RESEARCH PAPER

Influence of Varying Operational Parameters on the Defect Detection Performance of a High-Speed Ultrasonic Rail Inspection System During Field Tests

D. Datta[1] · **F. Lanza di Scalea**[1]

Abstract

Background Continuous monitoring is essential for detecting internal defects in rails and prevent derailment related accidents. Existing techniques do not facilitate continuous monitoring because they require specialized test cars and can only operate at speeds of up to 30 mph.

Objective The objective of this study is to evaluate the performance of a high-speed rail inspection system using a non-contact ultrasonic technique with the potential of operating at train revenue speeds.

Methods The technique utilizes air-coupled transducers that record the ultrasonic guided waves generated by the rail-wheel contact and does not require a controlled acoustic source of excitation. A modified version of the traditional Welch's periodogram technique is utilized to extract the Green's function between two points on the rail. The passively extracted Green's function is then analysed statistically to detect structural discontinuities (e.g., defects) in the rail.

Results Results from fields tests performed at the Transportation Technology Centre (TTC) in Pueblo, CO, USA, demonstrate possible test speeds as high as 80 mph. From these field tests, the performance of the system is evaluated using Receiver Operating Characteristic (ROC) curves for a range of different operational parameters including test speed, location of the sensors relative to the locomotive (source), signal-to-noise ratio (SNR) of the raw signals, SNR of the reconstructed transfer function, baseline distribution length in the statistical analysis, wheel-rail interactions, and redundancies introduced from multiple runs over the same track.

Conclusions This study presents the current stage of development and performance of the passive rail inspection system with full-scale experiments under field conditions. The results indicate the potential of the system to operate at high speeds as well as possible avenues of future improvement to the system.

Keywords Rail inspection · Output-only identification · Ultrasonic guided waves · Receiver operating characteristic curves · Green's function · Ultrasonic inspection

Introduction

Internal defects in rails are a major cause of train derailments around the world. In a recent study, the US Federal Railroad Administration (FRA) reported 4,961 train accidents during 2017-2020, of which 3,437 (69.2% of total accidents) were caused by derailments [1]. Internal defects in rails are one

of the primary causes of train derailments. The three major types of internal rail defects are [2]: (a) Transverse Fissure (TF), a fracture originating in the interior of the railhead marked by the presence of a nucleus; (b) Detail Fracture (DF), a progressive fracture originating at or near the surface of the railhead without a nucleus; and (c) Vertical Split Head (VSH), a crack propagating vertically through or near the middle of the railhead. TF and DF are commonly classified as Transverse Defects (TD), because of their orientation. TDs are the most common type of rail defects and typically account for the majority of train derailments [3].

The detection of these internal flaws is an essential task of railroad maintenance operations. Ultrasonic systems are commonly utilized for this purpose, most commonly using

F. Lanza di Scalea is a Member of SEM.

✉ F. Lanza di Scalea
flanza@ucsd.edu

1 Jacobs School of Engineering, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0085, USA

fluid-filled wheels (Rolling Search Units or RSUs) that host a series of piezoelectric transducers operated in a pulse-echo or pitch-catch mode [5–7]. While quite effective in covering the majority of the rail cross-section, RSUs operate at speeds (~ 30 mph) that are substantially lower than revenue speeds, hence requiring careful scheduling among normal train traffic operations.

Various research has been conducted to develop technologies that can potentially inspect the rail at higher speeds. High-speed inspections usually require, among other things, the ability to probe the rail in a non-contact manner. Wooh et al. [8] explored a high-speed rail inspection technique using air-coupled acoustic transducers based on the Doppler effect. Their technique, however, was limited to detecting surface-breaking cracks. Mandriota et al. [9] introduced a filter-based image processing technique for detecting rail defects. Their technique, again, was limited to surface defects visible to the camera. Other investigations on non-contact inspection techniques were based on Electro-Magnetic Acoustic Transducers (EMATs) [4, 10–12]. This technique requires a small lift-off distance of the transducers from the rail surface and hefty magnets to reach the required sensitivity. Ultrasonic non-contact techniques have been also studied utilizing either hybrid laser/air-coupled approaches [13–15] or completely air-coupled approaches [16]. These systems remain limited by the requirement for an active "pulsed" ultrasonic excitation.

More recently, UCSD researchers [17–19] have investigated the possibility for non-contact rail inspection without using an active ultrasonic source, but rather exploiting the natural acoustic excitations imparted by the rolling wheels of a travelling train. This is therefore a case of "passive" or "output-only" inspection that utilizes solely ultrasonic receivers. These receivers are air-coupled transducers that stay above the rail clearance envelope (3 in) for non-contact probing. The sensors are installed on a travelling train car enabling a new concept of "smart train." If successfully developed, this capability would (a) enable rail inspections at regular (revenue) speeds without traffic disruptions, and (b) maximize the Probability Of Detection (POD) while minimizing the Probability of False Alarms (PFA) by exploiting the redundancy afforded by the multiple train passes over the same section of rail. This technology uses concepts of passive reconstruction of a system's Green's function (or transfer function) that have been developed in various fields, including seismology, underwater acoustics, and also structural inspections [20–26]. Examples include ambient vibrations in bridges induced by traffic [27–30], aerodynamic vibration signatures from aircraft wings and wind-turbine blades [31, 32], wind-induced vibrations in high-rise buildings [33], among others.

This paper presents the latest performance evaluation of a high-speed rail inspection prototype based on this "passive" approach that was tested at the Transportation Technology Center (TTC) in Pueblo, CO, USA at speeds up to 80 mph. In particular, the paper presents Receiver Operating Characteristic (ROC) curves that quantify the ability of the system to detect rail discontinuities (welds, joints and TDs) in terms of POD versus PFA [16, 19, 33] with varying operational parameters. These parameters include: the length of the baseline distribution utilized in the statistical signal processing, the speed of the test run, the type of the wheel-rail interaction, the location of the transducer array with respect to the locomotive, the SNR of the reconstructed transfer function, and the number of test runs (redundancy). These studies build the foundations for future improvements of this system.

Background

"Passive" Transfer Function Extraction (Dual-Output System)

Consider the schematic in Fig. 1, showing a rail track dynamically excited by a rolling wheel W, and the responses measured by two air-coupled ultrasonic receivers at locations A and B. Assume that both receivers are only sensitive to waves propagating uni-directionally from left to right. The aim is to isolate the transfer function of the test structure (rail) between location A and location B which is denoted by $G_{AB}(f)$. The wheel excitation $W(f)$ is unknown, uncontrolled, and assumed to be piecewise-stationary, meaning that its statistics do not change during the observation time windows of $O_A(f)$ and $O_B(f)$ (as discussed later, the observation time windows are on the order of ~msec thus this assumption is reasonable). $WA(f)$ denotes the transfer function between the the wheel excitation and location A. Uncorrelated noise components $N_A(f)$ and $N_B(f)$ are also assumed to be present at each of the two outputs. Assuming all systems to be linear, the outputs at locations A and B with added noise can be written as:
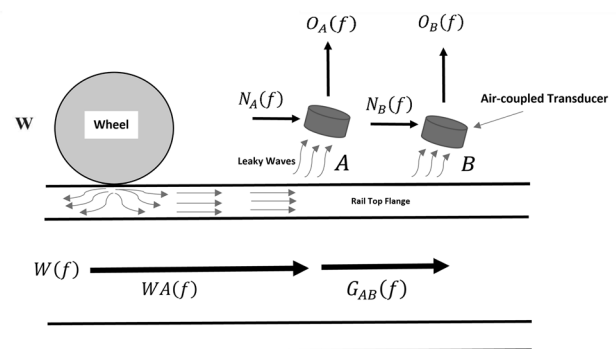

Fig. 1 Schematic diagram of passive transfer function reconstruction

$$O_A(f) = W(f).WA(f) + N_A(f) \quad receiver\,A \quad (1)$$

$$O_B(f) = W(f).WA(f).G_{AB}(f) \\ + N_B(f) \quad receiver\,B \quad (2)$$

The transfer function $G_{AB}(f)$ can be computed as a ratio of the cross-power spectrum between responses at A and B divided by the auto-power spectrum of the response at A. Both the cross-power spectrum and auto-power spectrum are computed in an ensemble average sense, by dividing the time-series into n segments, with a 50% overlap between segments to avoid loss of information near the ends of each segment [18]. A Hamming window is also applied to each time segment before computing the Fast Fourier Transform (FFT) which is a common practice in the signal processing domain to prevent side-lobe leakage [34]. Intra-segment averaging is used for computing the cross-power spectrum and inter-segment averaging is used to compute the auto-power spectrum for reasons that will be clear subsequently. Let us first compute the intra-segment cross-power spectrum between outputs at A and B as shown below:

$$\langle Cross\_Power \rangle_{intra-segment} = \langle O_{Ai}^*(f) \cdot O_{Bi}(f) \rangle = \langle |W(f)|^2 \cdot |WA(f)|^2 \cdot G_{AB}(f) \rangle + \\ \langle W^*(f) \cdot WA^*(f) \cdot N_B(f) \rangle + \langle W(f) \cdot WA(f) \cdot G_{AB}(f) \cdot N_A^*(f) \rangle + \\ \langle N_A^*(f) \cdot N_B(f) \rangle = |W(f)|^2 \cdot |WA(f)|^2 \cdot G_{AB}(f) \quad (3)$$

where * denotes the complex conjugate, $\langle \rangle$ denotes an ensemble average, | | denotes the absolute value, and i is an index for the different segments over which the averaging is done. The terms $\langle W^*(f) \cdot WA^*(f) \cdot N_B(f) \rangle$, $\langle W(f) \cdot WA(f) \cdot G_{AB}(f) \cdot N_A^*(f) \rangle$ and $\langle N_A^*(f) \cdot N_B(f) \rangle$ can be eliminated because the cross-power spectrum of uncorrelated signals (with no DC bias component) tends to zero in an averaged sense. Since the same segment (i) in responses at A and B is used, this averaging is termed as intra-segment averaging. Assuming the process to be ergodic, we can express the time-averaged cross-power spectrum as follows:

$$\langle Cross\_Power \rangle_{intra-segment} = \frac{1}{n} \sum_{i=1}^{n} O_{A,i}^*(f) \cdot O_{B,i}(f) \\ = |W(f)|^2 \cdot |WA(f)|^2 \cdot G_{AB}(f) \quad (4)$$

where n is the total number of segments. It is clear that the cross-power spectrum alone does not isolate the transfer function $G_{AB}(f)$ since it is 'colored' by the spectrum of the wheel-induced excitation $|W(f)|^2$ and the transfer function $|WA(f)|^2$ between the excitation source and transducer A. The cross-power spectrum is therefore normalized by the auto-power spectrum of the response at A. Let us first compute the auto-power spectrum using the same intra-segment averaging as discussed above. The auto-power spectrum at A can be written as:

$$\langle Auto\_Power_A \rangle_{intra-segment} = \langle O_{Ai}^*(f) \cdot O_{Ai}(f) \rangle = \langle |W(f)|^2 \cdot |WA(f)|^2 \rangle \\ + \langle W^*(f) \cdot WA^*(f) \cdot N_A(f) \rangle \\ + \langle W(f) \cdot WA(f) \cdot N_A^*(f) \rangle + \langle N_A^*(f) \cdot N_A(f) \rangle \\ = |W(f)|^2 \cdot |WA(f)|^2 + |N_A(f)|^2 \quad (5)$$

Again, assuming ergodicity, (equation (5)) may be rewritten as a time-average over n segments:

$$\langle Auto\_Power \rangle_{intra-segment} = \frac{1}{n} \sum_{i=1}^{n} O_{A,i}^*(f) \cdot O_{A,i}(f) \\ = |W(f)|^2 \cdot |WA(f)|^2 + |N_A(f)|^2 \quad (6)$$

The normalized cross-power spectrum can be obtained as:

$$\frac{\langle Cross\_Power \rangle}{\langle Auto\_Power \rangle} = \frac{\frac{1}{n} \sum_{i=1}^{n} O_{A,i}^*(f) \cdot O_{B,i}(f)}{\frac{1}{n} \sum_{i=1}^{n} O_{A,i}^*(f) \cdot O_{A,i}(f)} \\ = \frac{|W(f)|^2 \cdot |WA(f)|^2 \cdot G_{AB}(f)}{|W(f)|^2 \cdot |WA(f)|^2 + |N_A(f)|^2} \quad (7)$$

From (equation (7)), it is clear that the transfer function $G_{AB}(f)$ can not be isolated if the noise term $|N_A(f)|^2$ is non-zero. The term $|N_A(f)|^2$ is non-zero because the auto-power spectrum of an uncorrelated signal cannot be eliminated if it is taken over the same time segment. This problem can be resolved by averaging the same time signal over different segments (inter-segment averaging). This strategy eliminates the noise term $|N_A(f)|^2$. The inter-segment auto-power spectrum for response at A is computed as:

$$\langle Auto\_Power_A \rangle_{inter-segment} = \frac{1}{\bar{n}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} O_{A,i}^*(f) \cdot O_{A,j}(f) \# \quad (8)$$

where i, j are indices for different segments of the same signal and $\bar{n} = {}^nC_2 = \frac{n!}{2(n-2)!}$ is the number of the possible combinations of two different segments for a total of n segments. Substituting $O_A(f)$ from (equation (1)) into (equation (8)), the auto-power spectrum can be written as:

$$\langle Auto\_Power_A \rangle_{inter-segment} = |W(f)|^2 \cdot |WA(f)|^2 \\ + \frac{1}{\bar{n}} \sum_{i,j} N_{A,i}^*(f) \cdot N_{A,j}(f) \quad (9) \\ = |W(f)|^2 \cdot |WA(f)|^2$$

The noise term ($N_{A,i}^*(f) \cdot N_{A,j}(f)$) is eliminated because the cross-power spectrum of uncorrelated noise averaged over different segments tends to zero. Inter-segment averaging, however, introduces another problem. Strictly speaking, (equation (9)) is accurate only if the signals $W(f)$ and $WA(f)$ are correlated in both amplitude and phase over the different possible combinations of segments. This assumption is not true since conformity in phase cannot be assured for

different segments. A more reasonable assumption would be that $W(f)$ and $WA(f)$ are correlated in amplitude, but not phase, among different segments. This problem is handled by forcing different segments to be in the same phase. Let us rewrite the response at sensor A by separating the output into signal and noise components in (equation (1)) as:

$$O_A(f) = W(f) \cdot WA(f) + N_A(f) = S_A(f) + N_A(f) \tag{10}$$

where $S_A(f)$ includes the correlated signal at A and $N_A(f)$ is the uncorrelated noise. Assuming $S_A(f)$ to be time-invariant during the observation window, each inter-segment cross-power spectrum will therefore have an amplitude that is consistent and a phase that is random. Analytically:

$$
\begin{aligned}
\langle S_A^*(f) \cdot S_A(f) \rangle_{inter-segment} &= \frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} S_{A,i}^*(f) \cdot S_{A,j}(f) \\
&= \frac{1}{n} \sum_{i,j}^{n} |S_{A,i}(f)| e^{-i\phi_{A,i}} \cdot |S_{A,j}(f)| e^{i\phi_{A,j}} \\
&= \frac{1}{n} \sum_{i,j}^{n} |S_{A,i}(f)| \cdot |S_{A,j}(f)| e^{i\Delta\phi_{A,ij}}
\end{aligned}
\tag{11}
$$

The signals in each segment are shifted appropriately in such a way that their phases are aligned in all segments and therefore phase correlation is enforced in addition to amplitude correlation. The appropriate time-lag of shift for each segment pair is determined by the maximum peak of the cross-correlation function between the two segments. This time-lag for each segment pair is computed as:

$$\tau_{ij} = argmax\left(\int_{-\infty}^{+\infty} O_{A,i}^*(t).O_{A,j}(t+\tau)dt \right) \tag{12}$$

The final expression for the time-shifted, inter-segment averaged auto-power spectrum is written as:

$$
\begin{aligned}
\langle Auto\_Power \rangle_{\substack{inter-segment \\ shifted}} &= \frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} O_{A,i}^*(f) \cdot O_{A,j}(f).e^{-i2\pi f \tau_{ij}} \\
&= |S_A(f)|^2 + \frac{1}{n} \sum_{i,j} N_{A,i}^*(f) \cdot N_{A,j}(f) \\
&= |S_A(f)|^2 = |W(f)|^2.|WA(f)|^2
\end{aligned}
\tag{13}
$$

This technique efficiently removes the noise term since $\frac{1}{n}\sum_{i,j} N_{A,i}^*(f) \cdot N_{A,j}(f) = 0$ for uncorrelated noise. The normalized cross-power spectrum computed using this novel inter-segment auto-power spectrum successfully isolates the transfer function $G_{AB}(f)$ of the system as shown below:
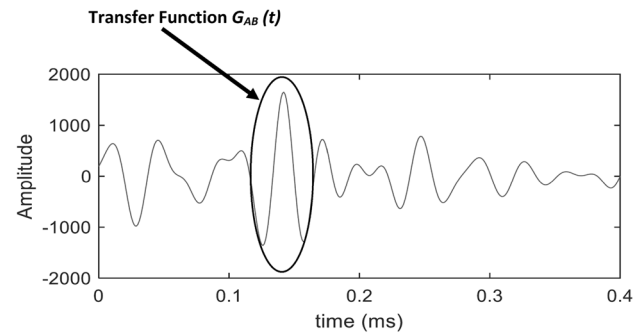

Fig. 2 A sample passively reconstructed transfer function in time domain

$$
\begin{aligned}
\frac{\langle Cross\_Power \rangle_{intra-segment}}{\langle Auto\_Power \rangle_{\substack{inter-segment \\ shifted}}} &= \frac{\frac{1}{n}\sum_{i=1}^{n} O_{A,i}^*(f) \cdot O_{B,i}(f)}{\frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} O_{A,i}^*(f) \cdot O_{A,j}(f).e^{-i2\pi f \tau_{ij}}} \\
&= \frac{|W(f)|^2 \cdot |WA(f)|^2 \cdot G_{AB}(f)}{|W(f)|^2 \cdot |WA(f)|^2} = G_{AB}(f)
\end{aligned}
\tag{14}
$$

The time-domain transfer function (impulse-response function) can be then retrieved from the frequency domain through an inverse Fourier Transform:

$$G_{AB}(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} G_{AB}(f).e^{ift}df \tag{15}$$

In the rail inspection prototype to be discussed, the transfer function is filtered in the frequency bands of 20 kHz - 40 kHz and 70 kHz – 120 kHz. These ranges were found to reconstruct a stable waveform during the tests. Figure 2 shows a typical reconstructed transfer function in time domain ($G_{AB}(t)$). The wave packet arrival at ~ 160 μs indicates the time taken by the wave to travel the distance from point A to B (Fig. 1) which is around 460 mm (18 inches).

Statistical Outlier Analysis

The transfer function so obtained defines a system governed by the properties of the rail segment between the two transducers. Discontinuities present in the rail between A and B (e.g. joints, welds and defects) would induce wave scattering hence alter the transfer function. Features from the transfer function (e.g. amplitude) can therefore be tracked statistically for changes along the length of the track. Outliers of these features with respect to a baseline distribution of features can then be related to the rail discontinuities.

Accordingly, for the prototype a statistical Damage Index (DI) was computed as the Mahalanobis Squared Distance [35] of the features of the transfer function with respect to a baseline distribution of features. The most general DI metric is defined below in a multivariate sense:

$$D.I. = (x - \overline{x})^T \cdot Cov^{-1} \cdot (x - \overline{x}) \quad (16)$$

where x is the feature vector extracted from a given location, $\overline{x}$ is the mean of the feature vector from the baseline distribution, Cov is the covariance matrix of the baseline distribution, and T represents the matrix transpose operator. In the field tests that were conducted, the feature vector $\{x\}_{4x1}$ consisted of the metric $variance^{-1}$ of the transfer function from four possible combinations of sensor pairs (discussed later). Since the rail geometry and wheel-rail interactions can change along a track, the baseline distribution was computed adaptively by considering a limited number of locations collected right before the current location. The length of this adaptive baseline distribution is one operational parameter of the results shown later. Finally, an "exclusive" version of the baseline was adopted, whereby extreme values of the DI (i.e. values larger than mean + twice the standard deviation) were removed from the baseline computation. This removal ensured that only pristine portions of rail were included in the computation improving the detection of outliers.

Field Tests

Test Setup

A set of field tests of a "passive" inspection prototype were conducted at TTC, in Pueblo, CO, USA in December 2018, June 2019 and December 2019. The prototype consisted of 12 ultrasonic capacitive air-coupled transducers (*CAP-2* by *VN Instruments Inc.*) with a central frequency of 120 kHz and arranged as shown in Fig. 3. The transducers prototype beam was mounted on the equalizer beam of a test car and arranged in 3 groups with each group having 4 transducers. This configuration resulted in four possible combinations of transfer functions from each group. The transducers were positioned at 3 inches from the rail's top surface at an angle of 6° with the vertical based on Snell's Law [36] to ensure unidirectional reception of the waves leaking from the rail into the air. Shown in Fig. 4, a laser system consisting of two sensors was attached to both ends of the prototype to detect cases of misalignment during a run. A high-speed camera (up to 100 frames per second) was installed alongside the prototype to continuously capture images of the rail. The camera images were also used to construct the "ground truth" map of rail discontinuities, consisting of locations of joints, welds, and internal defects (that were marked with
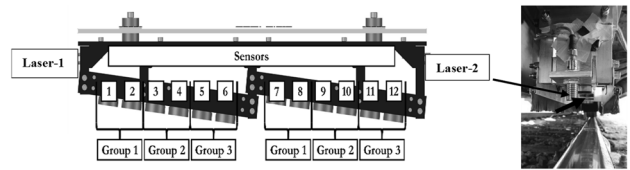

Fig. 3 Sensor arrangement in prototype with location of alignment lasers

paint). A GPS receiver was used to assign the signals and ground truth features to their specific locations on the track.

Test Methodology

The instrumented test car was towed by a locomotive. Tests runs were conducted on the High-Tonnage Loop (HTL) track and the Railroad Test Track (RTT) of TTC in 2018 and 2019. The difference between the field tests in 2018 and those in 2019 was that in the latter tests the prototype was placed closer to the locomotive wheels for improved signal strengths.

The HTL is a 2.7-mile-long test track that has numerous joints (about 22) and welds (about 275) since portions of the track are constantly replaced from damage due to heavy freight cars. At the time of the tests, the HTL also had three pre-identified defects (TDs) that were marked by spray-paint and were also identified by the camera for the ground truth. Speeds of 25 mph, 33 mph and 40 mph were tested on the HTL with three runs conducted at each speed. In addition, 12 continuously recorded runs were performed on the HTL at 40 mph and the results were compounded to study the role
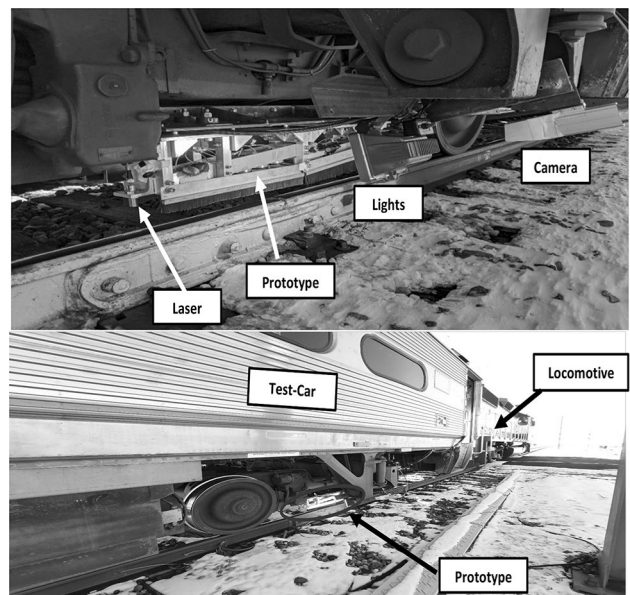

Fig. 4 Prototype and accessory hardware mounted on the test-car

of redundancy. Tests runs were also conducted on the RTT at even higher speeds of 60 mph, 70 mph and 80 mph with 3 runs conducted at each speed. The RTT is a 13.7-mile-long test track with 1801 welds and 45 joints based on the image-based ground truth library. The RTT had no known internal defects present at the time of the tests.

The sensor head was mounted to probe the inner rail for both the HTL and the RTT and the runs were conducted in a clockwise sense. Data was acquired continuously for the 3 runs at each speed without halting the train. Data recording was stopped at the end of the 3 runs at each speed and the train was brought back to the starting point to begin testing at a different speed. Different speeds were used to test the effects of varying excitation source strengths and their influence on the stability of the reconstructed transfer function. Real-time analysis was performed on a *National Instruments FPGA* module running on *LabView Real-time* platform. Extraction of the transfer function and computation of DI was performed in real-time, in tandem with data acquisition, as a quality control check on the acquired data.

Test Results

ROC Curves

The prototype performance was assessed with the help of ROC curves, similarly to previous evaluations of an "active" version of the non-contact rail inspection system [16]. The ROC curve plots the Probability Of Detection (POD) vs Probability of False Alarms (PFA) for different values of the DI threshold level. A high DI value indicates an outlier and may represent a possible rail discontinuity. A high value of DI in the vicinity of a known discontinuity would be a "true positive". Alternatively, a high value of DI in the vicinity of a pristine segment of track would be a "false positive". The POD gives an estimate of the "true positives" and is calculated by the equation below:

$$POD = \frac{D_i}{D_t} \qquad (17)$$

where D_i is the number of discontinuities detected during the test run and D_t is the total number of discontinuities present in the test track. Similarly, the PFA gives an estimate of the "false positives" and is computed by the equation:

$$PFA = \frac{D_p}{P_t} \qquad (18)$$

where D_p is the total number of discontinuities spuriously identified in pristine rail segments and P_t is the total number of pristine rail segments scanned. The ROC curves are computed by varying the DI threshold level such that

each threshold value corresponds to one point on the curve. A good performance is indicated by an ROC curve lying towards the top left of the graph, corresponding to high POD and low PFA values. To ensure robustness of the analysis, a minimum number of threshold crossings (7) was required within a fixed length of rail segment (18 inch) for a location to be flagged as a possible discontinuity. In order to compensate for the limited GPS resolution, a discontinuity search range of $\pm 10ft$ was adopted, meaning that any location flagged within the search range of a known discontinuity location was considered as a true detection. Any outliers flagged outside this search range of known discontinuities were considered as false positives. Figure 5 shows a sample DI trace along with the corresponding ROC curve obtained by varying the threshold levels. The next sections discuss the field test results obtained for varying operational parameters.

Signal-to-Noise Ratio of Raw Signals

The first parameter examined was the strength of the raw signals recorded by the air-coupled sensors from the wheel excitations with respect to the recordings' noise floor. When the train wheels do not acoustically excite the rails sufficiently, the signal in the sensors essentially consist of electronic and environmental noise. This happens, for example, when the train is moving at slow speeds.

The variance of the signal (σ_s^2) relative to the variance of the noise (σ_n^2) gives the signal-to-noise ratio (SNR) of the raw data and can be expressed in decibels (dB) as:

$$SNR_{raw}(dB) = 10log_{10}\left(\frac{\sigma_s^2}{\sigma_n^2}\right) \qquad (19)$$

It can be difficult to separate the raw data into signal and noise components because even at higher speeds, the noise component will be present. Hence, an approximate SNRwas calculated based on the assumption that any signal generated at speeds below 5 mph was noise, as shown below:

$$SNR_{raw-approx}(dB) = 10log_{10}\left(\frac{\sigma_{s>5mph}^2}{\sigma_{s<5mph}^2}\right) \qquad (20)$$

where $\sigma_{s<5mph}^2$ is the variance of the signal at speeds below 5 mph and $\sigma_{s>5mph}^2$ is the variance of the signal at speeds above 5 mph. Based on this SNR calculation, a cut-off dB level can be chosen. Track regions having signals above a threshold dB level can be classified as 'good' zones, whereas track regions having signals below the threshold dB level can be classified as 'bad' zones. As expected, the transfer function reconstructions and discontinuity detection performance in the 'good' zones was found more reliable compared to those from the 'bad' zones, as will be shown later. Figure 6 shows
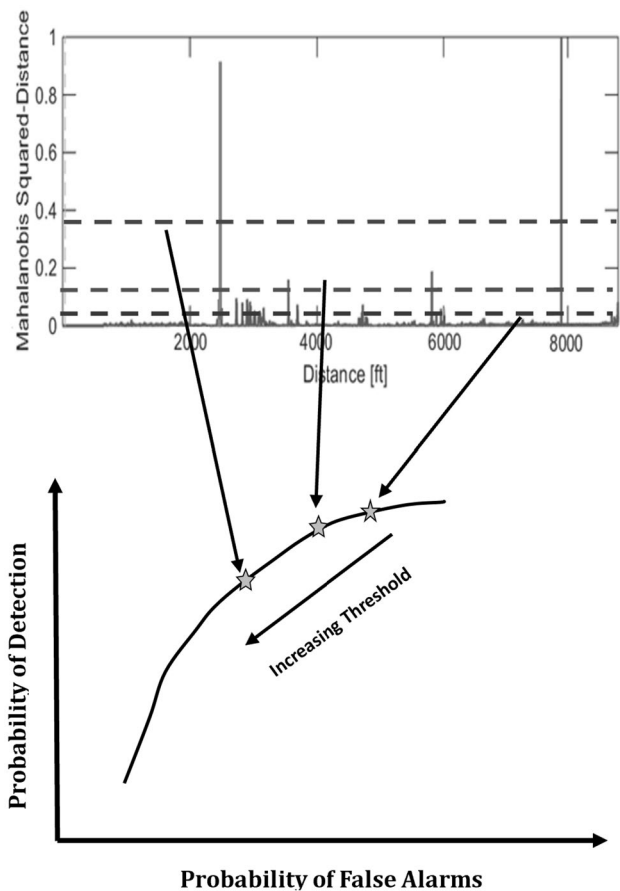
Fig. 5 ROC curve computation from varying DI thresholds

Fig. 6 Map of the HTL with regions of "high" and "low" acoustic signal strengths at different test speeds

high-frequency dominant tone which is independent of train speed [39]. Curve squeal, therefore, explains the high acoustic signal strengths at curved sections irrespective of train speed.

Figure 7(a) shows the ROC curves computed for the "joint" discontinuities at 40 mph comparing the entire run and good signal strength zones only (curves 1-4). The Area Under the Curve (AUC) is a measure of the system's overall performance for different thresholds, with a higher AUC value indicating better performance. Figure 7(b) shows the comparison of ROC curves of the "weld" discontinuities for the entire run selected from the SNR of the recordings. Clearly, the joint and weld detection performance of the system improves in regions with higher acoustic signal strengths, as indicated by the shift in the ROC curves towards the top-left and a corresponding increase in the AUC metric. Note that for the entire run considered, the joint detection accuracy is generally better than the weld detection. This is expected since joints are always expected to produce the most severe wave scattering (more pronounced outliers in the DI), whereas weld are expected to produce a limited wave scattering (and no wave scattering at all, at the wave frequencies considered, for a particularly "good" weld).

Note that the locations of all the three types of discontinuities (welds, joints, and internal defects) on the probed rail were already known. The library (ground truth) of these discontinuity locations was built by analyzing the camera images (of the rail) acquired during testing. The

the acoustic signal strengths based on a 6 dB SNR threshold for test runs at 40 mph and 25 mph on the HTL. From Fig. 6, it is evident that acoustic signal strength increases with increasing speed of the runs. Also, signal strengths seem to be consistently good in the curved sections of the track. Acoustic "noise" induced by the contact between the wheel flange and the rail gage [37, 38] corner in curved sections of the track could be one of the reasons for the increased recorded signal. However, since the inner rail of the HTL was probed, wheel flanging would only occur at curve 1 because the wheel on the inner rail has to travel a greater distance. Moreover, at shallow curves and lower speeds, wheel flanging may not occur at curve 1. This observation hence does not explain the higher signal strengths recorded at curve 1 for lower speeds (25 mph). Wheel flanging alone, therefore, cannot explain the high signal strengths observed in all the curves. It is possible that curve squeal could explain the high signal strengths at curves 2, 3 and 4. When a train manoeuvres a curve, the axle of the vehicle moves in a transverse direction which leads to a transverse slip between the wheel and the rail (lateral creep). This lateral creep induces a self-excited vibration with a single
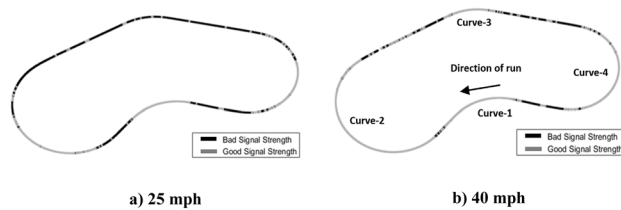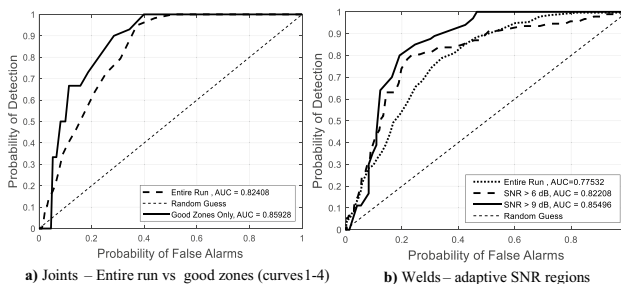

Fig. 7 ROC curves for joints and welds at 40 mph on the HTL track for different acoustic signal strength regions

known locations of the discontinuities were mapped through GPS coordinates to build a ground truth library. Locations flagged during data analysis as possible discontinuities were assigned to either joints, welds, or defects based on which known discontinuity was present in the vicinity of the flagged location. For example, if the system flagged a discontinuity in a rail segment and a known joint was present within a +-10 ft (search tolerance) of that flagged location, that detection was assigned a "true detection" for a joint. If there was no known discontinuity within the search tolerance limit, that detection was assigned a "false positive".

Achieving reliable rail excitation at lower speeds and tangent sections of the track is important to ensure sufficient signal-to-noise ratio. The use of a non-contact, controlled, and continuous acoustic source is currently being investigated. Another way of generating reliable excitations could be by mounting the prototype on a rail grinder vehicle. The grinding action is expected to impart significant acoustic energy to the rail potentially improving the signal-to-noise ratio.

Test Speed

Another important operational parameter for the prototype is the speed of the test run. Figure 8 shows the ROC curves for the three TDs present on the HTL track at different speeds. Interestingly, the passive defect detection improves significantly with the increase in speed due to higher acoustic signal strengths at high speeds. Best results were obtained at the speed of 40 mph where a 100 % detection rate (PD) was observed with a 17% possibility of false alarms (PFA). At 33 mph, the rate of detection drops to 67% (PD=67%) for the same rate of false alarms (PFA=17%). If the speed
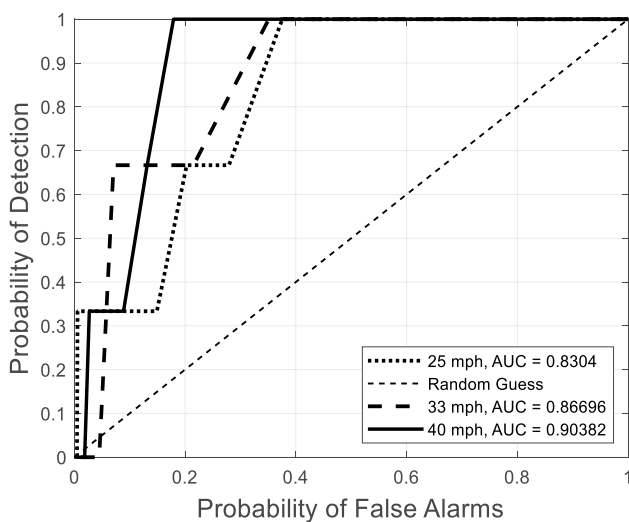
is lowered to 25 mph, the rate of detection further drops to 34%. Depending on the allowable rate of false alarms that can be tolerated, the DI threshold level can be selected that optimizes the detection performance of the system for a given set of operational parameters. The staggered nature of the ROC curves was because the PD was calculated with only 3 known location of defects which resulted in only 4 possible values (0, 1/3, 2/3, 1). The fact that speed seems to aid the performance is a comforting result since the objective is enabling inspections at revenue speeds. The sample size of 3 defects makes it difficult to draw any substantial conclusions and further tests need to be conducted on tracks with larger number of known defects to obtain statistically significant inferences. The PFA of 17% is still too high for industrial applications and needs further improvements.

Location of the Prototype

The location of the prototype with respect to the locomotive wheels was changed in two different sets of tests. Figure 9 shows the ROC curves for joints at 40 mph on the HTL for different locations of the prototype with respect to the locomotive. Placing the sensing array closer to the source of excitation led to a more stable transfer function reconstruction due to improved acoustic signal strengths, resulting in an improved detection performance.

Baseline Distribution Length

The length of the baseline in the outlier analysis plays a key role in the discontinuity detection performance. A longer baseline results in an increased number of points in computing the normal distribution which leads to a more averaged
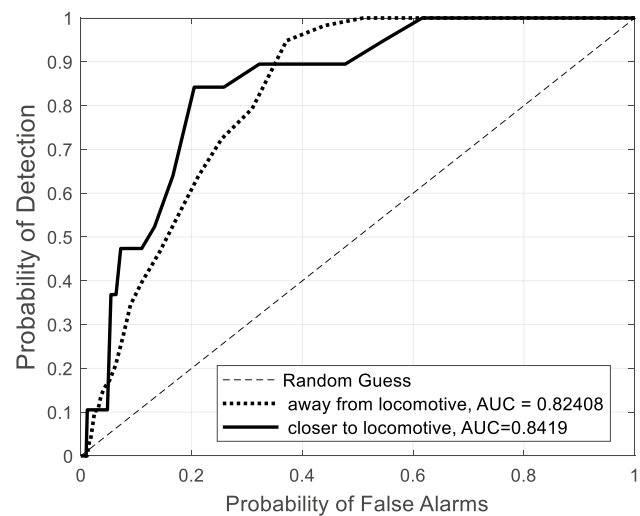

Fig. 8 ROC curves for defects (TD) at different speeds

Fig. 9 ROC curves for joints at 40 mph for different locations of the prototype
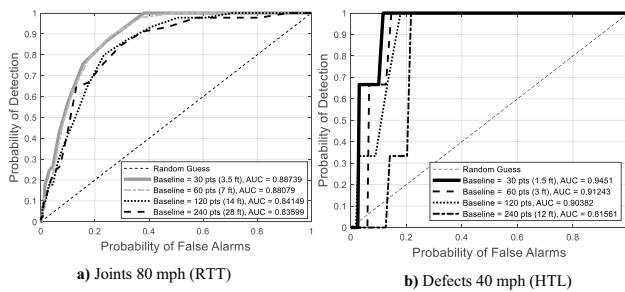
Fig. 10 ROC curves for joints and defects with different baseline distribution lengths

Fig. 11 Coherent signal and noise floor in the reconstructed transfer function

statistics of the rail. On the other hand, a shorter baseline results in lesser number of points in computing the normal distribution which leads to a more localized statistics of the rail. Therefore, a longer baseline is expected to result in a reduced sensitivity to discontinuity detection. On the other hands, a shorter baseline is expected to be more sensitive to discontinuities but may also lead to increased false alarms. The effects of changing the length of the baseline distribution was analysed with the help of ROC curves. Baseline distribution lengths of 30, 60, 120 and 240 points were analysed which correspond to approximately a physical distance of 3.5 ft, 7 ft, 14 ft and 28 ft respectively at 80 mph. The ROC curves of joints for different baselines at 80 mph testing speeds on the RTT are shown in Fig. 10(a). ROC curves of defects for different baselines at 40 mph on the HTL are shown in Fig. 10(b). From Fig. 10 it is evident that a reduction in the baseline distribution length increased the sensitivity of the system and improved the ROC curves as seen by an increase in the AUC. For example, for a certain DI threshold with 15% false alarms, a baseline distribution length of 240 points led to 34% detection rate for defects (Fig. 10(b)). Reducing the baseline length to 60 points led to a 100 % PD with 15% PFA.

SNR of Reconstructed Transfer Function

The defect detection performance of the passive rail inspection system depends on the quality of the reconstructed transfer function. A distinct wave arrival with minimum noise floor would be an ideal reconstruction. The relative amplitude of the arrival wave with respect to the noise floor is an indication of the quality of the transfer function (Fig. 11). The signal-to-noise ratio of the transfer function in dB is calculated as:

$$SNR_{TF}(dB) = 10 log_{10}\left(\frac{\sigma_t^2}{\sigma_n^2}\right) \tag{21}$$

where σ_t^2 is the variance of the transfer function within the arrival window and σ_n^2 is the variance of the noise outside the
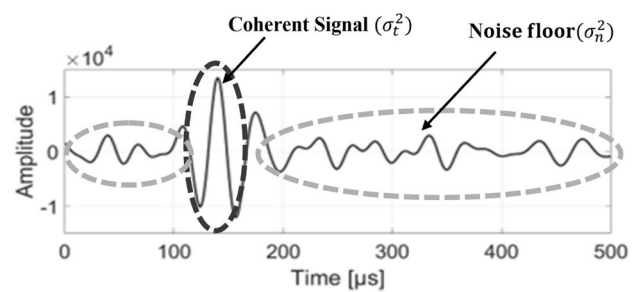
arrival window as shown in Fig. 11. Low SNR of the transfer function would indicate wave attenuation in the presence of discontinuities in the rail segment and can be used to predict defect locations. Figure 12 shows the predictions made by the passive inspection system based on a combination of DI values and the SNR of transfer functions from different pairs of group-1 sensors. Black diamonds indicate all four pairs of the transfer functions have SNR less than 3 dB. Red diamonds indicate 3 pairs of transfer functions have SNR less than 3 dB. Yellow diamonds indicate 2 pairs of transfer functions have SNR less than 3 dB and green diamonds indicate either one or none of the pairs have SNRs less than 3 dB. Black, red, yellow and green diamonds plotted are all above a threshold of 0.02% (2e-5) of the maximum DI value in the trace. Blue asterisks, orange triangles, yellow stars are the locations of the welds, joints and defects respectively picked up by the camera and represent the ground truth. Finally, the cyan asterisks represent the regions where the SNR of raw signals falls below 6 dB. Black, red and yellow diamonds represent the locations where the passive system predicts some form of discontinuity (welds, joints or defects). When these diamonds (black, red or yellow) align with any of the plotted ground truth (blue asterisk-welds, orange diamond-joints, yellow star-defects) a true detection can be assigned. When these diamonds occur in regions where there are no known discontinuities, a false alarm is raised. Zoomed views of a region of high signal strength (Zone-A) and a region of low signal strength (Zone-B) are shown in Fig. 12. Low signal strength zones have a comparatively higher rate of false alarms compared to the high signal strength zones.

Redundancies from Multiple Runs

The results discussed so far in this paper for the discontinuity detection performance of the passive inspection system are only for 1 test run considered at a time. In practical situations, when the prototype will be used for defect detection by mounting it on a revenue train, the same section of the track will be traversed by the train multiple times, resulting in a larger set of observations. Multiple observations on the
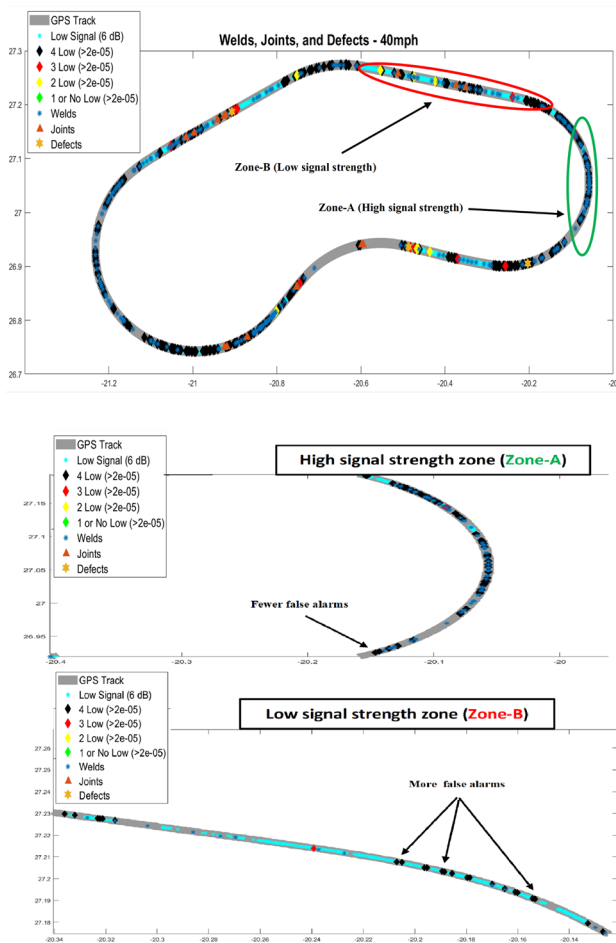
Fig. 12 GPS based map of passive system predictions, ground truth and signal strengths

Fig. 13 ROC curves for welds at 40mph with redundancies for multiple runs

same track are expected to cause redundancies in the data with consequent reduction of false alarms. This expectation is based on the assumption that false positives will occur at randomly distributed locations along the track, whereas true positives will occur at consistent locations for every run. Locations flagged as possible discontinuities in different runs based on the DI trace can then be overlaid on top of each other and non-coinciding points may be discarded as false positives, whereas coinciding locations could be tagged as true detections.

Figure 13 shows the results obtained on the 12-run dataset at 40 mph on the HTL for welds. If 33% or more of the number of runs detected a discontinuity at the same location within a search range ($\pm 10 ft$), that location was flagged as a positive detection. Therefore, the number of runs required to assign a positive detection was 1 out of 3 runs, 2 out of 6 runs and 3 out of 12 runs. The results obtained for 1-run and 3-run cases do not introduce any redundancies. The 3-run case still performs better than the 1-run case because of the additional odds of detecting the weld in at least 1 out
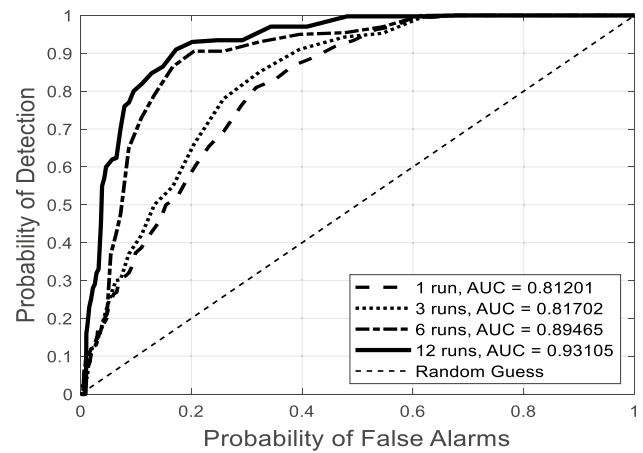
of 3 runs. The 3-run case, however, also increases the rate of false positives simultaneously and therefore a significant improvement is not achieved. The 6-run case improves the performance significantly because now redundancies are introduced, and the false positive detections need to align in at least two of the runs which is more unlikely. In general, it is observed that as the number of runs is increased, the ROC curve shifts towards the top-left indicating the expected reduction in the rate of false alarms. The results also do not improve as much when we compare the 6-run case with the 12-run case (sort of a saturation effect). This suggests that the system's performance tends to a maximum limit as the number of runs are increased. It is also worth noting that not all the false positives are removed by compounding multiple runs. This indicates the system keeps flagging similar locations in multiple runs where possibly unmarked defects or discontinuities are present. A more accurate ground truth is likely to convert many of the false positives to true positives.

Conclusions

This paper discusses the current state of the high-speed rail inspection system under development at the University of California San Diego on behalf of the Federal Railroad Administration. The system uses a passive ultrasonic sensing approach that utilizes non-contact air-coupled ultrasonic receivers and special signal processing algorithms to flag locations of discontinuities along a rail. The key potential advantages of this technology are: (1) the possibility to inspect the rail at regular train speeds (2) the possibility to enhance the detection performance due to the run redundancies. The field test performance is presented in terms of ROC curves quantifying the trade-off between PD and PFA

for various rail discontinuities (joints, welds, defects) and for different operational parameters. The SNR of the raw signal has an obvious effect on the performance, indicating "good" and "bad" portions of rail depending on the rail-wheel interaction conditions. This is a complex problem that needs to be investigated further. Higher test speeds were found to yield better performance of the system because of the higher energy and higher bandwidth levels introduced into the rail by the wheel excitations, which improved the signal-to-noise ratio of the passively reconstructed ultrasonic transfer function of the rail. Improved performance at higher speeds is very encouraging since one of the primary objectives of this project is enabling rail inspections at regular train speeds ("smart train" concept). The tests also show that the location of the sensors should be as close as possible to the locomotive wheels for enhanced excitation strength. In terms of statistical analysis, it is found that the length of the baseline distribution can affect the performance outcome, with an optimum length resulting in the best performance. Finally, it is found that multiple runs on the same track improve the performance, with the improvement saturating when a certain number of runs is reached (six in our tests).

Defects and welds will cause scattering of the travelling waves that allows their detection. Joints should result in total signal loss (total reflection). Note that all defects presented and identified in the manuscript are transverse defects. Transverse defects belong to the category of defects where the cracks primarily grow in a plane perpendicular to the train running direction in the transverse direction (parallel to the railroad ties). The proposed technique utilizes guided wave modes traveling in the longitudinal direction along the rail and therefore these waves are particularly sensitive to the transverse defects. Since guided waves are formed by constructive interference of bulk modes propagating in the transverse direction of the waveguide (rail), we have found in previous investigations that these modes are sufficiently scattered also by Vertical Split Head defects, enabling their detection (although with a smaller sensitivity compared to transverse-type defects).

Although this work has laid strong foundations for this approach, additional research and development efforts are needed to make a successful transition to industry. Further studies need to be conducted to investigate ways of improving signal strengths at lower speeds on straight sections of the track. One possible way to achieve this is to introduce a controlled and broadband excitation source, such as continuous impacts on the rail, using an automatically controlled hammer. Another technique is to use high powered acoustic horns for non-contact excitations. The fact that, in these cases, multiple excitation sources may co-exist should not be a problem for the Green's function reconstruction, as long as there is sufficient energy at the frequency band of interest. In other words, all multiple sources will superimpose to give

the excitation $W(f)$ in Fig. 1. The normalized cross-power spectrum described in (equation (14)) eliminates the effect of $W(f)$, regardless of the individual excitation sources present. The only requirement is that there is sufficient excitation energy in the frequency band of interest.

Further research also needs to be conducted to improve and optimize the signal feature vectors used in the outlier analysis with machine learning based techniques.

Finally, at this stage of research, the location of the defect is identifiable. Once the location is identified, other methods (e.g. hand-held verification) can be used for size determination.

Data Availability Data can be made available upon request.

Declarations

Conflict of Interests The author(s) hereby declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Federal Railroad Administration (FRA) Office of Safety Analysis (2020). Train Accidents and Rates. https://safetydata.fra.dot.gov/OfficeofSafety/default.aspx. Accessed Nov. 8, 2020

2. Track Inspector Rail Defect Reference Manual FRA https://railroads.dot.gov/elibrary/track-inspector-rail-defect-reference-manual. Accessed Nov. 8, 2020

3. Zakar F, Mueller E (2016) Investigation of a Columbus, Ohio train derailment caused by fractured rail. Case Stud Eng Fail Anal 7:41–49

4. Alers GA (1988) Railroad rail flaw detection system based on electromagnetic acoustic transducers. OTIFRA/ORD-88/09, U.S. Dept. of Transportation, Washington, DC

5. Berger H (1980) Nondestructive testing of railroad rail. Transportation Research Record 744:22–26

6. Anon F (1990) Rail-flaw detection. A science that works. Railway Track and Structures 86(5):30–32

7. Lanza di Scalea F (2007) Ultrasonic Testing Applications in The Railroad Industry (Special Applications of Ultrasonic Testing), in Non-Destructive Testing Handbook., 3rd Ed., American Society for Nondestructive Testing 535–552

8. Wooh S-C, Kim CC, Wei C (1999) Real-Time Processing of Continuous Doppler Signals for High-Speed Monitoring of Rail Tracks. Rev Prog Quant Nondestruct Eval 2245–2252

9. Mandriota C, Nitti M, Ancona N, Stella E, Distante A (2004) Filter-based feature selection for rail defect detection. Mach Vis Appl 15415:179–185

10. Alers GA, Manzanares A (1990) Use of Surface Skimming SH Waves to Measure Thermal and Residual Stresses in Installed Railroad Tracks. Rev Prog Quant Nondestruct Eval 1757–1764

11. Palmer SB, Dixon S, Edwards RS, Jian X (2005) Transverse and longitudinal crack detection in the head of rail tracks using Rayleigh wave-like wideband guided ultrasonic waves. Nondestructive Evaluation and Health Monitoring of Aerospace Materials, Composites, and Civil Infrastructure 5767:70–80

12. Rose JL, Lee CM, Hay TR, Cho Y, Park IK (2006) Rail inspection with guided waves. Proc., 12th Asia-Pacific Conf. on NDT

13. Coccia S, Phillips RR, Bartoli I, Salamone S, Rizzo P, Lanza di Scalea F (2012) On-line high-speed rail defect detection. Part II. Rep. No. DOT/FRA/ORD-12/02, Dept. of Transportation, Washington, DC

14. Coccia S, Bartoli I, Marzani A, Lanza di Scalea F, Salamone S, Fateh M (2011) Numerical and experimental study of guided waves for detection of defects in the rail head. NDT E Int 44:93–100

15. Rizzo P, Coccia S, Bartoli I, Lanza di Scalea F (2008) Noncontact rail monitoring by ultrasonic guided waves. Encycl Struct Heal Monit. https://doi.org/10.1002/9780470061626.SHM041

16. Mariani S, Lanza di Scalea F (2017) Predictions of defect detection performance of air-coupled ultrasonic rail inspection system. Struct Health Monitor. https://doi.org/10.1177/1475921717715429

17. Lanza di Scalea F, Zhu X, Capriotti M, Liang AY, Mariani S, Sternini S (2018) Passive extraction of dynamic transfer function from arbitrary ambient excitations: Application to high-speed rail inspection from wheel-generated waves. J Nondestruct Eval Diagnostics Progn Eng Syst. https://doi.org/10.1115/1.4037517/365839

18. Lanza di Scalea F, Sternini S, Liang AY (2018) Robust passive reconstruction of dynamic transfer function in dual-output systems. J Acoust Soc Am 143:1019

19. Datta D, Liang A, Cui R, Lanza di Scalea F (2020) Defect detection performance of a high-speed rail inspection system from passive acoustic identification. SPIE 11379:196–204. https://doi.org/10.1117/122558205

20. Gouédard P, Roux P, Campillo M, Verdel A (2008). Convergence of the two-point correlation function toward the Green's function in the context of a seismic-prospecting data set. https://doi.org/10.1190/1.2985822

21. Snieder R (2004) Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase. Phys Rev E 69:046610

22. Tippmann JD, Lanza di Scalea F (2014) Passive-only damage detection by reciprocity of Green's functions reconstructed from diffuse acoustic fields with application to wind turbine blades. J Intel Mater Syst Struct 26:1251–1258. https://doi.org/10.1177/1045389X14538539

23. Yang Y, Xiao L, Qu W, Lu Y (2017) Passive detection and localization of fatigue cracking in aluminum plates using Green's function reconstruction from ambient noise. Ultrasonics 81:187–195

24. Michaels JE, Michaels TE (2005) Detection of structural damage from the local temporal coherence of diffuse ultrasonic signals. IEEE Trans Ultrason Ferroelectr Freq Control 52:1769–1782

25. Duroux A, Sabra KG, Ayers J, Ruzzene M (2010) Extracting guided waves from cross-correlations of elastic diffuse fields: Applications to remote structural health monitoring. J Acoust Soc Am 127:204

26. Chehami L, De Rosny J, Prada C, Moulin E, Assaad J (2015) Experimental study of passive defect localization in plates using ambient noise. IEEE Trans Ultrason Ferroelectr Freq Control 62:1544–1553

27. Farrar CR, James GH (1997) System identification from ambient vibration measurements on a bridge. J Sound Vib 205:1–18

28. Mustafa S, Debnath N, Dutta A (2015) Bayesian probabilistic approach for model updating and damage detection for a large truss bridge. Int J Steel Struct 15:473–485

29. Salvermoser J, Hadziioannou C, Stähler SC (2015) Structural monitoring of a highway bridge using passive noise recordings from street traffic. J Acoust Soc Am 138:3864

30. Sabra KG, Winkel ES, Bourgoyne DA, Elbing BR, Ceccio SL, Perlin M, Dowling DR (2007) Using cross correlations of turbulent flow-induced ambient vibrations to estimate the structural impulse response. Application to structural health monitoring). J Acoust Soc Am 121:1987

31. Tippmann JD, Zhu X, Lanza di Scalea F (2015) Application of damage detection methods using passive reconstruction of impulse response functions. Philos Trans R Soc A Math Phys Eng Sci. https://doi.org/10.1098/RSTA.2014.0070

32. Ubertini F, Comanducci G, Cavalagli N (2016) Vibration-based structural health monitoring of a historic bell-tower using output-only measurements and multivariate statistical analysis 15:438–457. https://doi.org/10.1177/1475921716643948

33. Mariani S, Nguyen T, Zhu X, Lanza di Scalea F (2017) Field test performance of noncontact ultrasonic rail inspection system. J Transp Eng Part A Syst 143:04017007

34. Harris FJ (1978) On the use of windows for harmonic analysis with the discrete fourier transform. Proc IEEE 66:51–83

35. Yeager M, Gregory B, Key C, Todd M (2018) On using robust Mahalanobis distance estimations for feature discrimination in a damage detection scenario. 18:245–253. https://doi.org/10.1177/1475921717748878

36. Drosdoff D, Widom A (2005) Snell's law from an elementary particle viewpoint. Am J Phys 73:973

37. Kim JC, Yun YS, Noh HM (2019) Analysis of Wheel Squeal and Flanging on Curved Railway Tracks. Int J Precis Eng Manuf 20:2077–2087

38. Ho TK, Liu SY, Ho YT, Ho KH, Wong KK, Lee KY, Tam HY, Ho SL (2008) Signature analysis on Wheel-rail interaction for rail defect detection. IET Semin Dig. https://doi.org/10.1049/IC:20080342

39. Thompson D (2009) Curve Squeal Noise. Railw Noise Vib 315–342