frac{v_s}{L}. \tag{14}$$

While the peak width ΔW is determined by an HST speed, the carriage length, and carriage number

$$\Delta W = \frac{v_s}{N_c L}. \tag{15}$$

The amplitude spectrum of the HST input function [see Fig. 4(c)] has a characteristic feature of one typical ENBD spectrum, with the interval between two spectral peaks is 3.33 Hz, which is equal to the ratio between the HST speed (83.33 m/s) and the carriage length (25 m). In addition, we can observe that the envelope of $|H(f)|$ [see Fig. 4(c)] is determined by the single carriage input function's amplitude spectrum $|C(f)|$ [see Fig. 4(a)], indicating that the wheelset distance d_1, bogie distance d_2, and HST speed influence the envelope of $|H(f)|$.

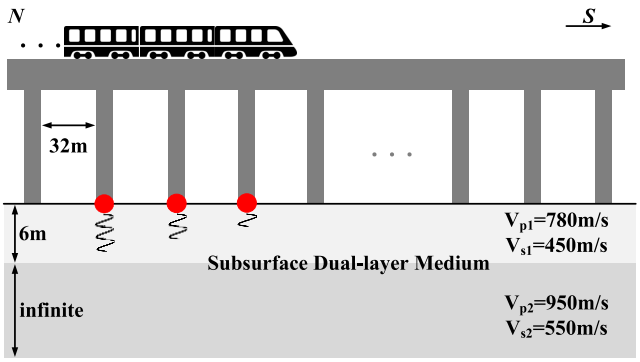

Fig. 5. Excitation process of an HST source at the bridge pier.

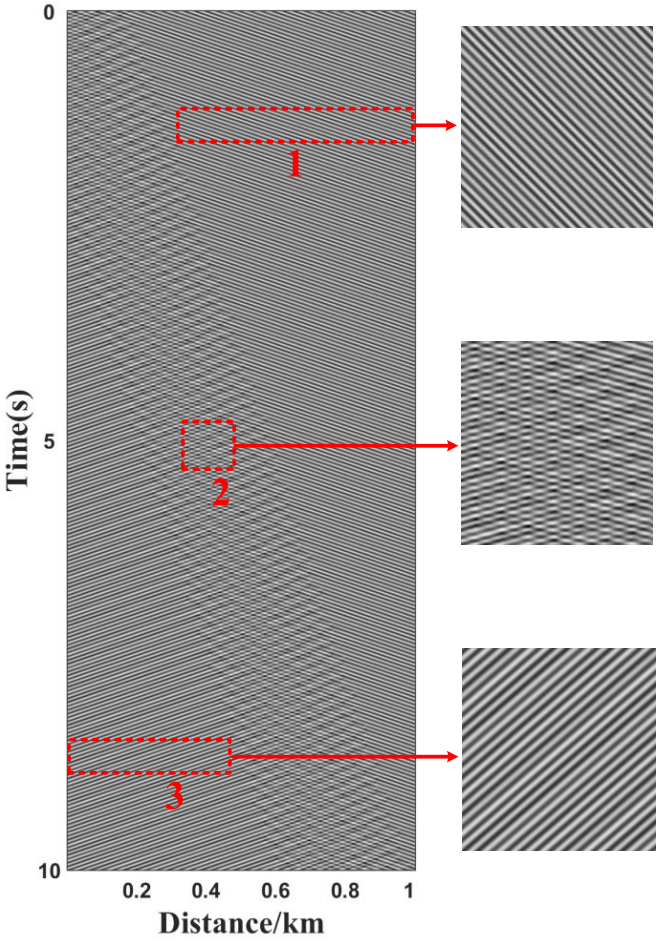

Fig. 6. Synthetic HST-induced dataset from 1 to 501 traces with an HST moving along the NS direction at the speed of 80 m/s.

The source function's amplitude spectrum is determined by the product of $|H(f)|$ and $|R(f)|$ [the amplitude spectrum of response function $r(t)$]. Although it is not easy to obtain the precise expression of $|R(f)|$, we have known that the HST source is different with the fixed pulse point source, which has a broad frequency band.

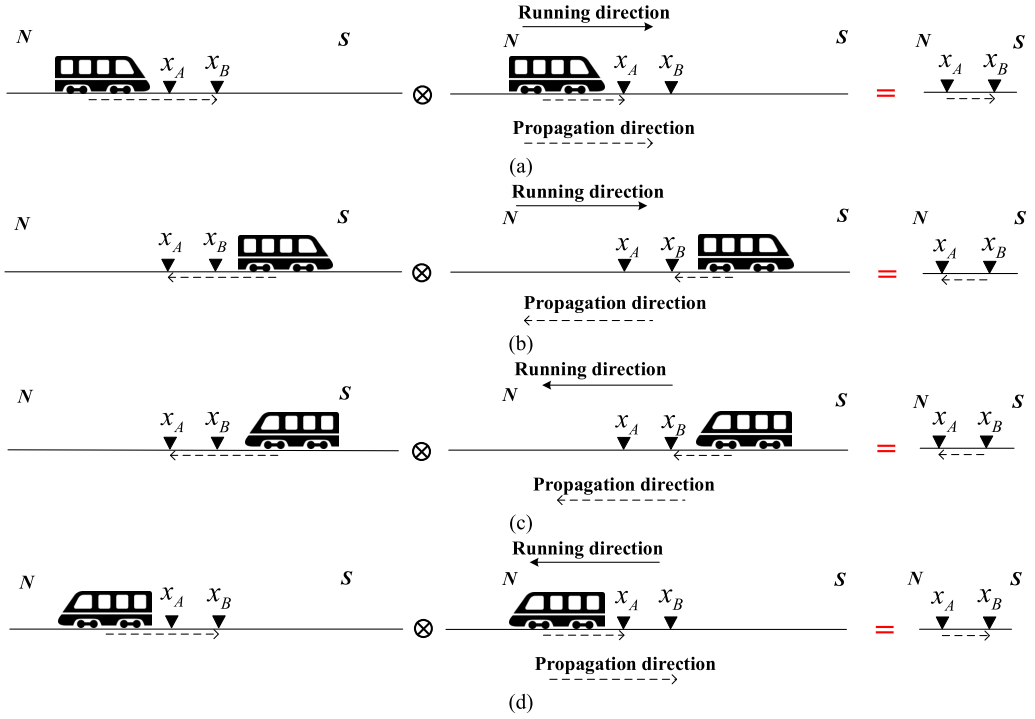

Fig. 7. SI-based virtual shot gather generation procedure for four situations. NS: (a) before the HST arrival at the geophone array and (b) after the HST departure from the geophone array. SN: (c) before the HST arrival at the geophone array and (d) after the HST departure from the geophone array.

B. Features of the HST-Induced Seismic Dataset

We use a finite difference method based on the 2-D elastic wave equation to synthesize the HST-induced seismic dataset. Fig. 5 illustrates the viaduct model and medium parameters. The distance between the piers is 32 m. The subsurface consists of two layers: a shallow layer with a depth of 6 m, a density of 1900 kg/m^3, longitudinal wave velocity $v_{p1} = 780$ m/s, and shear wave velocity $v_{s1} = 450$ m/s; and a deep layer with an infinite depth, a density of 2000 kg/m^3, longitudinal wave velocity $v_{p2} = 950$ m/s, and shear wave velocity $v_{s2} = 550$ m/s. The geophone array contains 1001 geophones and the geophone interval is 2 m.

Fig. 6 presents the synthetic HST-induced seismic dataset from 1 to 501 traces on the Z component, with the HST moving in the NS direction at a speed of 80 m/s. The first rectangular box shows the zoomed-in view dataset before the HST passes by, which corresponds to the situation shown in Fig. 2(a). The second rectangular box represents the zoomed-in view dataset when the HST passes by. It can be observed that the HST simultaneously stimulates both the NS wave and SN wave at each bridge pier, making it difficult to separate NS wave and SN wave. Therefore, it is necessary to remove the dataset corresponding to the HST passing by. The third rectangular box represents the zoomed-in view dataset after the HST passes by, which corresponds to the situation shown in Fig. 2(b). Consequently, the received dataset before (or after) the HST passes by is kept for further processing.

C. SI Based on the HST-Induced Seismic Dataset

SI involves computing the cross correlation of two received signals at different locations [18], [19], [20] and further generates the virtual shot gather. The SI is frequently used in background noise imaging because it does not need the precise source location, triggering time, and source features [18]. Furthermore, SI can improve the wavefield's SNR by using long-time records [7], [21]. The HST-induced seismic wavefield is one typical interference wavefield because many moving wheelsets stimulate seismic waves simultaneously. Therefore, SI can be used in this complex wavefield to generate one virtual shot gather, which can be treated as the shot gather stimulated by one fix point source.

However, we have some extra prior information about this HST source, such as HST moving direction, speed, and carriage structure. Therefore, the SI can be improved to fully utilize the prior information.

There are only two moving directions for HST (NS and SN direction), and we only keep the dataset before and after HST arriving at the geophone array. Therefore, four kinds of virtual shot gathers are generated corresponding to four situations in Fig. 2. Fig. 7 shows the SI-based virtual shot gather generation procedure for four situations, where $\otimes$ represents cross correlation. Position x_A is treated as the source position, while position x_B is treated as the receiver position.

Fig. 7(a) corresponds to the situation where the NS moving HST approaches the geophone array. We use $H(f)$ to represent HST source function in the frequency domain and use $G(x_{A-N}, f)$ to represent the Green function from the HST's position to x_A (HST is on the northside of geophone array), and the received signals (in the frequency domain) at position x_A and position x_B can be obtained

$$U(x_A, f) = G(x_{A-N}, f)H(f) \tag{16}$$

$$U(x_B, f) = G(x_{A-N}, f)G_{AB}(f)H(f) \tag{17}$$

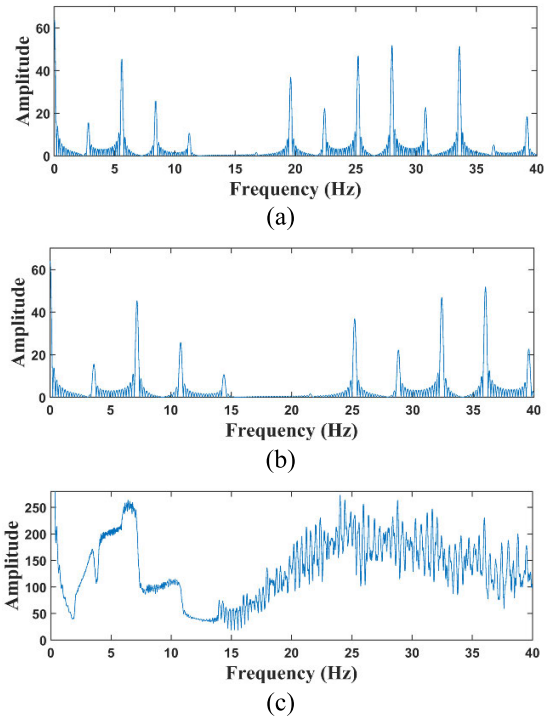

Fig. 8. Amplitude spectra of HST input functions (a) speed is 70 m/s, (b) 90 m/s, and (c) stacking results of 21 HST input functions with speeds ranging from 70 to 90 m/s at 1-m/s intervals.

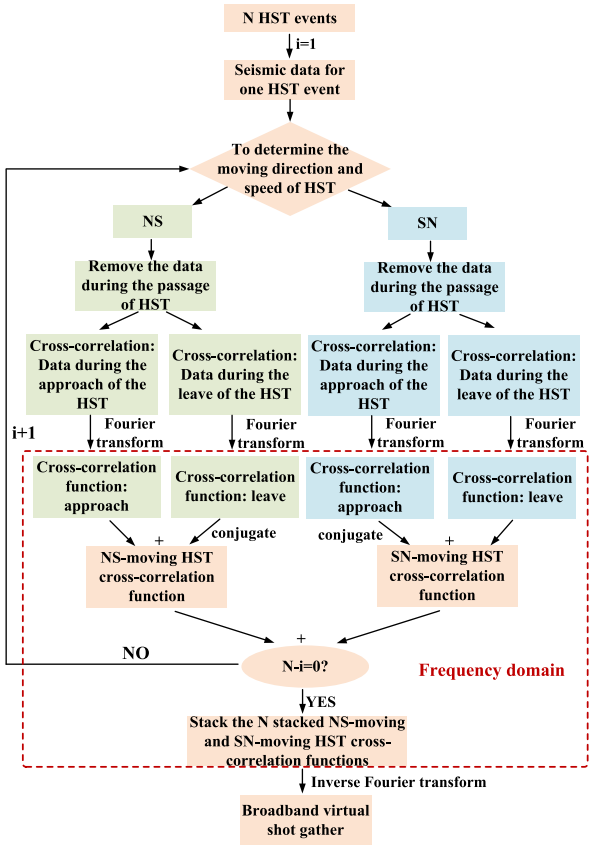

Fig. 9. Proposed workflow for constructing broadband virtual shot gather.

where $G_{AB}(f)$ represents the Green function between x_A and x_B. Then, the cross correlation (in the frequency domain) between the received signals in position x_A and position x_B can be obtained

$$C(x_{AB}, f)_{NS-approach} = U^*(x_A, f)U(x_B, f) = G_{AB}(f)|G(x_{A-N}, f)|^2|H(f)|^2. \quad (18)$$

Fig. 7(b) corresponds to the situation that the NS moving HST leaves the geophone array. We use $G(x_{B-S}, f)$ to represent the Green function from the HST's position to x_B (HST is on the southside of the geophone array), and the received signals (in the frequency domain) at position x_B and position x_A can be obtained

$$U(x_B, f) = G(x_{B-S}, f)H(f) \quad (19)$$

$$U(x_A, f) = G(x_{B-S}, f)G_{AB}(f)H(f). \quad (20)$$

Then, the cross correlation (in the frequency domain) between the received signals in position x_A and position x_B can be obtained

$$C(x_{AB}, f)_{NS-leave} = U^*(x_A, f)U(x_B, f) = G^*_{AB}(f)|G(x_{B-S}, f)|^2|H(f)|^2. \quad (21)$$

Fig. 7(c) and (d) corresponds to two situations where the SN moving HST approaches and leaves the geophone array, respectively. Thus, two corresponding cross-correlations

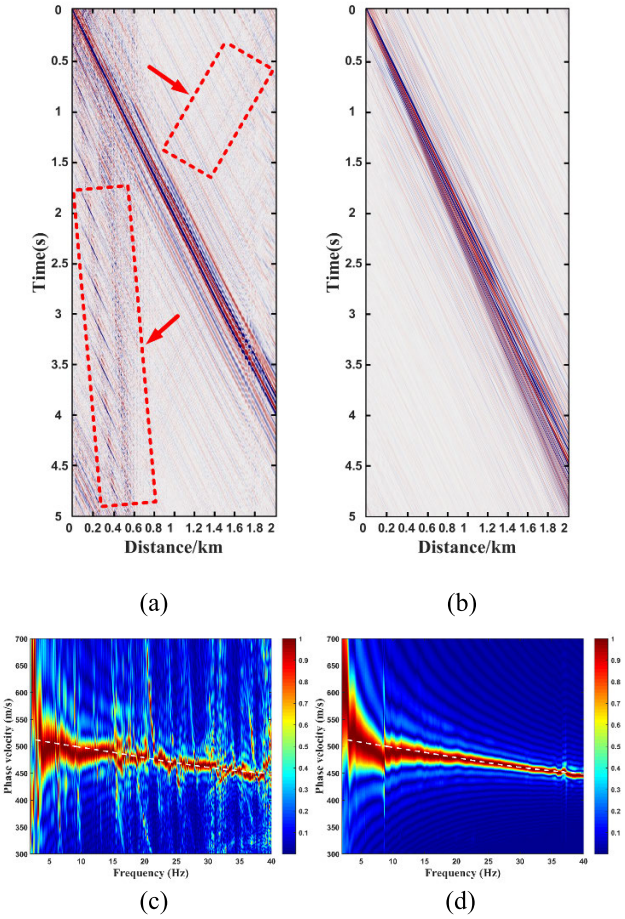

Fig. 10. Results of processing a synthetic HST event using different methods. Virtual shot gather based on (a) conventional method and (b) proposed method. (c) and (d) are the dispersion energy spectra corresponding to (a) and (b).

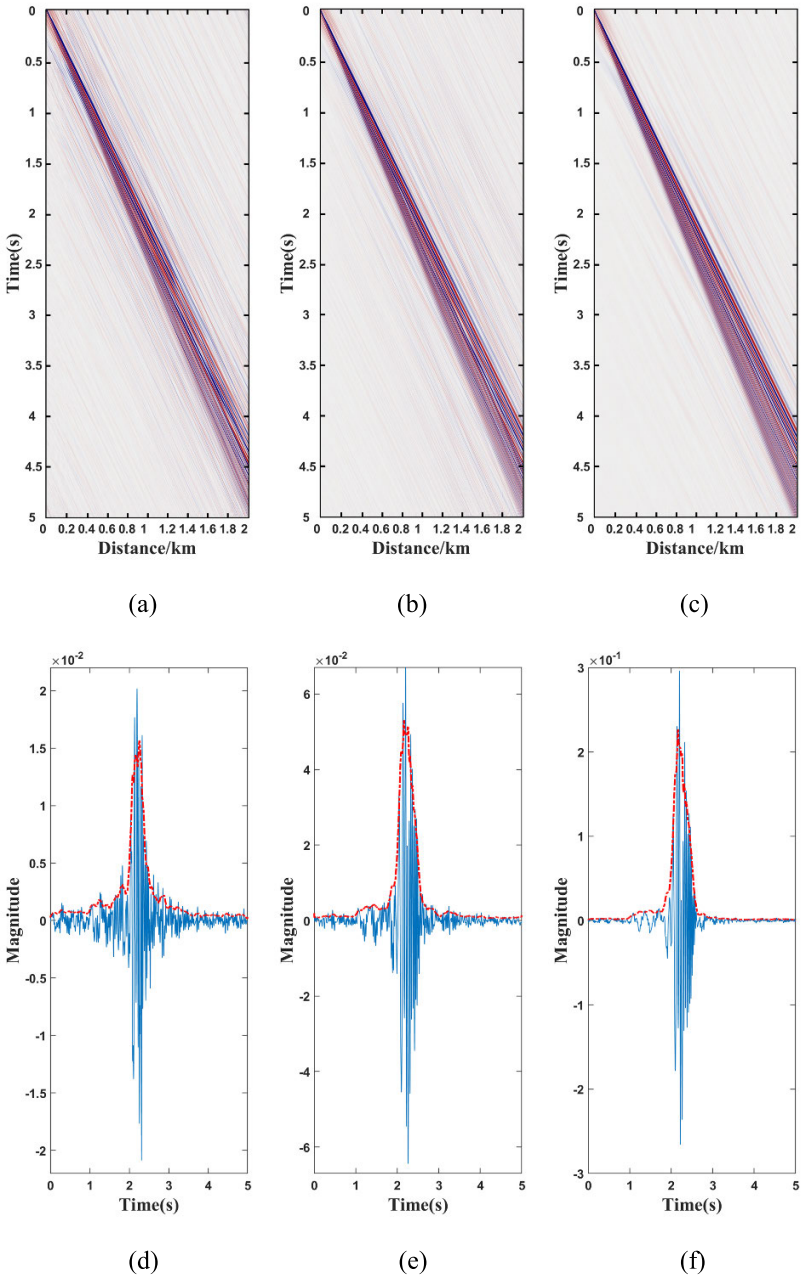

Fig. 11. Virtual shot gathers based on (a) single HST event, (b) four HST events, and (c) 16 HST events and (d)–(f) are the waveforms extracted from the virtual shot gathers [see (a)–(c)] at the 500th trace.

between the received signals in position x_A and position x_B can be obtained in the frequency domain

$$C(x_{AB}, f)_{SN-approach} = U^*(x_A, f)U(x_B, f) = G^*_{AB}(f)|G(x_{B-S}, f)|^2|H(f)|^2 \quad (22)$$

$$C(x_{AB}, f)_{SN-leave} = U^*(x_A, f)U(x_B, f) = G_{AB}(f)|G(x_{A-N}, f)|^2|H(f)|^2. \quad (23)$$

Equations (18) and (21)–(23) do not have the same amplitude spectra but share similar phase spectra.

D. Virtual Shot Gather Stacking Method

From (18) and (21)–(23), four kinds of virtual shot gathers can be obtained. Equations (18) and (23) have the same phase spectra. The phase spectra of (21) and (22) are the conjugate expression of (18)'s phase spectrum. As (13) shows, the HSTs with different moving speeds result in different spectrum peak positions for the HST input function, which can be observed from the amplitude spectra of the HST input functions at velocities of 70 [see Fig. 8(a)] and 90 m/s [see Fig. 8(b)]. Therefore, the source functions in (18) and (21)–(23) may differ for the HST with different speeds. We summed the HST input function's amplitude spectra related to 21 HSTs with different speeds and shown it in Fig. 8(c). It can be observed

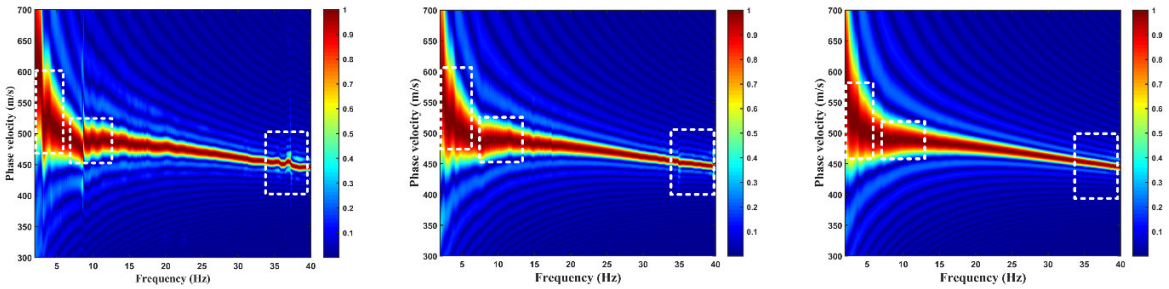

Fig. 12. Dispersion energy spectra based on (a) single HST event, (b) four HST events, and (c) 16 HST events.

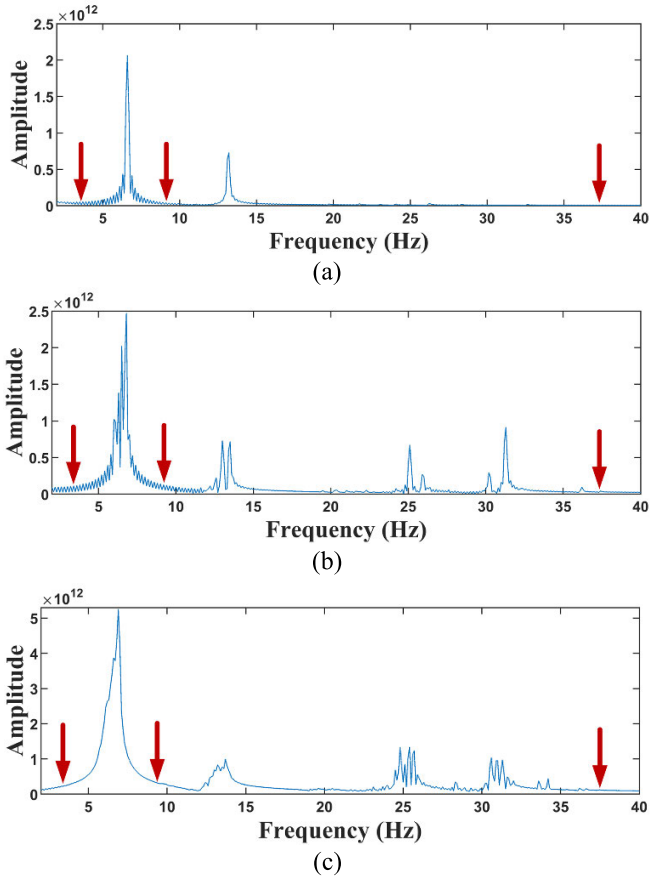

Fig. 13. Amplitude spectrum of the tenth trace of the virtual shot gather based on (a) single HST event, (b) four HST events, and (c) 16 HST events.

that the frequency band is widened. Therefore, to obtain a cross correlation function with broadband spectrum, (18) and (21)–(23) can be stacked. The cross correlation functions relating to NS moving HST and SN moving HST can be stacked in the frequency domain

$$C(x_{AB}, f) = \sum_{n=1}^{N_{NS}} \left[C_n(x_{AB}, f)_{NS-approach} + C_n^*(x_{AB}, f)_{NS-leave}\right] + \sum_{n=1}^{N_{SN}} \left[C_n^*(x_{AB}, f)_{SN-approach} + C_n(x_{AB}, f)_{SN-leave}\right] \quad (24)$$

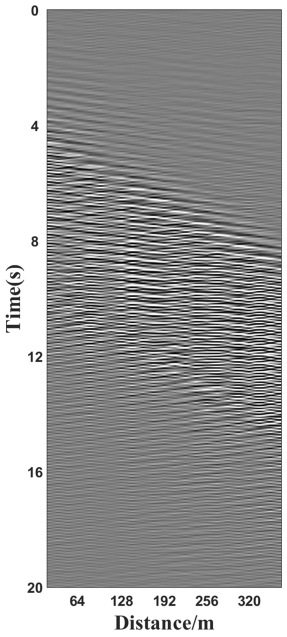

Fig. 14. Real HST event moves at a speed of 79 m/s in the NS direction.

where N_{NS} and N_{SN} represent the number of NS moving HST and SN moving HST. To make the cross correlation function $C(x_{AB}, f)$ have a wider bandwidth, the HST events with different speeds should be used. After obtaining cross correlation functions, the amplitude spectrum of (24) can be used to normalize (24). The proposed broadband virtual shot gather constructing method is summarized in Fig. 9.

III. Data Examples

A. Synthetic Data Example

We use the medium model in Fig. 5 and the 2-D finite difference of the elastic wave equation to generate the seismic dataset induced by the HST with a speed of 80 m/s. A geophone array with 1001 geophones and a 2-m interval is used to receive HST-induced seismic waves. Fig. 6 shows the first 500 traces of the synthetic HST event.

For the conventional method, the workflow involves removing the dataset corresponding to the HST passage by and then performing cross correlation on the remaining HST dataset to obtain the virtual shot gathers. The virtual shot gather

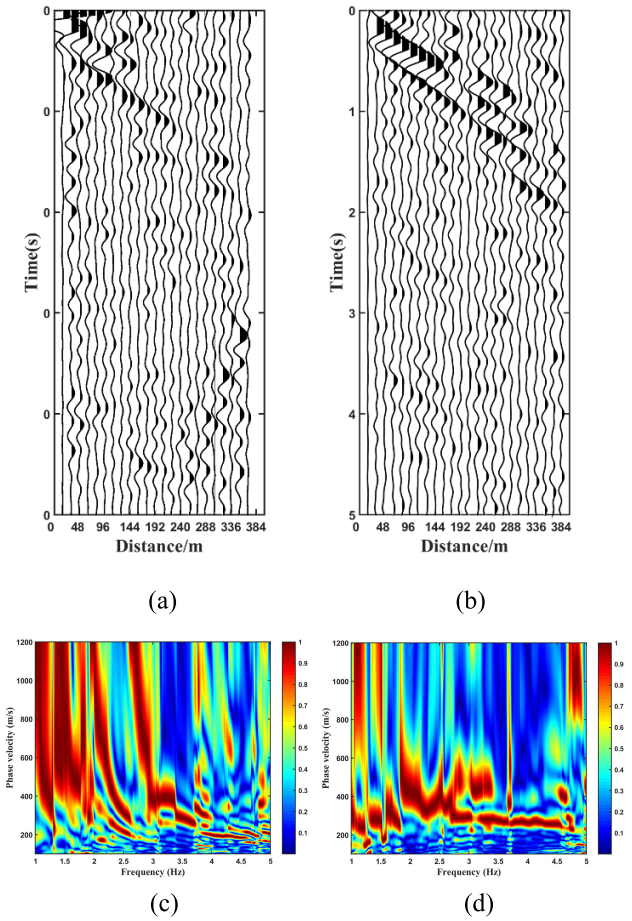

Fig. 15. Results of processing a real HST event using different methods. Virtual shot gathers based on (a) conventional method and (b) proposed method. (c) and (d) Dispersion energy spectra corresponding to (a) and (b).

based on the conventional method [see Fig. 10(a)] indicates that some interference noise exists (highlighted by the red dashed rectangular box). Moreover, as shown in the dispersion energy spectrum [see Fig. 10(c)] obtained using the phase shift method [22], [23], [24], there is a considerable amount of noise interference, and the energy is discontinuous in the low-frequency region. The white dashed in Fig. 10(c) represents the real dispersion curve [25], [26]. It can be seen that the dispersion spectrum obtained using the conventional method does not fit well with the real dispersion curve.

By applying the proposed method to the dataset after removing the data during the passage of the HST, we can obtain a broadband virtual shot gather. As shown in Fig. 10(b), this method reduces noise interference and improves the resolution of the virtual shot gathers. The dispersion energy spectra shown in Fig. 10(d) exhibit less noise interference and better continuity compared to the dispersion spectrum obtained using the conventional method, particularly in the low-frequency region. Similarly, the white dashed in Fig. 10(d) also represents the real dispersion curve. Compared to the conventional method, the dispersion spectrum obtained using the proposed method fits better with the real dispersion curve.

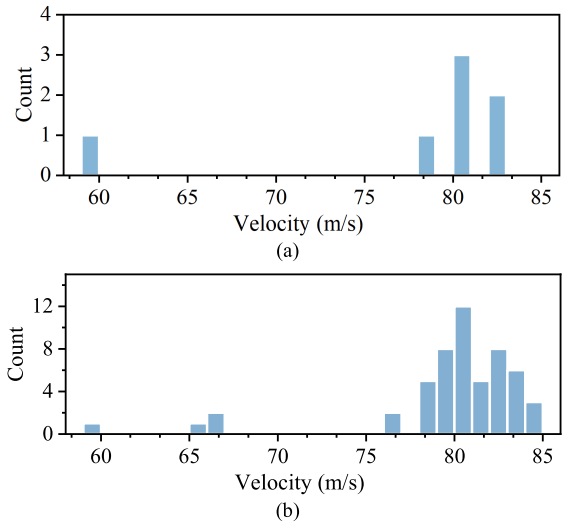

Fig. 16. Velocity distribution of (a) seven real HST events and (b) 53 real HST events.

Next, we simulate the seismic wave field induced by 16 different-speed HST events. Fig. 11(a) shows the virtual shot gather result based on a single HST event with a speed of 80 m/s, while Fig. 11(b) and (c) shows the virtual shot gather based on four synthetic HST events with different moving speeds (75, 78, 81, and 84 m/s) and 16 synthetic HST events with different moving speeds (75–90 m/s), respectively. It can be observed that the quality of the stacked virtual shot gather based on 16 events is greater than that for a single event and four events. Therefore, increasing the HST event number with different speeds can improve the quality of the virtual shot gather.

The waveforms extracted from the virtual shot gathers at the 500th trace [at 1 km in Fig. 11(a)–(c)], as shown in Fig. 11(d)–(f), indicate that the proposed method has the narrowest time width, which indicates that the proposed method has the widest band in the frequency domain. Moreover, Fig. 12(a)–(c) shows the dispersion energy spectra of Fig. 11(a)–(c), respectively. Compared to dispersion energy spectra based on a single HST event [see Fig. 12(a)] and four HST events [see Fig. 12(b)], the dispersion energy spectrum based on 16 HST events has better continuous energy and a higher SNR. This confirms that, as discussed in Section II-D, the virtual shot gather based on multiple HST events with different speeds has a wider frequency band compared to the virtual shot gather based on one HST event.

Fig. 13(a)–(c) shows the amplitude spectrum of the 10th trace of the virtual shot gathers [see Fig. 11(a)–(c)]. The positions indicated by the red arrows correspond to the positions of the white dashed rectangular boxes in the dispersion spectrum [see Fig. 12(a)–(c)]. The amplitude spectrum of a single HST event [see Fig. 13(a)] shows missing frequency components at the positions indicated by the red arrows, resulting in discontinuities in the dispersion spectrum [see Fig. 12(a)] at these missing frequencies. Fig. 13(b) and (c) shows the amplitude spectrum of the 10th trace of the virtual shot gathers based on 4 and 16 HST events, respectively. Compared to the amplitude spectrum of the virtual shot gather based on a

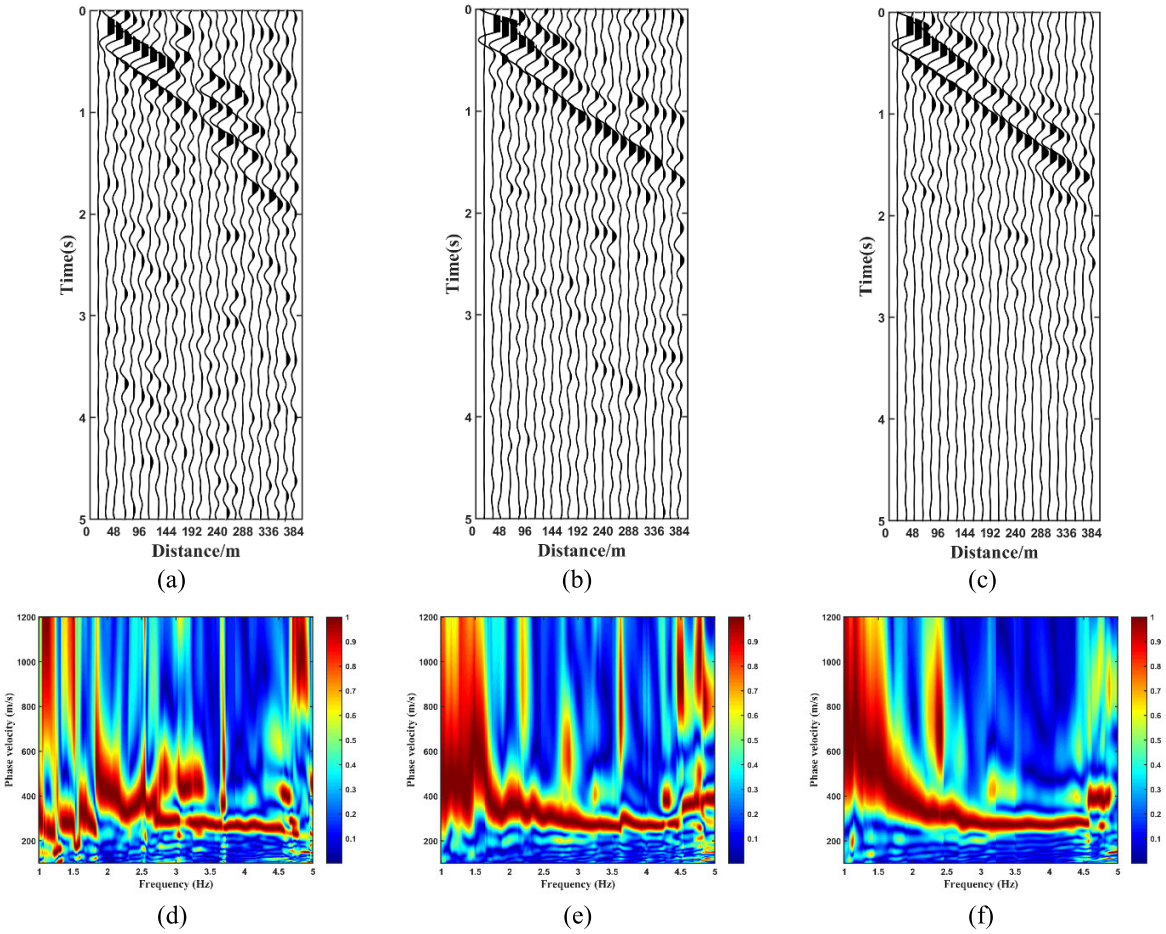

Fig. 17. Virtual shot gathers and dispersion energy spectra for different numbers of real HST events: the virtual shot gathers based on (a) one event, (b) seven events, and (c) 53 events; the dispersion energy spectra based on (d) one event, (e) seven events, and (f) 53 events.

single HST event, they exhibit a wider frequency bandwidth and more abundant frequency components. Therefore, adding multiple HST events with different speeds can enrich the frequency components of the virtual shot gathers.

B. Real Data Example

Fig. 1 shows the position of geophone array for real dataset acquisition near the Beijing–Guangzhou HSR in Baoding. Due to the spacing between geophones being 16 m, only information within the frequency band range of 1–5 Hz was considered.

Fig. 14 presents one real HST dataset. First, the conventional method is applied to construct the virtual shot gather. After removing the dataset during the HST passage by, the virtual shot gather is obtained by cross correlation on the remaining dataset. As indicated in Fig. 15(a), there is substantial noise interference in the virtual shot gather. The dispersion energy spectrum for the virtual shot gather in Fig. 15(a) is shown in Fig. 15(c), exhibits energy discontinuities within the 1–5-Hz frequency ranges, and contains significant noise throughout the entire frequency range.

We apply the proposed method to this real HST event. Fig. 15(b) and (d) shows the virtual shot gather and the corresponding dispersion energy spectrum, respectively. Fig. 15(b) shows that the method reduces noise interference and exhibits better waveform continuity. Compared to the dispersion spectra [see Fig. 15(c)] obtained by the conventional method, it can be seen that the continuity of the dispersion energy spectra [see Fig. 15(d)] at low frequencies is improved and contains less noise interference. These results demonstrate that our proposed method can retrieve a broadband virtual shot gather from HST vibrations.

The virtual shot gather's frequency bandwidth can be widened by stacking multiple synthetic HST events with different speeds. Fig. 16(a) and (b) shows the velocity distribution of 7 and 53 real HST events, respectively, indicating that real HST events move at different speeds. Therefore, we use the real dataset to test the effect of HST event number on virtual shot gather. Fig. 17(a)–(c) shows the virtual shot gathers based on 1, 7, and 53 HST events, respectively. The quality of the virtual shot gathers increases significantly with the increase in the HST event number. The corresponding dispersion energy spectra in Fig. 17(d)–(f) show that the continuity and SNR of the dispersion energy spectra are greatly improved with the increase of HST event number.

Fig. 18(a)–(c) shows the amplitude spectrum of the 19th trace of virtual shot gather [Fig. 17(a)–(c)], indicating that the bandwidth of the virtual shot gathers based on 53 HST events is broader compared to those based on 1 and 7 HST events. Thus, stacking multiple real HST events with different speeds can widen the frequency bandwidth of the virtual shot gather.

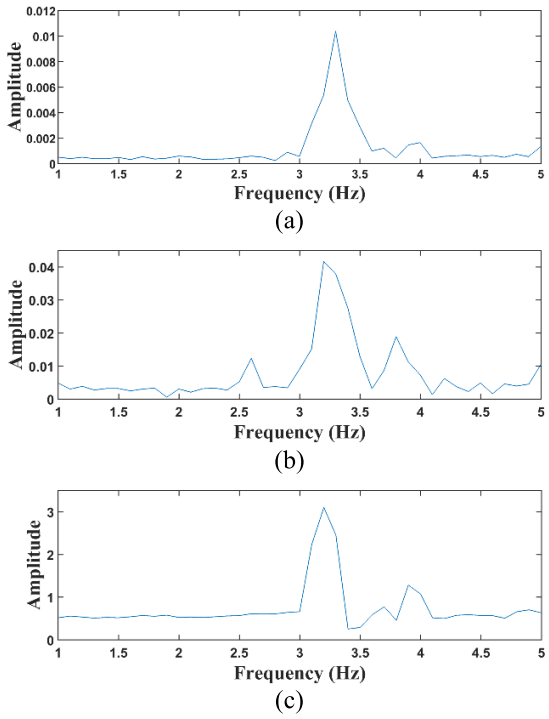

Fig. 18. Amplitude spectrum of the 19th trace of the virtual shot gather based on (a) one event, (b) seven events, and (c) 53 events.

IV. Conclusion

The virtual shot gather can be obtained based on the HST-induced seismic dataset. However, insufficient attention has been given to the HST moving direction, the propagation direction of seismic waves, and the influence of the HST event number. This article introduces a method for obtaining broadband virtual shot gather from the HST-induced seismic dataset. The method involves removing the data during the passage of HST and then using SI to obtain two virtual shot gathers from the data during the approach and leave of the HST. These two virtual shot gathers are then stacked to generate a virtual shot gather with wider bandwidth. By applying our method to both synthetic HST-induced seismic datasets and real HST-induced seismic datasets, we were able to obtain broadband virtual shot gather, which provides a high-quality dataset for shallow surface imaging.

Acknowledgment

The authors thank Dr. Xueliang Li from the Institute of Geology and Geophysics, Chinese Academy of Sciences (IGGCAS), for their assistance in the data acquisition. They also thank the members of the Joint Research Group of HST Induced Seismology for some useful discussions.

References

[1] J. Zhao, Y. Yang, T. Li, and W. Jin, "Application of empirical mode decomposition and fuzzy entropy to high-speed rail fault diagnosis," in *Proc. 8th Int. Conf. Intell. Syst. Knowl. Eng.*, 2013, pp. 93–103.

[2] G. Li, W. Jin, P. Jiang, X. Fu, D. Xiong, and P. Gu, "Research on fault diagnosis technology of high-speed rail for large scale monitoring data," *J. Syst. Simul.*, vol. 26, no. 10, pp. 2458–2464, Oct. 2014.

[3] F. Zhu and W. Jin, "Application of method to determine basic probability assignment in HSR equipment faults diagnosis," *China Railway*, vol. 5, no. 4, pp. 82–85, Apr. 2018.

[4] Z. Wang, C. Chen, W. Bai, and Y. Li, "Analysis of the responses of seismic waves excited by the pile foundations of the viaduct when the high-speed trains passage," *Chin. J. Geophys.*, vol. 65, no. 7, pp. 2622–2635, Jul. 2022.

[5] A. Ditzel, G. Herman, and P. Hölscher, "Elastic waves generated by high-speed trains," *J. Comput. Acoust.*, vol. 9, no. 3, pp. 833–840, Sep. 2001, doi: 10.1142/s0218396x01001133.

[6] X. He, T. Wu, Y. Zou, Y. F. Chen, H. Guo, and Z. Yu, "Recent developments of high-speed railway bridges in China," *Struct. Infrastruct. Eng.*, vol. 13, no. 12, pp. 1584–1595, Mar. 2017.

[7] J. Shao, Y. Wang, and L. Chen, "Near-surface characterization using high-speed train seismic data recorded by a distributed acoustic sensing array," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5912911, doi: 10.1109/TGRS.2022.3153831.

[8] N. M. Shapiro, M. Campillo, L. Stehly, and M. H. Ritzwoller, "High-resolution surface-wave tomography from ambient seismic noise," *Science*, vol. 307, no. 5715, pp. 1615–1618, Mar. 2005, doi: 10.1126/science.1108339.

[9] M. Behm, G. M. Leahy, and R. Snieder, "Retrieval of local surface wave velocities from traffic noise—An example from the la barge basin (Wyoming)," *Geophys. Prospecting*, vol. 62, no. 2, pp. 223–243, Mar. 2014, doi: 10.1111/1365-2478.12080.

[10] N. Nakata, R. Snieder, T. Tsuji, K. Larner, and T. Matsuoka, "Shear wave imaging from traffic noise using seismic interferometry by cross-coherence," *Geophysics*, vol. 76, no. 6, pp. SA97–SA106, Nov. 2011.

[11] D. A. Quiros, L. D. Brown, and D. Kim, "Seismic interferometry of railroad induced ground motions: Body and surface wave imaging," *Geophys. J. Int.*, vol. 205, no. 1, pp. 301–313, Feb. 2016.

[12] H. Zhang, B. Wang, J. Ning, and Y. Li, "Interferometry imaging using high-speed-train induced seismic waves," *Chin. J. Geophys.*, vol. 62, no. 6, pp. 2321–2327, Jun. 2019.

[13] Y. Liu, Y. Yue, Y. Luo, and Y. Li, "Effects of high-speed train traffic characteristics on seismic interferometry," *Geophys. J. Int.*, vol. 227, no. 1, pp. 16–32, May 2021, doi: 10.1093/gji/ggab205.

[14] Y. Liu, Y. Yue, Y. Li, and Y. Luo, "On the retrievability of seismic waves from high-speed-train-induced vibrations using seismic interferometry," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3050205.

[15] B. You, B. Mi, B. Guan, H. Zhang, and Y. Liu, "High-quality surface wave retrieval from vibrations generated by high-speed trains moving on viaducts," *J. Appl. Geophysics*, vol. 212, May 2023, Art. no. 105005.

[16] H.-H. Hung and Y.-B. Yang, "Elastic waves in visco-elastic half-space generated by various vehicle loads," *Soil Dyn. Earthq. Eng.*, vol. 21, no. 1, pp. 1–17, Jan. 2001, doi: 10.1016/s0267-7261(00)00078-6.

[17] V. V. Krylov, "On the theory of railway-induced ground vibrations," *Le J. De Phys. IV*, vol. 4, no. C5, pp. C5–769, May 1994, doi: 10.1051/jp4:19945167.

[18] G. T. Schuster, *Seismic Interferometry*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[19] K. Wapenaar, D. Draganov, R. Snieder, X. Campman, and A. Verdel, "Tutorial on seismic interferometry: Part 1—Basic principles and applications," *Geophysics*, vol. 75, no. 5, pp. 75A195–75A209, Sep. 2010, doi: 10.1190/1.3457445.

[20] J. F. Claerbout, "Synthesis of a layered medium from its acoustic transmission response," *Geophysics*, vol. 33, no. 2, pp. 264–269, Apr. 1968, doi: 10.1190/1.1439927.

[21] H. Zhu, "The study and application of seismic interferometry," Ph.D. dissertation, Jilin Univ., Changchun, China, 2015.

[22] G. Dal Moro, M. Pipan, E. Forte, and I. Finetti, "Determination of Rayleigh wave dispersion curves for near surface applications in unconsolidated sediments," in *Proc. SEG Tech. Program Expanded Abstr.*, vol. 46, Jan. 2003, pp. 1247–1250.

[23] C. B. Park, R. D. Miller, and J. Xia, "Imaging dispersion curves of surface waves on multi-channel record," in *SEG Technical Program Expanded Abstracts*. Houston, TX, USA: Society of Exploration Geophysicists (SEG), 1998.

[24] R.-T. Dai, G.-Z. Shao, X.-Y. Liu, Z.-M. Ren, X.-T. Heng, and X.-D. Ren, "Study of an automatic picking method for multimode dispersion curves of surface waves based on an improved U-Net," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5924214, doi: 10.1109/TGRS.2022.3224919.

[25] E. A. Olafsdottir, S. Erlingsson, and B. Bessason, "Tool for analysis of multichannel analysis of surface waves (MASW) field data and evaluation of shear wave velocity profiles of soils," *Can. Geotechnical J.*, vol. 55, no. 2, pp. 217–233, Feb. 2018, doi: 10.1139/cgj-2016-0302.

[26] E. A. Olafsdattir, "Multichannel analysis of surface waves for assessing soil stiffness," M.Sc. thesis, Univ. Iceland, Reykjavik, Iceland, 2016.

Xiaokai Wang (Member, IEEE) received the B.S. degree in information engineering and the Ph.D. degree in communication engineering from Xi'an Jiaotong University (XJTU), Xi'an, China, in 2006 and 2012, respectively.

He was a Post-Doctoral Fellow at the Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China, from 2012 to 2014. He joined the Department of Computational Geophysics, XJTU, in 2015. He was a Visiting Scholar of the Bureau of Economic Geology, University of Texas at Austin, Austin, TX, USA, from 2016 to 2017. Since 2018, he has been with the School of Information and Communication Engineering, XJTU, where he is currently a Professor. His research interests include seismic attributes extraction, time–frequency analysis, and high-speed rail seismology.

Shengpei Xia received the B.S. degree in communication engineering from Ningxia University, Yinchuan, China, in 2020. He is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an, China.

His research interests include high-speed rail seismology and seismic interferometry.

Xinyue Pan received the B.S. degree in information engineering from Xi'an Jiaotong University, Xi'an, China, in 2021, where she is currently pursuing the M.S. degree in communication engineering.

Since 2024, she has been with China Academy of Aerospace Aerodynamics, Beijing, China. Her research interests include forward modeling of seismic wavefield.

Baoli Wang received the Ph.D. degree in information and communication engineering from Xi'an Jiaotong University, Xi'an, China, in 2012.

He is currently a Professor with CCTEG Xi'an Research Institute, Xi'an. His research interests include seismic data processing and imaging.

Dawei Liu received the bachelor's degree in communication engineering from Chang'an University, Xi'an, China, in 2013, the master's degree in electronic and communication engineering from Xi'an Jiaotong University (XJTU), Xi'an, in 2018, and the Ph.D. degree from the School of Information and Communication Engineering, Xi'an jiaotong University, in 2022.

He was a Post-Doctoral Scholar at Purdue University, West Lafayette, IN, USA, from 2022 to 2023. He is a currently a Post-Doctoral Fellow with the University of Alberta, Edmonton, AB, Canada. His research interests include tensor decomposition, deep learning, and time–frequency analysis and their applications in the seismic data processing.

Dr. Liu is a reviewer for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *Geophysics*, *Acta Geophysica*, and *Petroleum Science*.

Wenchao Chen (Member, IEEE) received the B.S. and M.S. degrees in seismic exploration and information technology from Chang'an University, Xi'an, China, in 1993 and 1996, respectively, and the Ph.D. degree in electromagnetic field and microwave technology from Xi'an Jiaotong University, Xi'an, in 2000.

From 2000 to 2002, he was a Post-Doctoral Fellow at the Department of Computation Science, Northwestern Polytechnical University, Xi'an. Since 2002, he has been with the Institute of Wave and Information, Xi'an Jiaotong University, where he is currently a Professor. His research interests include seismic and GPR signal processing, sparse representation, and ground penetrating radar.

Dr. Chen has served as a member of SEG and the Chinese Geophysical Society. He was also a referee for several journals, including IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *Geophysics*, *Interpretation*, and *Journal of Applied Geophysics*.