U.S. Department
of Transportation

**Federal Railroad
Administration**

On-line High-speed Rail Defect Detection

Office of Research and
Development
Washington, D.C. 20590

DOT/FRA/ORD-04/16

October 2004
Final Report - Part 1

NOTICE

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United State Government assumes no liability for its contents or use thereof.

NOTICE

The United States Government does not endorse products or manufacturers. Trade or manufacturer's names appear herein solely because they are considered essential to the objective of this report.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
October 2004	Final Report - Part 1	August 2003 - August 2004

4. TITLE AND SUBTITLE
On-line High-speed Rail Defect Detection

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DTFR53-02-G-00011

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
F. Lanza di Scalea, I. Bartoli, P. Rizzo, and J. McNamara

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California, San Diego
Department of Structural Engineering
9500 Gilman Drive, M.C. 0085
La Jolla, CA 92093-0085

8. PERFORMING ORGANIZATION REPORT NUMBER
UCSD SSRP-2004/10

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Department of Transportation
Federal Railroad Administration
Office of Research and Development
1120 Vermont Avenue NW
Washington, D.C. 20590

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
DOT/FRA/ORD-04/16

12. DISTRIBUTION/AVAILABILITY STATEMENT
This document is available to the public through the National Technical Information Service, Springfield, VA 22161

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report presents the results of phase 2 of the project "On-line high-speed rail defect detection" aimed at improving the reliability and the speed of current defect detection in rails. Ultrasonic guided waves, traveling in the rail running direction, were examined for this task. The advantages of this approach include: the inherent sensitivity to transverse-type cracks even in the presence of surface shells , the potential for extremely high testing speed, and the ease of field deployment since the sensors can be positioned at distances as far as 3.5 inches from the rail surface. Two methods of implementation were investigated. The first method used an impact hammer for wave generation and air-coupled sensors for wave detection. This setup proved effective for detecting "large" (>15% H.A.) transverse cracks at frequencies below 50 kHz. The second method used a pulsed laser for wave generation. This setup proved effective for detecting "small" (<15% H.A.) transverse surface-breaking cracks as shallow as 1 mm at frequencies between 100 kHz and 600 kHz. The inspection ranges (distance covered in one test) were substantially larger than current methods.

15. SUBJECT TERMS
Transverse cracks, ultrasonic guided waves, non-contact ultrasonic testing, laser ultrasonics, air-coupled sensors, continuous Wavelet transform, discrete Wavelet transform, signal de-noising, Finite Element Modeling, automatic defect classification

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Francesco Lanza di Scalea
Unclassified	Unclassified	Unclassified	UU	110	19b. TELEPHONE NUMBER *(Include area code)* 858-8221458

Reset

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

PREFACE

This report presents an ultrasonic method for defect detection in rails that substantially differs from conventional ultrasonic rail inspection. The main difference is in the use of ultrasonic guided waves traveling along the rail running direction. These waves have the potential for improved reliability in detecting transverse cracks and for improved inspection speed when compared to conventional rail inspections. Extensive experimental and numerical tests have been conducted to demonstrate the capabilities of guided waves for the detection of both "small" (<15% H.A.) and "large (>15% H.A.) transverse cracks in the rail head. Advanced signal processing based on wavelet transforms and automatic pattern recognition have been implemented in the testing. Finally, impact hammers, pulsed lasers and air-coupled sensors have been investigated as practical ways to perform the testing in the field.

This work has been performed by the University of California, San Diego under university research grant DTFR53-02-G-00011 awarded by the Federal Railroad Administration of the U.S. Department of Transportation. The work was sponsored by FRA's Office of Research and Development. The authors are grateful to Mahmood Fateh, the FRA's Technical Representative, and Dr. David Jeong of the Volpe National Transportation Systems Center at Cambridge, MA for their comments on the draft version of the report, based on which revisions have been made.

METRIC/ENGLISH CONVERSION FACTORS

ENGLISH TO METRIC

LENGTH (APPROXIMATE)

1 inch (in)	=	2.5 centimeters (cm)
1 foot (ft)	=	30 centimeters (cm)
1 yard (yd)	=	0.9 meter (m)
1 mile (mi)	=	1.6 kilometers (km)

AREA (APPROXIMATE)

1 square inch (sq in, in^2)	=	6.5 square centimeters (cm^2)
1 square foot (sq ft, ft^2)	=	0.09 square meter (m^2)
1 square yard (sq yd, yd^2)	=	0.8 square meter (m^2)
1 square mile (sq mi, mi^2)	=	2.6 square kilometers (km^2)
1 acre = 0.4 hectare (he)	=	4,000 square meters (m^2)

MASS - WEIGHT (APPROXIMATE)

1 ounce (oz)	=	28 grams (gm)
1 pound (lb)	=	0.45 kilogram (kg)
1 short ton = 2,000 pounds (lb)	=	0.9 tonne (t)

VOLUME (APPROXIMATE)

1 teaspoon (tsp)	=	5 milliliters (ml)
1 tablespoon (tbsp)	=	15 milliliters (ml)
1 fluid ounce (fl oz)	=	30 milliliters (ml)
1 cup (c)	=	0.24 liter (l)
1 pint (pt)	=	0.47 liter (l)
1 quart (qt)	=	0.96 liter (l)
1 gallon (gal)	=	3.8 liters (l)
1 cubic foot (cu ft, ft^3)	=	0.03 cubic meter (m^3)
1 cubic yard (cu yd, yd^3)	=	0.76 cubic meter (m^3)

TEMPERATURE (EXACT)

[(x-32)(5/9)] °F = y °C

METRIC TO ENGLISH

LENGTH (APPROXIMATE)

1 millimeter (mm)	=	0.04 inch (in)
1 centimeter (cm)	=	0.4 inch (in)
1 meter (m)	=	3.3 feet (ft)
1 meter (m)	=	1.1 yards (yd)
1 kilometer (km)	=	0.6 mile (mi)

AREA (APPROXIMATE)

1 square centimeter (cm^2)	=	0.16 square inch (sq in, in^2)
1 square meter (m^2)	=	1.2 square yards (sq yd, yd^2)
1 square kilometer (km^2)	=	0.4 square mile (sq mi, mi^2)
10,000 square meters (m^2)	=	1 hectare (ha) = 2.5 acres

MASS - WEIGHT (APPROXIMATE)

1 gram (gm)	=	0.036 ounce (oz)
1 kilogram (kg)	=	2.2 pounds (lb)
1 tonne (t)	=	1,000 kilograms (kg)
	=	1.1 short tons

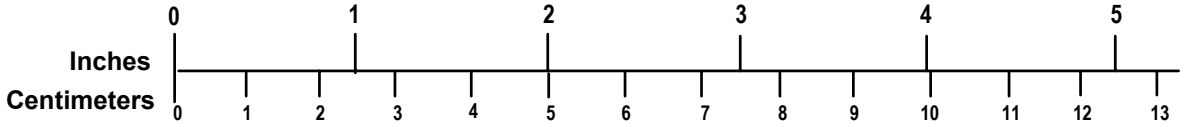
VOLUME (APPROXIMATE)

1 milliliter (ml)	=	0.03 fluid ounce (fl oz)
1 liter (l)	=	2.1 pints (pt)
1 liter (l)	=	1.06 quarts (qt)
1 liter (l)	=	0.26 gallon (gal)
1 cubic meter (m^3)	=	36 cubic feet (cu ft, ft^3)
1 cubic meter (m^3)	=	1.3 cubic yards (cu yd, yd^3)

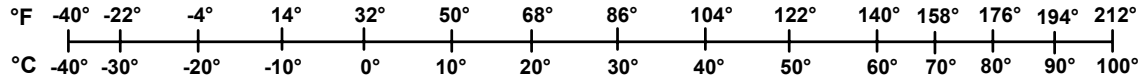
TEMPERATURE (EXACT)

[(9/5) y + 32] °C = x °F

QUICK INCH - CENTIMETER LENGTH CONVERSION



QUICK FAHRENHEIT - CELSIUS TEMPERATURE CONVERSION



For more exact and or other conversion factors, see NIST Miscellaneous Publication 286, Units of Weights and Measures. Price $2.50 SD Catalog No. C13 10286

Updated 6/17/98

TABLE OF CONTENTS

Section **Page**

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

1. Motivation for Research And Summary of Phase I Activities

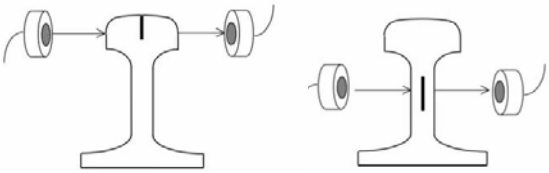
The methods of rail defect detection used today include magnetic induction testing and ultrasonic testing, with the second approach being more widely employed. Magnetic induction exploits the perturbations in the rail magnetic field induced by the presence of geometrical discontinuities such as cracks. Contact brushes are required to induce an electric current. The disadvantages of the magnetic induction method include the contact requirement for the brushes, the elevated sensitivity to joint switches and other structural elements of the rail, and the limited inspection coverage.

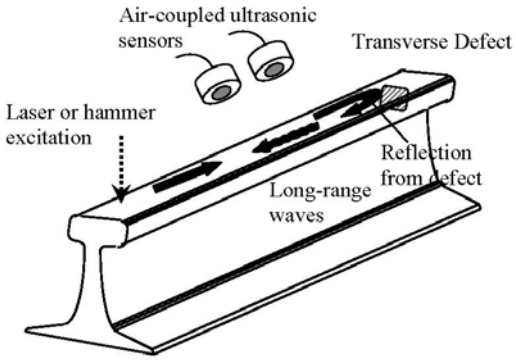
Conventional ultrasonic testing of rails uses wheels or sleds filled with water that host an array of piezoelectric sensors. The sensors are typically operated from the top of the rail head in a pulse-echo mode with orientations at zero degrees (normal incidence) and at 70 degrees from the normal to the running surface. The zero degree orientation targets horizontal defects and the 70 degree orientation targets transverse defects that tend to grow along a 20 degree orientation. The disadvantages associated with these methods include the requirement for bulky wheels and the contact conditions. Both factors limit inspection speed and area coverage. For example, in the U.S. rail inspections are typically carried out at speeds of 10 – 20 mph with a "stop and confirm" mode. These values are substantially lower that the speed of regular passenger and freight trains that easily reach 70 mph. The contact conditions are sensitive to environmental changes (such as temperature fluctuations) that can affect the output of the test. Also, certain areas of the rail are not accessible by the inspection including the critical head flanges. Finally, the presence of shallow horizontal cracks near the surface of the rail head prevents the ultrasonic beams from reaching the target internal defects. Hence the limited defect detection reliability of current ultrasonic rail inspections. According to the FRA Safety Data Statistics (FRA 1992-2002), there were 2,553 rail caused incidents during the period 1992-2002, accounting for about 25% of the total track caused accidents/incidents. It is thus clear that conventional rail defect detection methods are inadequate. Unfortunately, rail safety concerns can only become more serious given the unavoidable aging of the transportation infrastructure and the increasing rail tonnage. In response to the FRA's Strategic 5-Year R&D Plan to address railroad track safety, the University of California at San Diego (UCSD) is performing research in the area of Rail Defect Detection under grant DTFR53-02-G-00011 titled "On-line High-speed Rail Defect Detection." The objectives of this work are 1) **to increase the defect detection reliability** of current rail inspection methods, and 2) **to increase the overall testing speed** of current rail inspection method. UCSD's research activities in the area of rail defect detection started in 2000 under a research grant from the National Science Foundation. San Diego Trolley, Inc. is actively participating to the work by providing UCSD with in-kind support in terms of donations of rail sections for laboratory testing and access to rail lines for field testing of the technologies being developed in the laboratory.

Phase I of grant DTFR53-02-G-00011 (08/01/02-08/31/03) was completed in August 2003 and a draft report was submitted to FRA's Office of Safety in July 2003 (Lanza di Scalea 2003). The present report summarizes the results of Phase II work (09/01/03-08/31/04). A summary of activities and major findings of Phase I work follows.

In Phase I two new methods based on ultrasonic stress waves were examined for defect detection. These technologies consisted in 1) a high-frequency (~MHz) method probing the rail in its cross-sectional plane ("Cross-Sectional Method"), and 2) and a low-frequency (~kHz), method probing the rail along its longitudinal, running direction ("Long-Range Method"). Both methods made use of extensive signal processing developed under the general environment of the Wavelet Transform.

In the **Cross-Sectional Method** (Fig. 1-left column), air-coupled ultrasonic transducers were used to excite and detect the ultrasonic waves. It was found that by exploiting the rail cross-sectional geometry, resonance conditions could be achieved with the use of proper excitation frequencies. These resonance conditions made possible to probe the rail in a non-contact fashion at the expenses of signal detection sensitivity. Ultrasonic testing in this method is done across the rail and thus requires scanning of the sensors along the running direction similarly to current ultrasonic rail inspections. The method proved successful for the detection of longitudinal internal defects in the head, web and head flange of 115-lb American Railway Engineering and Maintenance of Way Association (A.R.E.M.A.) rails tested in the laboratory. However, theoretical estimates indicated that the maximum inspection speeds achievable by the Cross-Sectional Method are lower than the speed range of conventional rail testing cars, although substantial speed improvements could be achieved by heavier use of signal processing for de-noising the ultrasonic signals thereby reducing the number of necessary signal averages in a signal test. In addition, the necessity for positioning the sensors on both sides of the rail, rather than above the rail head, constitutes a severe limit for field deployment since the rail sides are not always accessible.





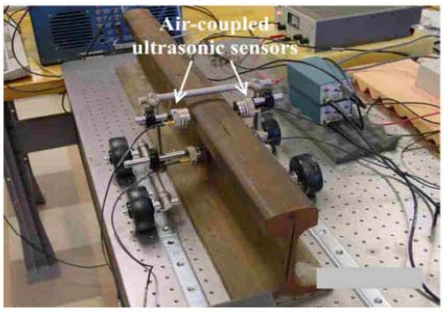


Figure 1 – Rail defect detection: the Cross-Sectional Method (left column) and the Long-Range Method (right column).

In the **Long-Range Method** (Fig. 1-right column), ultrasonic waves traveling in the rail longitudinal, running direction are employed. These waves are "guided" by the rail geometry and thus propagate as "guided waves." Laboratory and field testing of 115-lb A.R.E.M.A. rails were performed to characterize speed and attenuation of three primary guided-wave modes, namely the longitudinal, the lateral and the vertical modes. The interaction of these modes with rail defects were then examined in terms of wave reflection coefficients as a function of the wave frequency. The method proved successful in detecting transverse head defects of various sizes and orientations including transverse surface cracks as shallow as 1 mm from the head surface. These tests were carried out by using a combination of impulse hammers, pulsed lasers and air-coupled ultrasonic sensors. The primary advantages of the Long-Range Method that emerged from Phase I study are 1) increased reliability of detection of transverse-type defects, even in the presence of shallow, horizontal surface cracks that often mask critical internal defects in conventional inspections, 2) potential for extremely high testing speeds, since guided waves propagate at the speed of sound in steel (~5,000 mph), 3) extended rail coverage in a single test, since several feet of rail can be inspected at once, and 4) ease of field deployment since sensors can be positioned above the rail head and as far as 3.5 inches from the rail surface thus meeting the recommended clearance envelope for rail inspection systems. Table 1 summarizes the overall performance of the two methods investigated in Phase I.

Table 1 – Overall performance of the Cross-Sectional Method and the Long-Range Method for rail defect detection.

	Testing Speed*	**Types of Defects Targeted**	**Smallest Detectable Defect**	**Ability for Real-time Quantitative Inspection**	**Ease of Field Deployment**
Cross-Sectional Inspection	Low/Moderate (10 km/h or 6 mph)	Longitudinal	Internal head or web crack 20 mm in planar extent	High w/ signal processing	Low (sensors located on rail sides)
Long-range Inspection	Very High (8,000 km/h or 5,000 mph)	Transverse	Internal head crack 15% of head area Surface head crack 1 mm in depth	High w/ signal processing	Very High (sensors located above rail head at distances > 3 in)

*assumes the most unfavorable conditions of one ultrasonic transmitter and one ultrasonic receiver and slowest-traveling wave mode

Based on Phase I results, research activities in Phase II were aimed at furthering the development of the Long-Range Method that showed the strongest potential for meeting the proposed program objectives of 1) increased defect detection reliability and 2) increased testing speed.

In the following, the definition of “large” defects refers to transverse cracks larger than 15% of the rail head cross-sectional area (HA) and that of “small” defects refers to transverse cracks smaller than 15% HA.

2. Detection of "Large" (> 15% HA) Transverse Defects in Rail Head

2.1 Finite Element Modeling

The aim of this portion of the work was to demonstrate the use of a commercial finite element package, ABAQUS EXPLICIT, to model the interaction of a broadband vertical bending mode with transverse-type head defects in rails as small as 15% of the head cross-sectional area. Reflection coefficient spectra in the 20 kHz – 45 kHz range were obtained for four different sizes and three different orientations of transverse head flaws. A preliminary study of Lamb waves in a free plate helped drawing modeling guidelines for the rail. These results have been submitted for publication (Bartoli et al. 2004).

It is known that the classical theories of Euler-Bernoulli and Timoshenko for rail vibrations are only accurate at frequencies below 0.5 kHz and 1.5 kHz, respectively. At higher frequencies these theories cannot account for the significant cross-sectional deformations of the rail (Thompson 1993 and 1997). Numerical approaches have been proposed to predict the acoustic and ultrasonic modal properties of rails as dispersion curves. Thompson (1993) used the finite element method to model transient vibrations in rails with beam and plate elements at frequencies as high as 6 kHz. A double Timoshenko beam model that allowed relative displacement between the head and the foot of the rail was used by Wu and Thompson (1999) to model vertical waves, again below 6 kHz.

In the area of rail defect detection by long-range ultrasonic waves, however, frequencies in the range 10 kHz – 50 kHz have shown the best promise (Wilcox et al. 2003; Cawley et al. 2003; Lanza di Scalea and McNamara 2003; McNamara 2003). In this case, complications arise from the multimode and dispersive character of high frequency guided waves in rails that sometimes makes the mode identification challenging. For example, at 50 kHz there are around 20 vibrational modes theoretically propagating in typical rails. Gavric (1994 and 1995) relaxed the complexity of finite element modeling of waves in acoustic waveguides including rails by proposing a discretization in the rail cross-sectional plane only. The three-dimensional problem was thus successfully treated as a bi-dimensional one with significant savings in computational efforts. The same bi-dimensional method was employed by Wilcox et al. (2002) where a cyclic symmetry condition was imposed in the wave propagation direction. With the cyclic symmetry approximation, Gavric's bi-dimensional method could be implemented in standard finite element programs. These techniques are very efficient to calculate the wave modal solutions in terms of dispersion curves and cross-sectional mode shapes.

However, predicting the interaction of the waves with structural defects generally requires a three-dimensional model. Sophisticated numerical methods have been used to predict reflections of ultrasonic waves from rail defects, including three-dimensional time marching models (Wilcox et al. 2003; Cawley et al. 2003) and adaptive mesh refinement models (Trivedi et al. 2004). The reflection coefficients of guided waves from defects are the basis of defect detection in rails by long-range inspection. Defects that develop transversely to the rail

running direction are notoriously the most dangerous in rails. According to the Federal Railroad Administration safety data for the decade 1992-2002, transverse defects were the first leading cause of track failures in the US with $91M in associated damage and repair costs (FRA 1992-2002). Reflection coefficients for longitudinal, lateral and vertical waves interacting with transverse head defects of four different sizes and three different orientations were recently determined experimentally by a broadband, impulse excitation of a 115lb A.R.E.M.A. rail (Lanza di Scalea and McNamara 2003; McNamara 2003). In these studies a joint time-frequency analysis based on the Wavelet transform was employed to analyze the defect reflections in the 10 kHz – 40 kHz frequency range.

2.1.1 Finite Element Modeling with Application to Lamb Waves in a Plate

The effectiveness of conventional finite element modeling of elastic waves propagating in structural components has been shown in the past. The case of Lamb waves in free plates is a classical example (Moser et al. 1999). The package used in the present study, ABAQUS EXPLICIT, uses an explicit integration based on a central difference method.

The stability of the numerical solution is dependent upon the temporal and the spatial resolution of the analysis. To avoid numerical instability ABAQUS EXPLICIT recommends a stability limit for the integration time step Δt equal to:

$$\Delta t = L_{min} / c_L \quad (1)$$

where L_{min} is the smallest dimension of the smallest finite element of the model and C_L is the bulk longitudinal wave velocity through the material. This limit represents the time of travel of a longitudinal wave across the element.

The maximum frequency of the dynamic problem, f_{max}, limits both the integration time step and the element size. A good rule is to use a minimum of 20 points per cycle at the highest frequency (Moser et al. 1999), that is:

$$\Delta t = 1 / (20\, f_{max}) \quad (2)$$

The size of the finite element, L_e, is typically derived from the smallest wavelength to be analyzed, λ_{min}. For a good spatial resolution 10 nodes per wavelength are normally required (Alleyne and Cawley 1991), although some studies (Moser et al. 1999) recommend a more stringent condition of 20 nodes per wavelength. The latter condition can be written as:

$$L_e = \lambda_{min} / 20 \quad (3)$$

A pilot study of Lamb waves propagating in a free, isotropic plate was conducted with the purpose of verifying the proposed limits of temporal and spatial resolution in ABAQUS EXPLICIT as well as to show the potential of the Wavelet transform to analyze dispersive waves in a broad frequency domain. This work was also aimed at identifying criteria to be extended to the case of wave propagation in railroad tracks.

The well known Rayleigh-Lamb solutions of Lamb waves are:

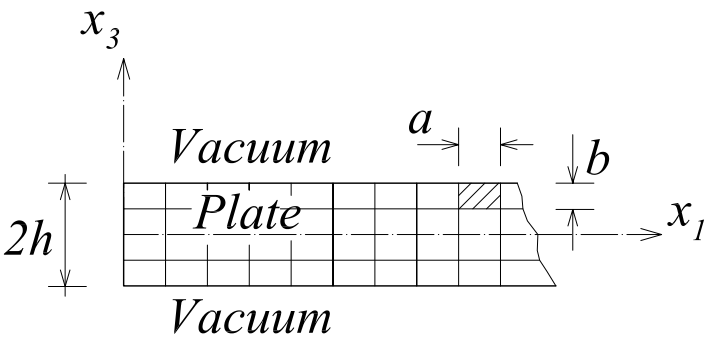
$$\frac{\tan(qh)}{\tan(ph)} = -\frac{4k^2 pq}{(q^2 - k^2)^2} \quad \text{for the symmetric modes, and} \tag{4}$$

$$\frac{\tan(qh)}{\tan(ph)} = -\frac{(q^2 - k^2)^2}{4k^2 pq} \quad \text{for the antisymmetric modes} \tag{5}$$

where k is the wavenumber and $2h$ is the plate thickness. The variables p and q are found from the circular frequency, ω, the wavenumber, k, and the bulk longitudinal and shear wave velocities, c_L and c_T, following:

$$p^2 = \left(\frac{\omega}{c_L}\right)^2 - k^2 \qquad \text{and} \qquad q^2 = \left(\frac{\omega}{c_T}\right)^2 - k^2 \tag{6}$$

The ABAQUS model of the plate is shown in Fig. 2.



	a (mm)	b (mm)
Mesh 1	10	10
Mesh 2	10	5
Mesh 3	5	5
Mesh 4	5	2
Mesh 5	2	2
Mesh 6	1	1

Figure 2 – ABAQUS model of the plate with meshes examined.

The material is steel, with the following nominal properties: Young's modulus E=209×10^9 N/m^2; Poisson's ratio ν=0.3; density ρ=7800 Kg/m^3; bulk longitudinal wave velocity c_L=5.98 km/sec; bulk shear wave velocity c_T=3.20 km/sec. Damping effects were neglected. A thickness of 20 mm was chosen for the plate based on the medium width of the web of the railroad track subsequently examined. In Fig. 2 the plate is infinite along x_2 so the problem is a plane strain one. A finite length of 3 m was considered for the model in the x_1 direction.

The plate was discretized by 4-node bilinear plane strain quadrilateral elements with two degrees of freedom per node. Both square and rectangular elements of various sizes were considered as schematized in the table of Fig. 2, where *a* and *b* are the element sizes along the longitudinal direction x_1 and the thickness direction x_3, respectively. The integration time step was set to Δt = 0.2 μsec throughout the analyses.

In the frequency range of interest (tens of kilohertz), only the zero-order symmetric mode, S_0, and anti-symmetric mode, A_0, are present. These two modes were selectively excited in the model by applying appropriate nodal displacements. A triangular forcing function $d(t)$ was applied as a uniform x_1-displacement distribution on the left edge of the plate to excite S_0 (Fig. 3a). A similar displacement distribution, only in the vertical direction x_3, was applied to the same nodes to excite A_0 (Fig. 3b). Typical results obtained by the model are shown in Fig. 3(a) and Fig. 3(b) for S_0 and A_0. The variable monitored was the vertical, x_3-displacement monitored at mid-span at the top surface of the plate (node 1). In Fig. 3(a) the reflection from the right-end of the plate is also visible in addition to the first arrival owing to the larger group velocity of S_0 when compared to A_0 at these frequencies.

The continuous Wavelet transform was used to perform the joint time-frequency analysis of the time signals. As a good compromise between time and frequency resolution, the Gabor wavelet transform (GWT) was employed with a center frequency of 2π and a shape factor of 5.336. GWT scalograms, representing the energy of the signal in the joint time-frequency domain, are shown in Fig. 3 for the corresponding time histories. The GWT analysis allows to extract the group velocity dispersion curves as well as the reflection coefficients from geometrical discontinuities in a broad frequency range at once (Lanza di Scalea and McNamara 2004). The analysis simply requires a single generation point and a single detection point. Group velocity dispersion curves for the plate model were calculated by comparing the scalograms of node 1 with those of node 2, located on the same top edge of the plate at a distance L = 750 mm from node 1. The frequency-dependent velocity was then obtained through the arrival times of the scalogram peaks for the two nodes following:

$$C_g(f) = \frac{L}{t_2(f) - t_1(f)} \tag{7}$$

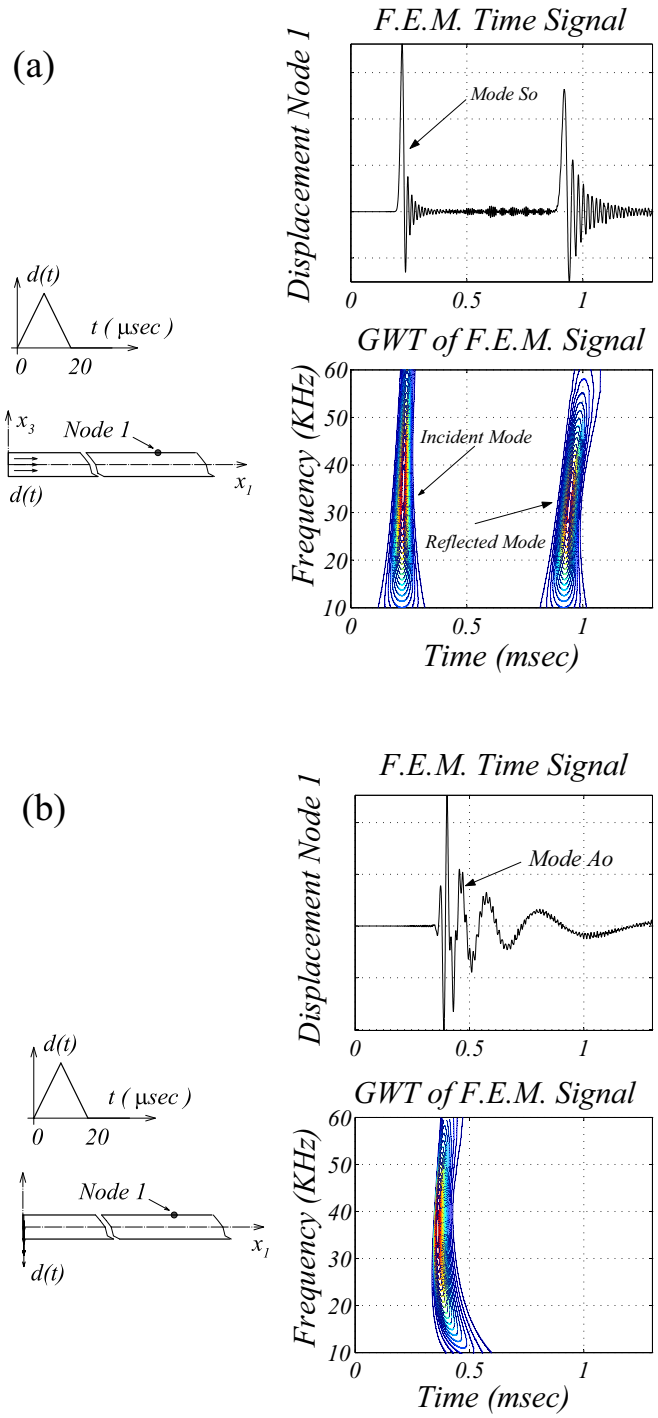


Figure 3 – ABAQUS simulations of Lamb waves in a plate excited by an impulse. (a) Case of the S_0 mode; (b) case of the A_0 mode.

The dispersion results from the finite element model are shown in Fig. 4 and compared to the exact Rayleigh-Lamb solutions from eqs. (4) and (5). In this figure the convergence of the model is shown for the different discretization meshes employed. The analysis was aimed at examining the effect of the element size both in the longitudinal direction x_1 and in the cross-sectional direction x_3. There is a lack of recommendations in the literature on the mesh refinement requirements along the cross-section of the waveguides. As the present study shows, an appropriate mesh refinement in the cross-section is as important as the mesh refinement in the longitudinal, wave traveling direction. The appropriate cross-sectional refinement is dependent on the mode shapes of the guided wave in a similar way as the longitudinal refinement is dependent on the wavelength from eq. (3). Fig. 4 shows that by doubling the refinement in the thickness direction (mesh 2 compared to mesh 1), there is an appreciable improvement in the group velocity prediction. By increasing the mesh refinement also in the longitudinal direction (mesh 3), a more substantial improvement is obtained with maximum errors below 3% of the exact solution. The increasingly refined meshes 4 through 6 coincided with the theoretical result in the frequency range examined.

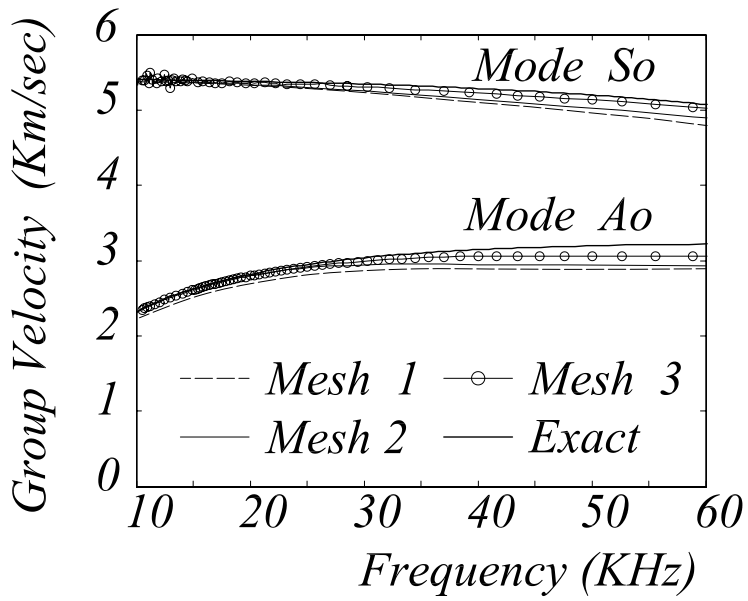


Figure 4 – Group velocity dispersion curves extracted from the model for varying mesh refinements. Plot also shows the Rayleigh-Lamb exact solution.

Results were also obtained in terms of energy reflection coefficients from the plate edge. Both a perpendicular edge and a 35 deg inclined edge were considered in analogy with the reflections from perfectly transverse and inclined defects in the rail. It is well known that for a perpendicular edge reflection no mode conversion is expected, meaning that a given incident mode (whether S_0 or A_0) would produce the same reflected mode below the cut-off frequencies of higher modes. An inclined reflector will instead produce mode-converted reflections necessary to satisfy stress-free boundary conditions. Energy reflection coefficients were obtained from the GWT scalogram ridges of the vertical displacements of node 1 in Fig. 3 for the first arrival and the edge reflection. To compensate for dispersion effects, the areas under the scalogram ridges were considered rather than the scalogram peaks. Reflection coefficient spectra were then calculated using:

$$R(f) = \frac{\int_{\Delta t_2} GWT_2(t, f)dt}{\int_{\Delta t_1} GWT_1(t, f)dt} \tag{8}$$

where $GWT(t, f)$ is the scalogram map, Δt is the entire duration of the pulse at frequency f, and the subscripts *1* and *2* refer to the first arrival and the edge reflection, respectively. The results are presented in Fig. 5 for the different mesh refinements and monitoring the x_1 longitudinal displacement in all cases. The case of mesh 6 (a=b=1mm) is here considered an exact solution. For the perpendicular edge, top of Fig. 5, the reflection coefficient for both S_0 and A_0 is one throughout the frequency range and no substantial effect of the mesh refinement can be seen. For the 35 degree inclined edge, mode conversions take place. The same-mode, S_0 reflection decreases with increasing frequency because the remaining energy is converted to A_0. A similar trend is observed for the A_0 mode incident on the inclined edge. Contrarily to the perpendicular reflection, the mesh refinement has a substantial effect on the inclined reflection in all cases. The reason for this is that it is more challenging to accurately represent mode converted modes including the non propagating (evanescent) modes that exist near the reflector (Torvik 1967). The importance of the mesh refinement in the cross-sectional direction, in addition to the longitudinal direction, is reaffirmed by comparing mesh 3 (a=b=5mm) to mesh 4 (a=5mm; b=2mm) for the oblique reflections. Convergence up to 60 kHz is obtained by further refining the longitudinal direction with mesh 5 (a= b=2mm). This suggests that an appropriate cross-sectional spatial resolution requires 10 nodes across the plate thickness, thus:

$$L_{e,cs} = d / 10 \tag{9}$$

where $L_{e,cs}$ is the cross-sectional size of the finite element and d is the plate thickness.

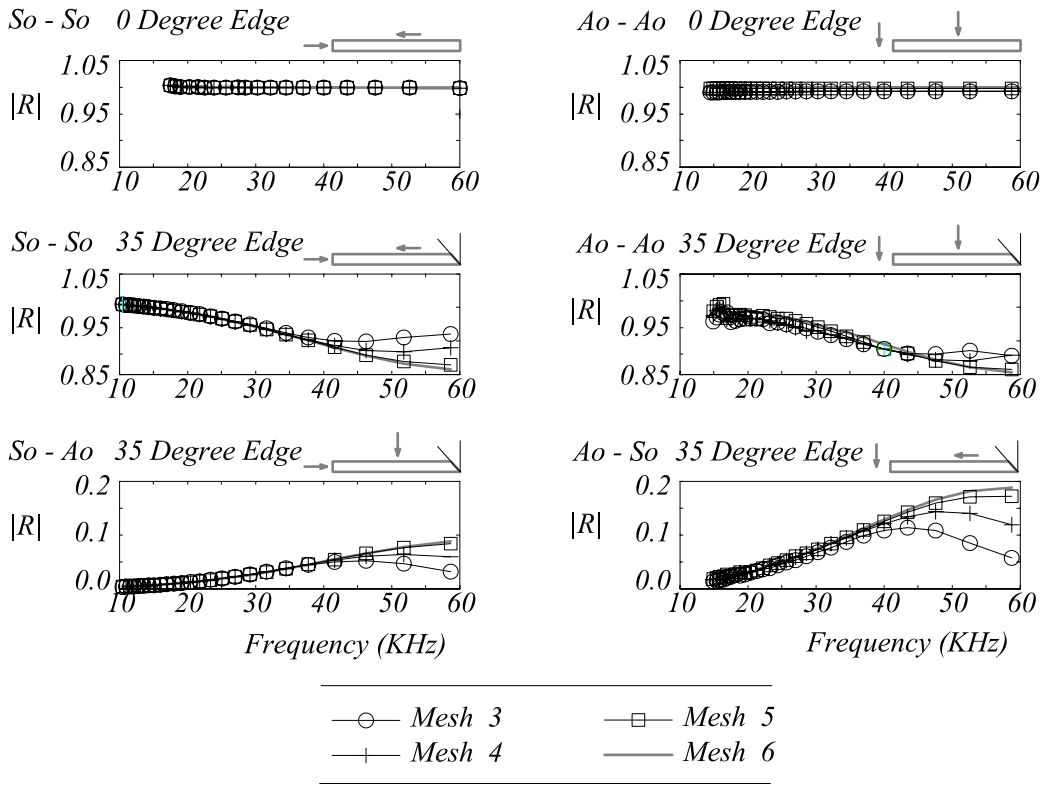


Figure 5 – Energy reflection coefficients in a plate from a perpendicular edge and from a 35 deg inclined edge.

The plate results confirm qualitatively the validity of the criteria expressed in eqs. (1)-(3). For a maximum frequency of 60 kHz, for example, the minimum wavelength is for A_0 and it is given by $\lambda_{min} = c / f_{max} = 40$ mm, considering a theoretical phase velocity of $c = 2.45$ km/sec. From eq. (3), the corresponding limit on the element size is $L_e = 2$ mm that corresponds to the size of the mesh 5 element. As for the temporal resolution, with an element size of 2 mm eq. (1) recommends a maximum integration time step of $\Delta t = 0.33$ μsec that is satisfied with the value of 0.2 μsec used for the analysis.

It should also be noted that if lower frequencies are of interest, these convergence criteria along the longitudinal direction can be relaxed. For example, taking 45 kHz as the highest frequency of interest (as in the rail application that follows), Fig. 5 shows that mesh 4 gives adequate results. In this case the largest error, in Fig. 5f, is within 8 % of the exact solution. Therefore, considering that $\lambda_{min} = 54.4$ mm for $f_{max} = 45$ kHz, ten nodes per wavelength in the longitudinal direction provide an acceptable accuracy. Mesh 4 still satisfies the cross-sectional criteria of eq. (9).

2.1.2 Finite Element Modeling with Application to Rail Defect Detection

The use of guided waves in rails, as discussed above, is of interest in the context of long-range defect detection. The potential is to detect defects as far away as tens or even hundreds of feet from the sensoring system. In this study the attention was focused on the vertical bending

vibrational mode that provides good defect sensitivity and it is easily generated and detected from the top of the rail. Access to the top of the rail only is required for any practical defect detection system that operates at high speeds in rails. The vertical bending mode can be generated by an impulse excitation at the top of the rail head in the vertical direction.

Both numerical and experimental analyses were conducted to study the reflection of the vertical bending mode from transverse-type defects in the rail head. These are among the most dangerous flaws in rails as mentioned above. Fig. 6 schematizes the 115 lb A.R.E.M.A. rail section that was studied. The elastodynamic properties of the rail steel material were the same as the ones assumed for the plate. Four defect sizes were studied numerically, herein indicated by the percentage of the head section that was cut, i.e. “15% defect”, “50% defect”, “85% defect”, and “100% defect.” For all four sizes, three different orientations were considered, namely a perfectly transverse direction, a direction inclined at 20 deg from the transverse direction, and one inclined at 35 deg from the transverse direction. In total, twelve cases of head defects were examined by the finite element analysis.

Results from the experimental testing are here shown only for the 100% defect (cut through the entire head) at the three orientations of transverse, 20% oblique and 35% oblique. Also, only same-mode reflections were measured. The experimental setup is shown in Fig. 7. The main novelty compared to the experimental work presented by Lanza di Scalea and McNamara (2003) and McNamara (2003) is the use of a high-frequency impulse hammer (PCB 086D80 mini hammer) that effectively excited frequencies as high as 50 kHz with appreciable energy. The hammer struck the top of the rail head at its left end along the vertical direction *y*. A typical time history of the hammer impulse is shown in Fig. 7. The same time history was used as a force input in the finite element analysis. An accelerometer (PCB 352C67) recorded the vertical bending mode at the top of the rail head at a distance L_2= 812.8 mm from the defect. The perfectly transverse 100% defect was located at L_1= 1828 mm from the impact location whereas the 100% inclined defects were located at L_1= 1574.8 mm from the impact location.

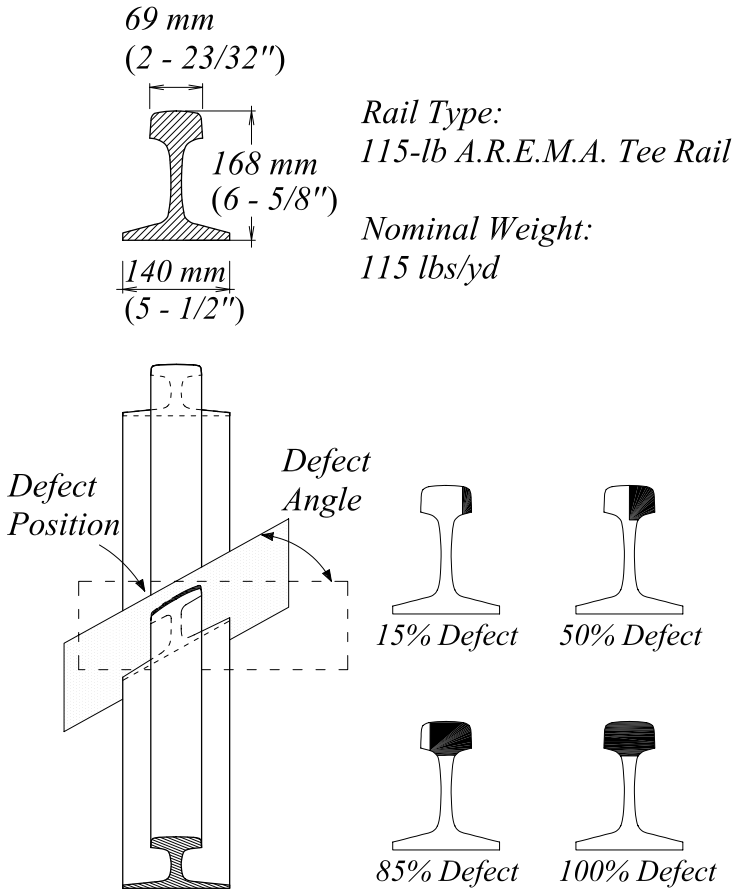


Figure 6 – Geometry of the rail studied with the different transverse head defects.

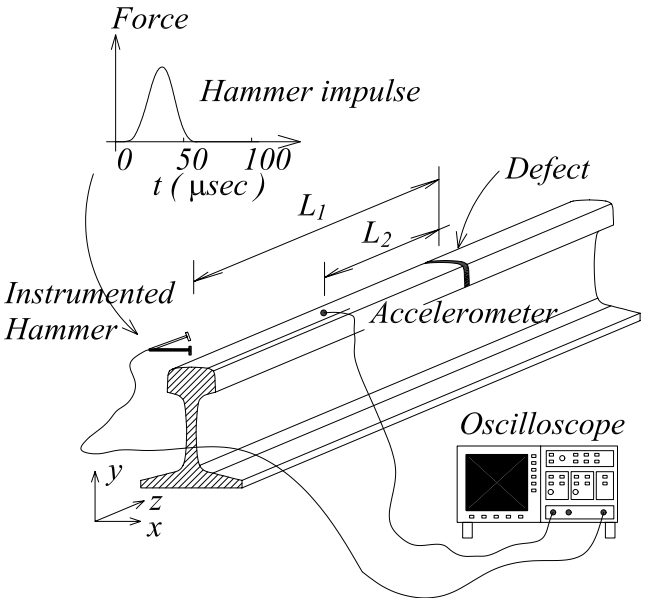


Figure 7 – Experimental setup for measuring reflections of the vertical bending mode from transverse head defects.

In the finite element model, the rail was discretized in ABAQUS EXPLICIT by 8-node brick elements with linear deformation, lumped mass matrix, and 3 degrees of freedom per node. Fig. 8 shows the mesh employed with a typical head defect. The size of the elements used was 4 mm in the longitudinal (running) direction *z*, and between 2 mm and 4 mm in the cross-sectional directions, *x* and *y*. The longitudinal direction refinement followed the ten node-per-wavelength rule found acceptable for the plate below 45 kHz (in this case λ_{min} = 44.4 mm considering c = 2 km/sec). The cross-sectional refinement of the rail head also followed the ten-node-per-thickness rule expressed in eq. (9). To save computational efforts, the cross-sectional refinement criteria applied to the rail head was not strictly observed for the rail web and base as seen in Fig. 8. This approximation was considered acceptable since the defects were all located in the rail head. The rail section modeled was 2.4 m long. This length successfully isolated the defect reflection from the far end reflection. The total number of nodes was on the order of 300,000. The integration time step was set at 0.2 μsec. As for the plate case, this value satisfied the temporal criteria in eq. (1). All analyses were carried out on a 2.6 GHz Pentium IV with 1 GByte of RAM. A typical analysis for each of the defect cases took 150-200 minutes.

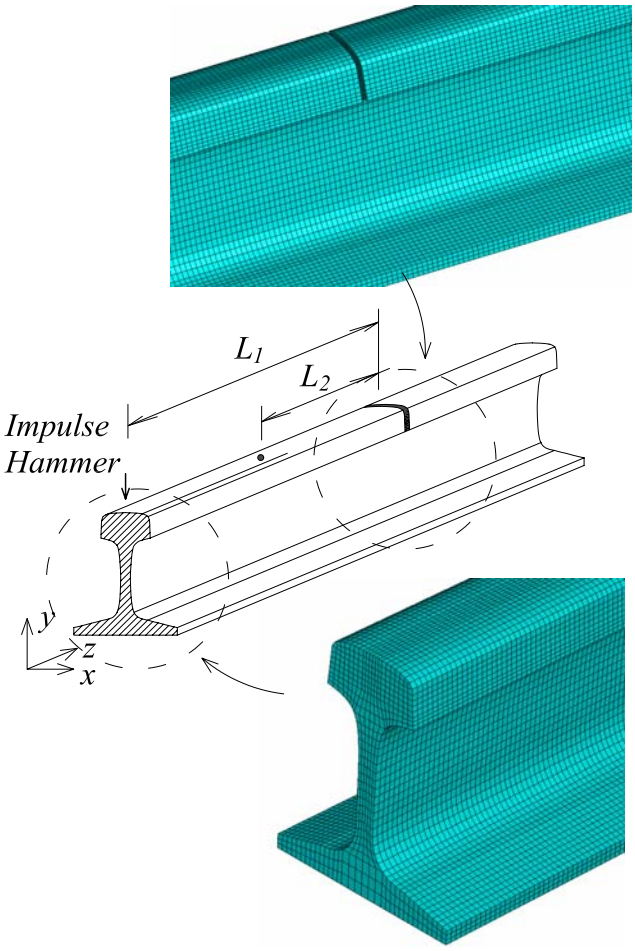


Figure 8 – Finite element mesh employed for the rail.

The accelerations of the node at the top of the rail located at L_2 from the defects were monitored in the three Cartesian directions. Fig. 9 shows two time signals for the vertical acceleration with the corresponding GWT scalograms. The signal in Fig. 9a was obtained by the model whereas the one in Fig. 9b is experimental. The two signals were obtained with the 100% defect in the perfectly transverse orientation. As indicated by the scalograms, the first arrival of the vertical bending wave and its reflection from the defect can be clearly identified. The group velocity dispersion curves were extracted from eq. (7) where now the subscripts *1* and *2* refer to the first arrival and the defect echo, respectively. Also, the distance L in eq. (7) is $2L_2$ in Figs. 7 and 8. The group velocity results are shown in Fig. 9e confirming the expected character of the vertical bending mode with a slight dispersion at the low frequencies. Also, the agreement between the numerical and the experimental results is satisfactory with a maximum discrepancy of 4%.

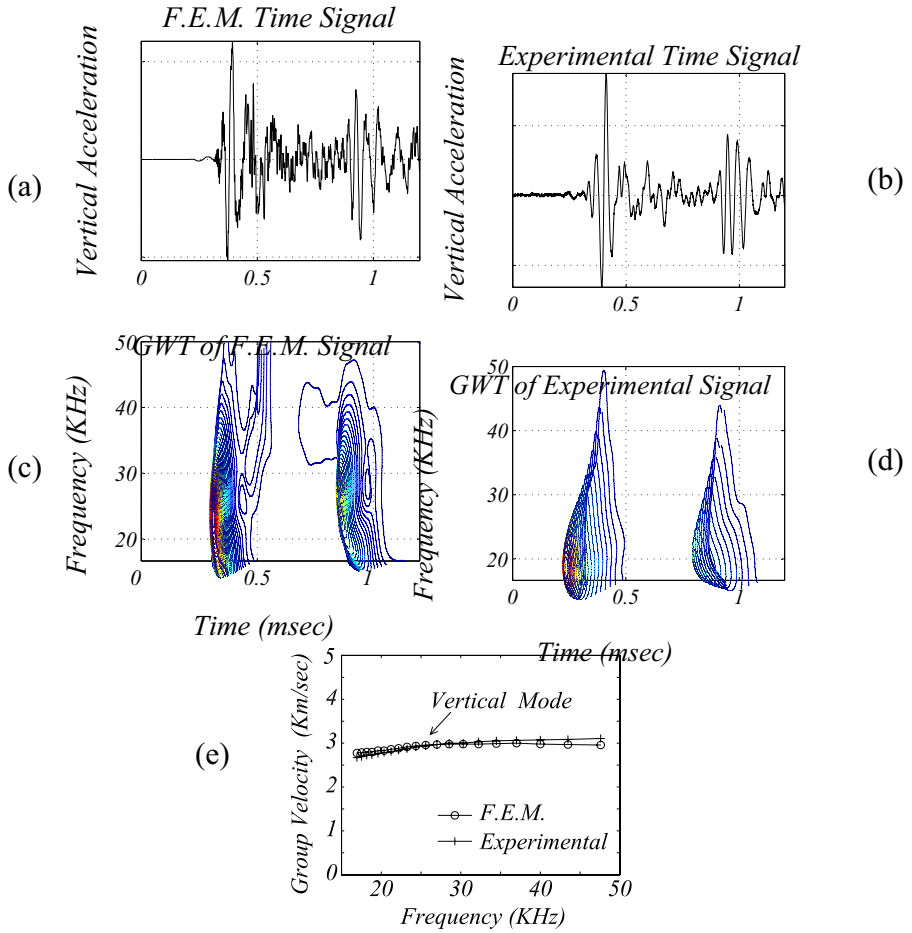


Figure 9 – Time histories and GWT scalograms of the vertical acceleration at the top of the rail head from the simulation (a and c) and from the experiment (b and d). Resulting group velocity shown in (e).

Energy reflection coefficients for all 12 defect cases were obtained by using eq. (8) and the results are presented in Figs. 10 and 11. Fig. 10 shows the same-mode reflection results, meaning the reflected vertical bending mode when the same mode is incident.

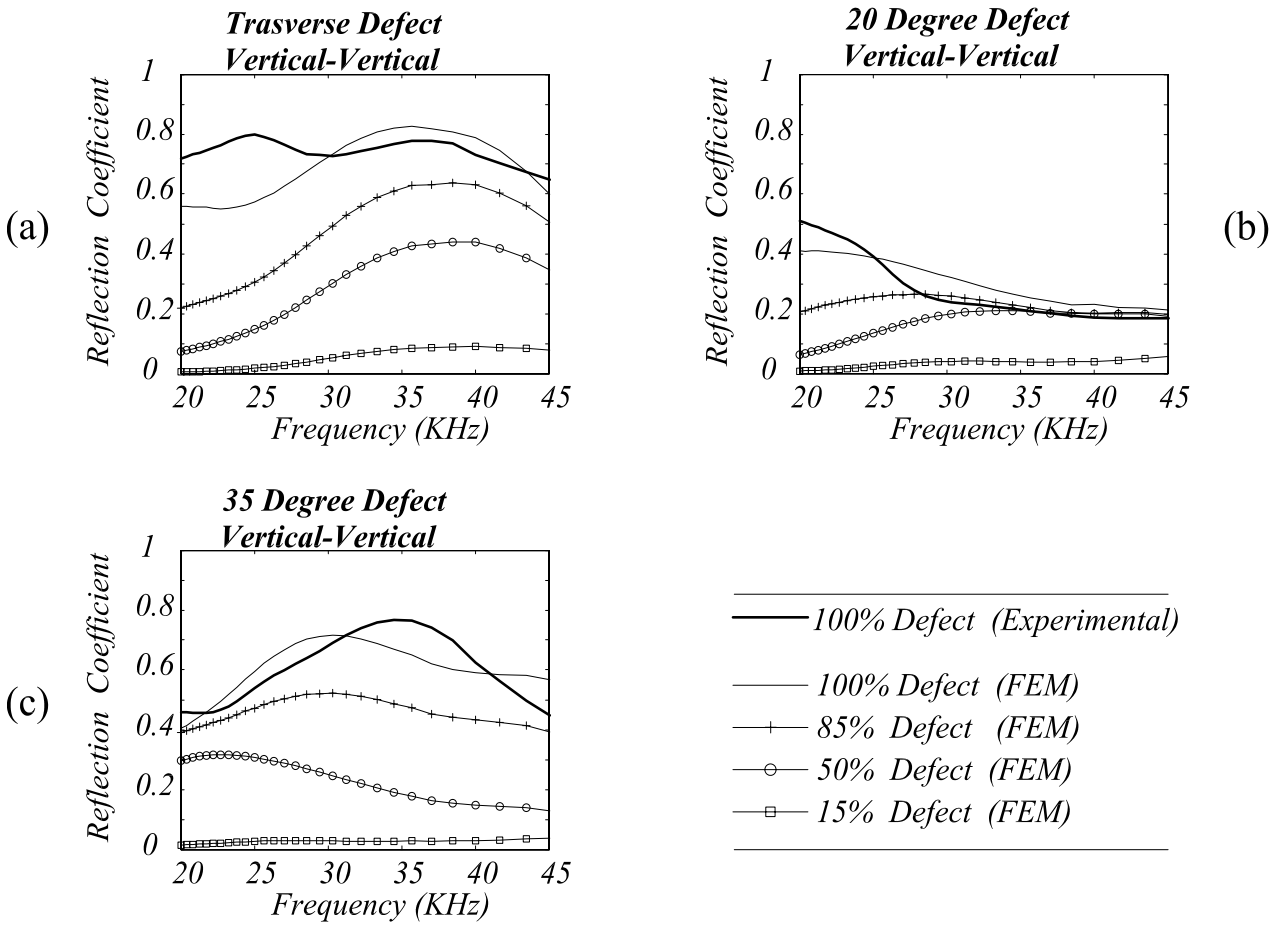


Figure 10 – Same-mode reflection coefficients for the vertical bending mode from the perfectly-transverse defects (a), the 20 deg inclined defects (b), and the 35 deg inclined defects (c).

The experimental result for the 100% defect is presented as a solid curve in these figures. It should be noted that the experimental data presented here were corrected for material attenuation effects since no attenuation was included in the model. The following correction equation can be easily derived:

$$R_{corrected}(f) = R(f) \times 10^{\frac{\alpha(f)(2L_2)}{10}} \quad (10)$$

where R(f) is the raw, experimentally-derived reflection coefficient spectrum from eq. (8), $\alpha(f)$ is the linear attenuation spectrum of the vertical mode in dB/m, and L_2 is the accelerometer-defect distance in meters. By using this correction equation the results can be converted to general reflection coefficients that are not bound by the experimental setup used. In other words, the results correspond to what would be measured if the detection accelerometer was positioned at the same location as the defect, resulting in stronger reflections than the raw data would indicate. The values of $\alpha(f)$ used in eq. (10) were those

previously measured for the vertical bending mode by Lanza di Scalea and McNamara (2003). In that study the attenuation was found highly varying with frequency, and generally smaller than 1 dB/m below 50 kHz.

Fig. 10 shows a slight discrepancy between the numerical and the experimental data. This is probably the result of slight differences between the real defects obtained by saw cutting and the flaws simulated in the model. Appreciable reflections are seen in the entire frequency range examined (20 kHz – 45 kHz), confirming the suitability of these ultrasonic frequencies for long-range defect detection. The general trend that emerges from Fig. 10 is an expected increase in reflection strength with increasing defect size for all three orientations of transverse (Fig. 10a), 20 deg oblique (Fig. 10b) and 35 deg oblique (Fig. 10c). For the 20 deg oblique defects, however, the reflection strength is less dependent on defect size above 30 kHz. A more unexpected result is a general decrease in reflection strength for the 20 deg oblique defects when compared to the other two orientations of transverse and 35 deg oblique. The effect can be explained by severe mode conversions for the 20 deg cuts that take energy away from the same-mode, primary reflection.

The mode-converted reflection coefficients are presented in Fig. 11 where the reflected lateral bending mode is now being monitored with the usual incident vertical mode. Mode conversion between a symmetrical mode (vertical bending) and anti-symmetrical mode (lateral bending) is expected when either the size or the orientation of the defect are such that the flaw is not symmetrical with respect to the longitudinal plane of symmetry of the rail. This occurs for all defect cases examined with the exception of the 100% defect in the perfectly-transverse orientation. The difficulty of monitoring mode conversion is that the lateral bending mode has the same group velocity as the vertical bending mode above 25 kHz, and thus the two are indistinguishable in the time domain. In this study the reflected lateral mode was identified by monitoring the transverse acceleration at the detection node (direction x in Fig. 8), rather than the vertical acceleration, along direction y. Since this node is on the plane of symmetry of the rail, the acceleration along x is zero for the vertical mode and it is only "activated" by the reflected lateral mode. Mode-converted reflections were computed from the usual eq. (8), where now GWT_2 refers to the scalogram of the transverse acceleration for the reflection (lateral mode), and GWT_1 refers to the scalogram of the vertical acceleration for the first arrival (vertical mode).

The results in Fig. 11 show that mode-converted reflections are generally weaker than the primary, same-mode reflections in Fig. 10. Also, mode conversions are stronger at the low frequencies, rapidly decaying above 30 kHz. Fig. 11a for the transverse orientation confirms that mode conversion is absent for the largest 100% defect. In the same figure, the reflection is stronger for the 50% defect when compared to the 85% defect. This apparently unexpected behavior can be explained by noting that the 50% defect is more un-symmetric than the 85% defect, and it thus generates a larger mode conversion. The same trend is observed for the 20 deg oblique reflections in Fig. 11b. In this case the strongest reflections occur for the largest 100% defect. The 35 deg oblique orientation, Fig. 11c, indicates a more intuitive trend of increasing reflections with increasing defect size.

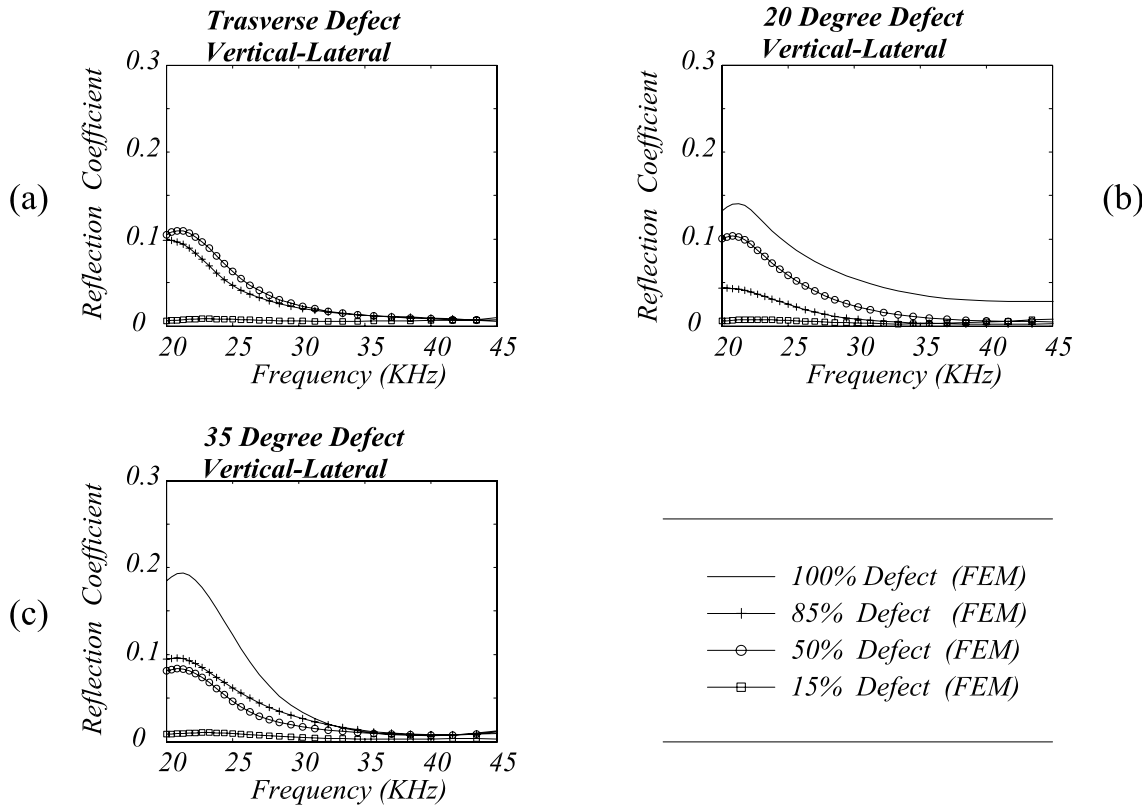


Figure 11 – Mode-converted reflection coefficients for the vertical bending mode from the perfectly-transverse defects (a), the 20 deg inclined defects (b), and the 35 deg inclined defects (c).

2.1.3 Summary and Conclusions of the Finite Element Study

This section demonstrated the use of a commercial finite element package, ABAQUS EXPLICIT, to model guided ultrasonic waves propagating at frequencies of tens of kilohertz. The waves, generated by impulsive excitation, were examined in a broad frequency range with the aid of joint time-frequency (wavelet) analysis.

The case of Lamb waves propagating in a 20mm-thick free plate was first examined with the purpose of checking the validity of existing recommendations for the spatial and temporal resolution of the finite element analysis as applied to ABAQUS. Modal properties of the Lamb waves in terms of group velocity as well as reflection coefficients from perpendicular and inclined edges were studied. The modal solutions and the reflection from the perpendicular edge were found less sensitive to the mesh refinement than the reflections from the oblique edge. The reason is the increased accuracy required to account for the presence of mode conversions and evanescent modes. Adequate accuracy in modeling edge reflections in the plate was achieved with 10 nodes per wavelength below 45 kHz and 20 nodes per wavelength below 60 kHz. As for the cross-sectional refinement, it was found that 10 nodes across the thickness are required for both the 45 kHz and the 60 kHz limits.

These guidelines were then applied to modeling guided waves propagating in 115lb A.R.E.M.A. rails in the context of rail defect detection by long-range ultrasonic inspection. The vertical bending mode was examined for its ease of generation and detection in the field. This mode was generated by an impulse excitation at the top of the rail head in the vertical direction. The defects examined included four different sizes of transverse head flaws at three different orientations, for a total of twelve cases. The model extracted same-mode reflections (vertical incident and reflected) and mode-converted reflections (vertical incident and lateral reflected). A limited set of experimental data gathered in the laboratory was also obtained for comparison with the finite element predictions. The study shows that appreciable reflections from defects as small as 15% of the rail head are found in the 20 kHz – 45 kHz range. As expected, the same-mode reflection strength generally increases with increasing defect size, although this dependence is not marked for the 20 deg oblique flaws above 30 kHz. Also, the same-mode reflections from the 20 deg flaw orientations are weaker than those from the 35 deg orientations. Below 30 kHz, the mode-converted reflections are found to be substantial, their strength rapidly decreasing at the higher frequencies. The amount of mode conversion depends on the symmetry of the flaw with respect to the longitudinal plane of symmetry of the rail. This consideration helps explaining why in some cases stronger mode conversions are predicted for smaller defects.

It should be noted that the numerical model assumed an unrestrained rail, whereas the experimental tests were conducted on rails laid on wood sleepers with no ties or fasteners. However, the boundary conditions are expected to affect the wave propagation properties for frequencies below 10 kHz. In the frequency range of interest in this study, larger than 20 kHz, most of the wave energy remains confined within the rail head and it does not “see” the sleepers. Based on this consideration, the findings could be reasonably extended to the field where the rail is fastened on the sleepers.

Finite element modeling becomes an essential tool to predict the wave interactions with a variety of defects that would be impractical to replicate experimentally. The reflection coefficient spectra from this variety of defect sizes and orientations can then be fed to an Automatic Defect Classification (ADC) algorithm that operates in a supervised learning mode. The first implementation using Support Vector Machines for ADC of transverse defects in rails was recently demonstrated in the laboratory by McNamara et al. (2004). The results of the ADC study are presented in the Section 4 of this report.

2.2 Hammer Impact Generation – Air-coupled (Non-contact) Detection

The experimental setup used for this system is shown in Fig. 12. An impulse hammer (PCB 086D80 mini hammer) was employed to excite guided waves at frequencies below 50 kHz by striking vertically the top of the rail head. A pair of air-coupled transducers with broadband response in the DC-2MHz range was used to detect the waves. Using two sensors allowed to extract both reflection coefficients and transmission coefficients for the target defects. Pictures of the sensors located prior to and across a transverse defect are shown in Figs. 12b and 12c, respectively. In these pictures the large lift-off distance of the sensors can be appreciated. In this portion of the study results are presented for a transverse crack that runs for the entire section of the rail head.

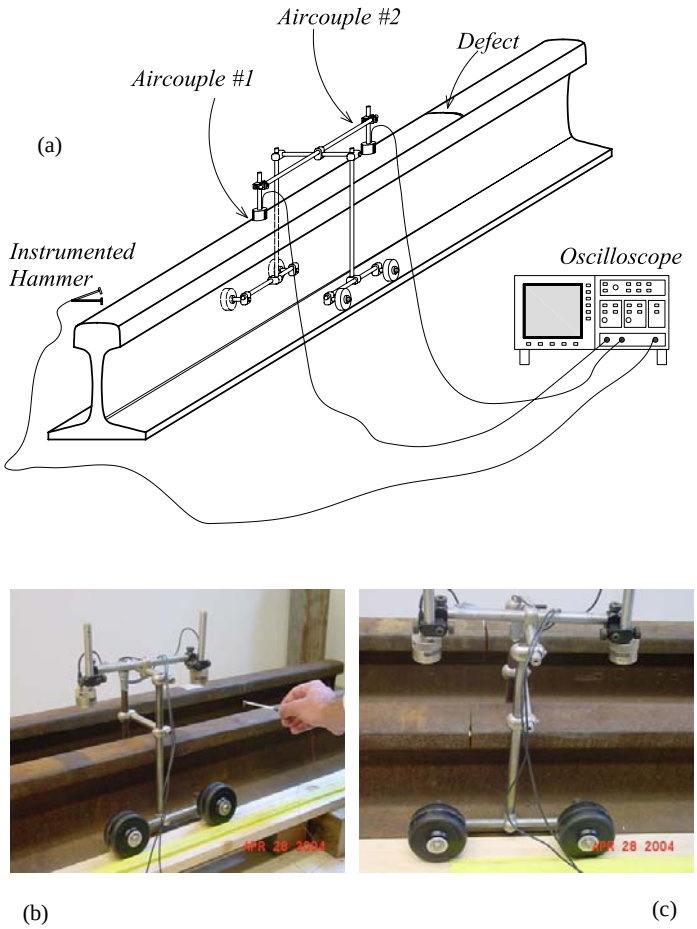


Figure 12 - System for detecting transverse cracks in rails by generating ultrasonic guided waves with a hammer and detecting the waves with a pair of air-coupled sensors.

2.2.1 Effect of Sensor Inclination

Determining the role of the sensor orientation is critical to maximize the sensitivity to the defects as well as to characterize the system's tolerance to accidental misalignments in the

field. The optimum air-coupled detection angle from the normal to the rail surface, α, is given by Snell's law of refraction:

$$\alpha = \arcsin\left(\frac{c_p}{c_{air}} \sin\theta_p\right) \tag{11}$$

where c_p is the phase velocity of the guided wave in the rail, c_{air} =330 m/sec is the wave velocity in air and $\theta_p = 90^o$ for a wave propagating parallel to the rail surface. Considering that a vertical-type guided wave propagates at c_p = 3,000 m/sec in the frequency range of interest, eq. (1) gives $\alpha = 6.3^o$ (Fig. 13). This is the optimum sensor angle to detect crack reflections. The convention here is positive angle towards the defect (i.e. away from the generating hammer).

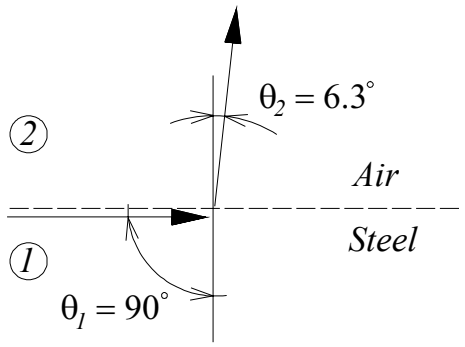


Figure 13 – Snell's law of refraction applied to air-coupled detection of guided waves.

However, it is also important to detect the wave incoming from the hammer that travels in an opposite direction from the crack reflection. In this case the optimum detection angle would be – 6.3º, implying sensor oriented towards the hammer. A compromise is to orient the sensor parallel to the rail surface ($\alpha = 0^o$). Fig. 14a shows the signal detected by such an orientation where the sensor lift-off distance, *h*, is fixed at 76 mm (3") and the distance sensor-defect is 1.120 mm (44"). The first arrival and the reflection from the defect are clearly visible. Fig. 14b is the continuous wavelet transform (CWT) scalogram of the time signal revealing its time-frequency spectrum.

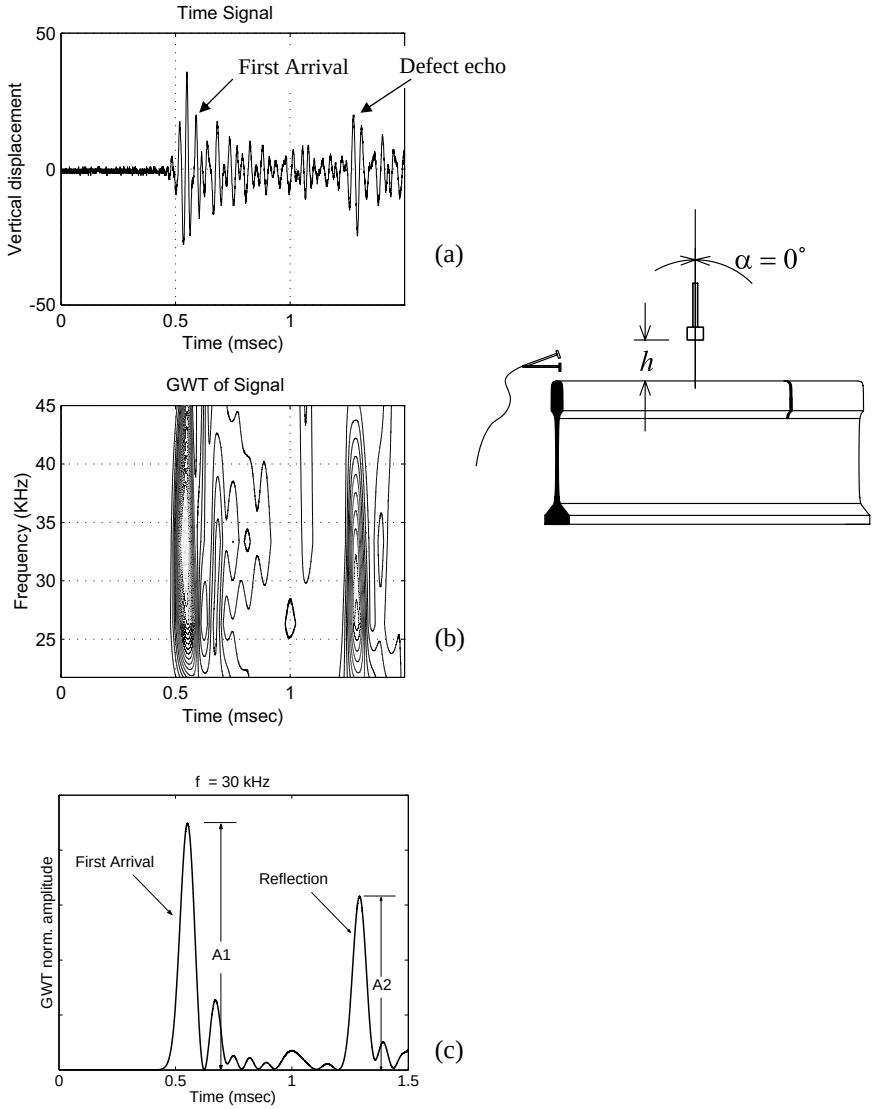


Figure 14 - Detection of transverse crack by reflection measurements (detection angle 0º). (a) time history; (b) continuous wavelet scalogram; (c) wavelet peaks at 30 kHz.

Fig. 14c illustrates the extraction of the reflection coefficient at a frequency of 30 kHz from the wavelet scalogram. The reflection coefficient at frequency *f*, *R*(*f*), is here defined as the ratio between the reflected energy, A_2, and the incoming energy, A_1, following:

$$R(f) = \frac{A_2(f)}{A_1(f)} \tag{12}$$

A value of *R(f)*=1 thus means that all of the incoming energy is reflected from the defect, whereas *R(f)*=0 means that no energy is reflected and it is all transmitted.

Fig. 15 presents the results for the sensor inclinations α = 10º and 40º. Fig. 16 presents the results for α = -20º and -40º.

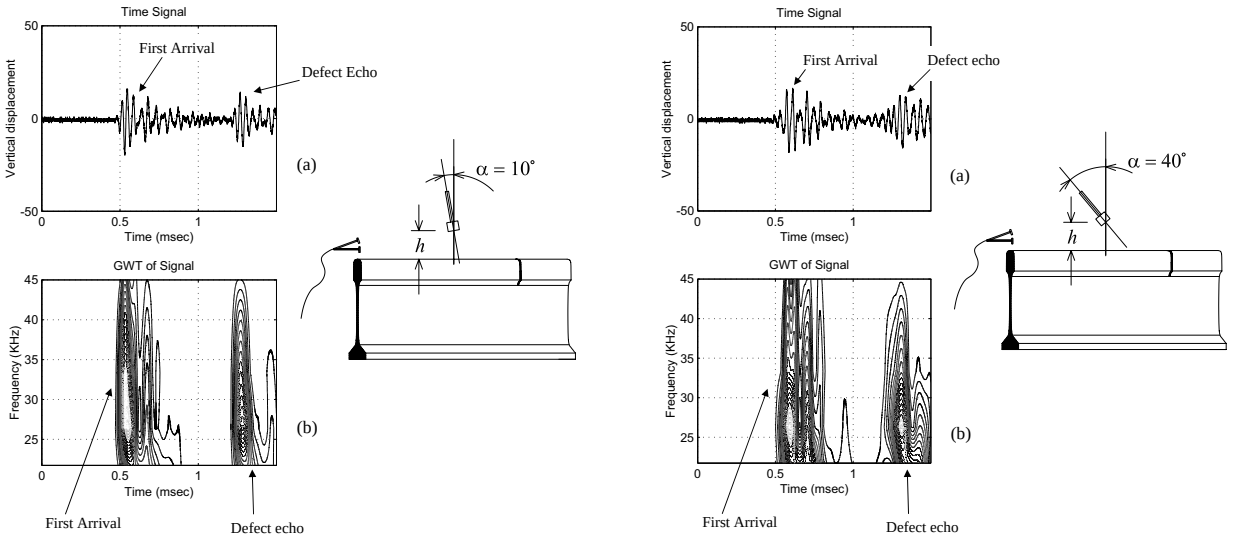


Figure 15 - Detection of transverse crack by reflection measurements (detection angles 10º and 40º). (a) time history; (b) continuous wavelet scalogram.

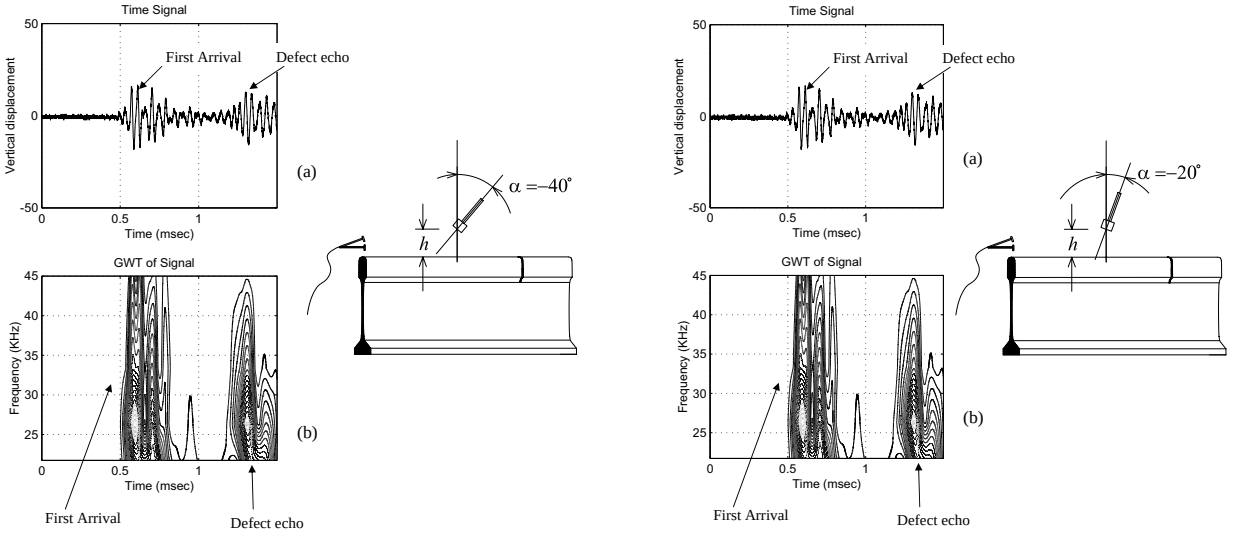


Figure 16 - Detection of transverse crack by reflection measurements (detection angles -20 º and -40º). (a) time history; (b) continuous wavelet scalogram.

Reflection coefficients were measured by varying the sensor orientation in the range – 50º to 50º. The results are shown in Fig. 17. Here *R* was calculated at a frequency of 30 kHz that provided strong reflections. It is remarkable that it is possible to detect defect reflections in a large range of sensor orientations, especially considering the large lift-off distance of 76 mm (3") used in the tests. As expected, the reflection strength increases when the sensor is oriented towards the defect. This suggests that in those cases where signal-to-noise ratio may be a problem, such as when reflections are sought from small defects, the sensors should be oriented away from the generation hammer. This point will be demonstrated further when

discussing the laser/air-coupled system in the second part of this paper. An optimum angle to detect the defect reflection is identified at 35º in Fig. 17.

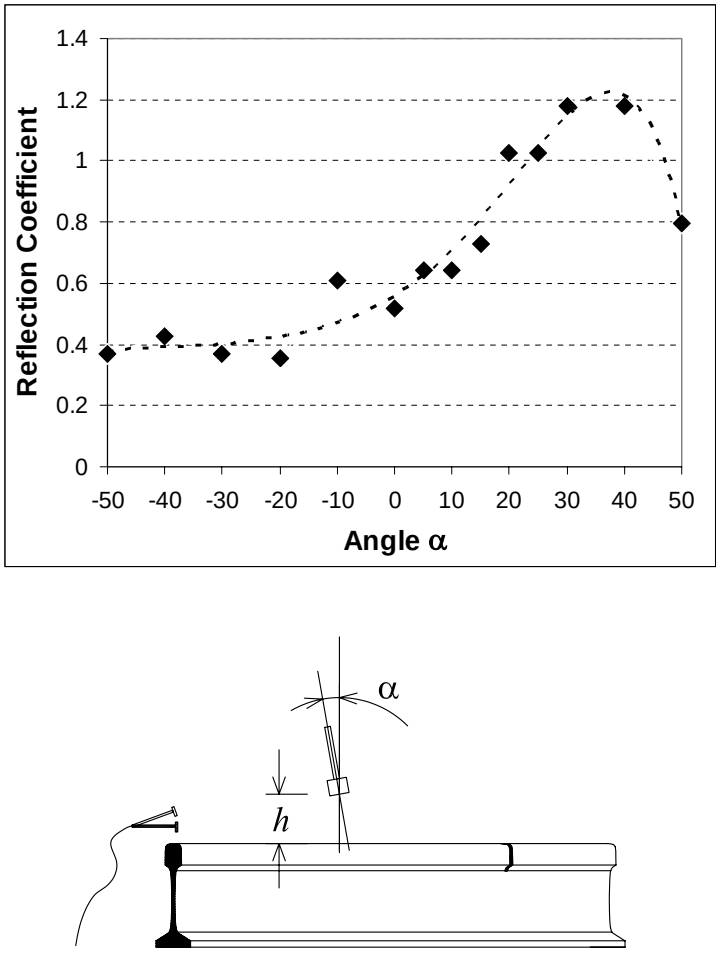


Figure 17 - Variation of the reflection coefficient for transverse crack as a function of detection angle.

The optimum detection angle depends on the phase velocity of the guided waves which, in turn, is a function of the wave frequency. Consequently, the value of 35º may not be appropriate for frequencies beyond the 45 kHz limit examined here. Values of *R* larger than 1 are a result of the more favorable detection of defect-reflected signals rather than incoming signals at large sensor inclinations. Also, the variations of *R* are very small for sensor angles between – 10º and 10º. This implies that some transducer misalignment would be well tolerated by the system in the field.

2.2.2 Effect of Sensor Lift-off Distance

Positioning the sensor as far away as possible from the top of the rail head was the target of this study. At a minimum, the distance must satisfy the clearance envelope of 65 mm (2.5") that is generally recommended for any new rail inspection system. Larger sensor distances, however, tend to degrade the defect detection sensitivity due to the losses in air, particularly severe for high-frequency waves.

Reflection coefficients from the large transverse defect positioned 1.120 mm (44”) away from the probes were obtained while varying the sensor lift-off distance, h. The results of these tests, only summarized here for the sake of space, indicated some unexpected trends for $h <$ 25.4 mm (1”). It was found that for small h standing waves form between the rail and the sensor face producing interference patterns that complicate the defect detection task. The formation of these standing waves in air (“lift-off echoes”) is shown in Fig. 18.

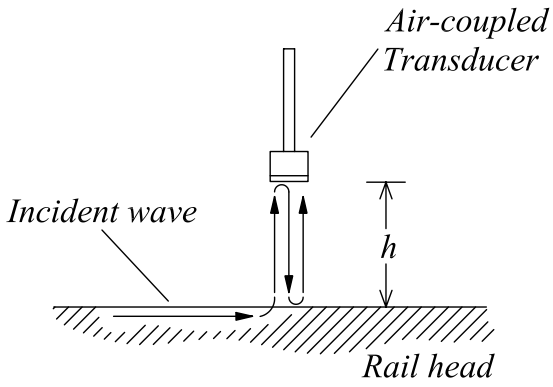


Figure 18 – Formation of resonances in the lift-off air gap.

The time interval between multiple resonances in air is given by:

$$\Delta t = \frac{2h}{c_{air}} \tag{13}$$

Let us consider, for example, the case of h = 19 mm (0.75”) shown in Fig. 19. Eq.(13) yields Δt = 0.12 msec that matches with the lift-off echoes seen in the scalogram of Fig. 19b. These echoes can be seen following the first arrival, as well as following the defect echo.

When the sensor is positioned further away from the rail head, at h = 38 mm (1.5”), the lift-of echoes should be spaced at Δt = 0.23 msec. This estimate matches with the experimental results in Fig. 20. It is important to note that the lift-off echoes can overlap with the defect reflection, giving raise to a constructive interference condition or to a destructive interference condition. In the former case the defect size would be overestimated whereas in the latter case the defect size would be underestimated or even go undetected. It is thus clear that the lift-off echoes are a spurious effect that can affect the reliability of the defect detection.

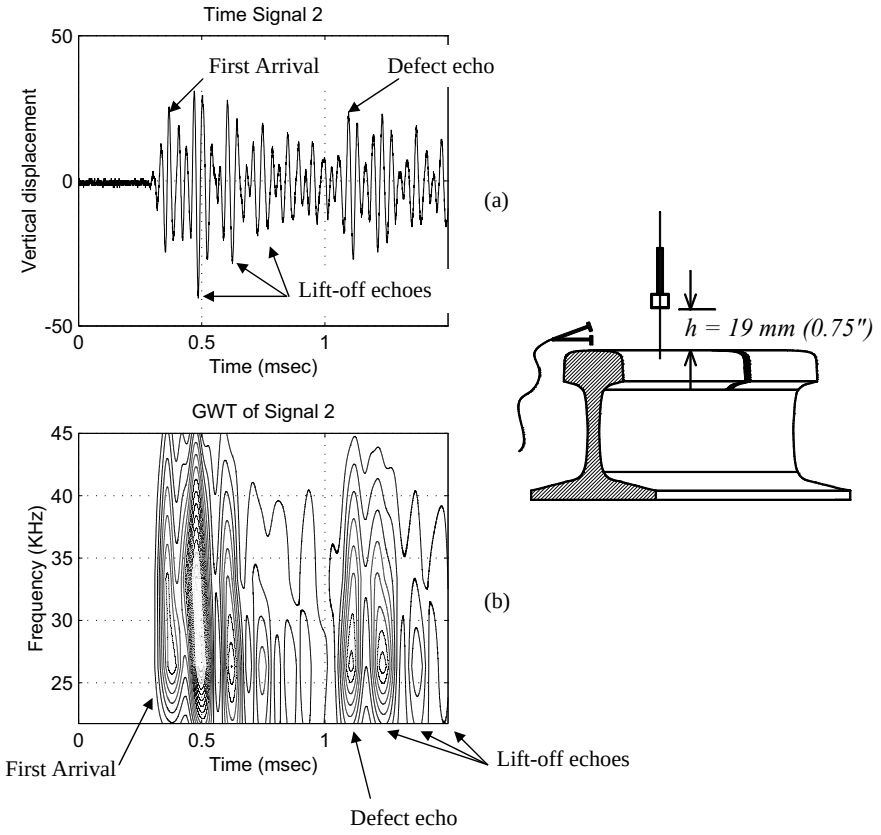


Figure 19 – Defect reflection measurements for sensor lift-off distance of *h* = 19 mm.

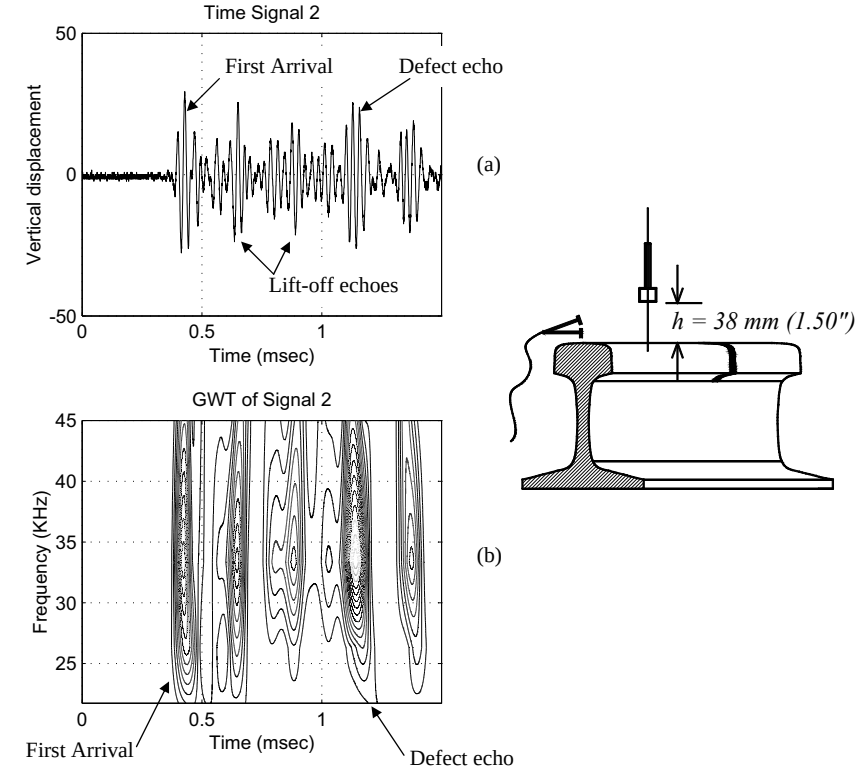


Figure 20 – Defect reflection measurements for sensor lift-off distance of *h* = 38 mm.

The problem is alleviated for large lift-off distances. Fig. 21 shows the case of h = 64 mm (2.5") where the lift-off echo is confined in time and does not influence first arrival and defect reflection. At h = 89 mm (3.5"), Fig. 22, lift-off echoes disappear as they get quickly damped.

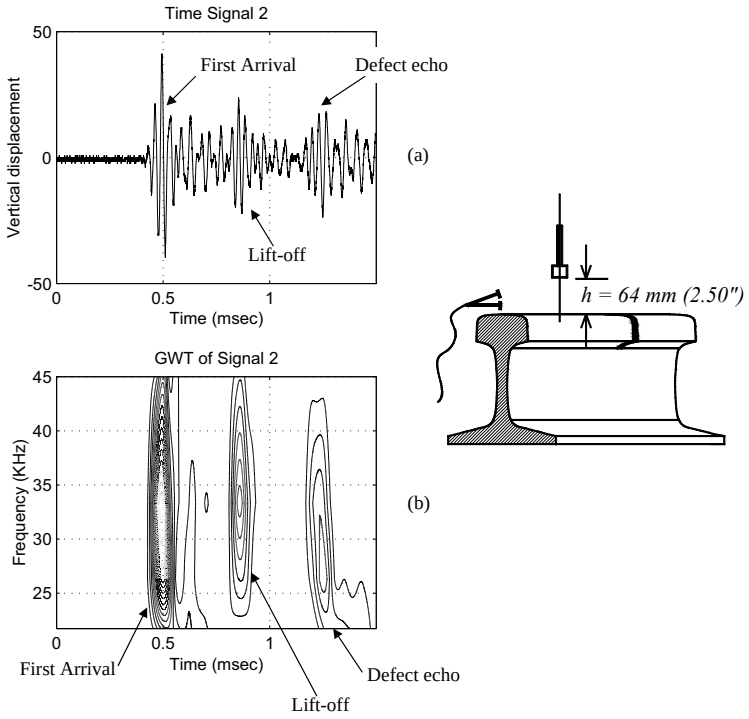


Figure 21 – Defect reflection measurements for sensor lift-off distance of h = 64 mm.

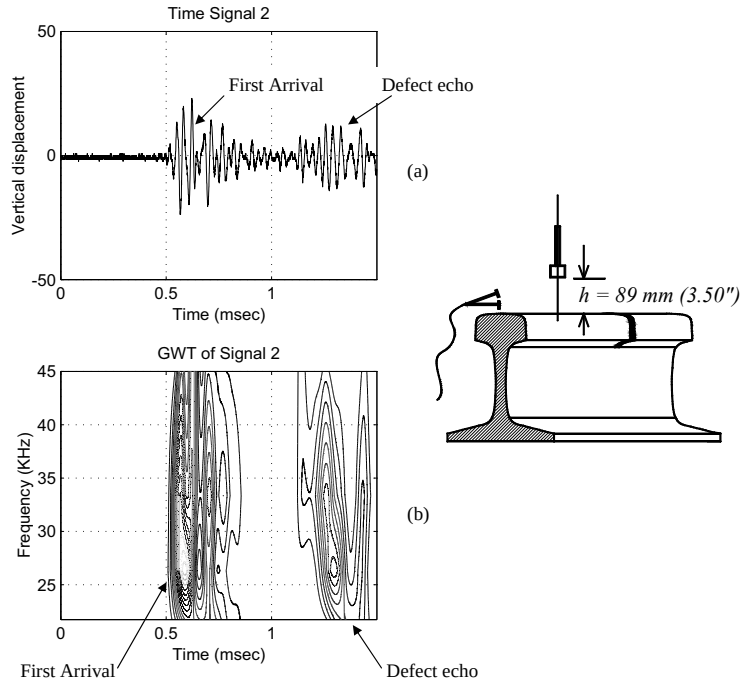


Figure 22 – Defect reflection measurements for sensor lift-off distance of h = 89 mm.

Fig. 23 shows the variation of the reflection coefficient as a function of sensor lift-off distance for the large transverse head defect. The coefficient *R* was computed from eq. (12) as before, considering a frequency of 30 kHz. It can be seen that larger *R* values are obtained for smaller lift off distances. This is a result of the lift-off echoes superimposing constructively with the defect reflection. The effect is particularly severe at h = 25.4 mm (1”). The opposite case is found at h = 64 mm (2.5”). It can be seen that the *R* values are more stable for lift-off distances larger than 76 mm (3”). Also, at these large distances the defect reflections are only slightly smaller than those measured at shorter distances.

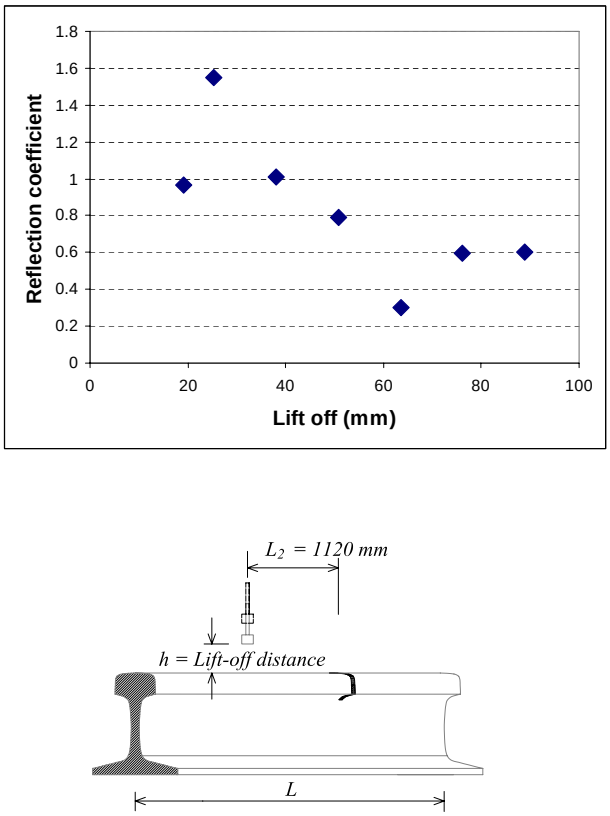


Figure 23 - Variation of the reflection coefficient for transverse crack as a function of sensor lift-off distance.

It can thus be concluded that lift-off distances between 76 mm (3”) and 90 mm (3.5”) should be selected for obtaining a stable reflection from the large transverse defect in the DC - 50 kHz frequency range examined under hammer excitation. Previous studies (Lanza di Scalea and McNamara 2003; McNamara et al. 2004; McNamara 2003; Bartoli et al. 2004) indicate that such frequencies are appropriate for detecting transverse head defects as small as 15% of the head cross-sectional area. The potential thus exists for a truly “non-contact” system that stays outside the clearance envelope of 65 mm (2.5”) generally recommended for new rail inspection systems.

2.2.3 Transmission Approach

By using a pair of sensors as shown in Fig. 12, it is possible to detect a defect by monitoring the transmission coefficient between the two sensors. This represents a redundant defect detection tool in addition to the reflection measurements discussed above. The transmission coefficient, *T(f)*, can be extracted from the CWT scalograms of the signals detected by sensor #1 and sensor #2 following:

$$T(f) = \frac{A_{sensor\#2}(f)}{A_{sensor\#1}(f)} \tag{14}$$

where A indicates the energy of the detected signals at frequency f and sensor #1 is the closest to the impact hammer. When both sensors are located prior to the defect, they should detect the same signal (neglecting attenuation losses in the rail material) and thus $T = 1$. When the sensors are on either side of the defect, part of the energy is reflected back to sensor #1 and T is substantially reduced. Once the sensor pair passes the defect, T returns to its initial value of 1. Monitoring transmission coefficients at short ranges can thus provide a defect detection tool in addition to monitoring reflection coefficients at long ranges. Hence the potential for an improved reliability of defect detection.

Figs. 24 through 27 demonstrate the detection of the 100% head transverse crack by transmission measurements for various positions of the sensor pair relative to the crack. The sensor orientation angle was 0º and lift-off distance was 76 mm (3”). Figs. 24a and 24b show the time signals detected by the two sensors when they are both positioned prior to the crack, with L_2 = 1100 mm. The corresponding wavelet scalograms are shown in Figs. 24c and 24d, respectively. It can be seen that the energy detected by the two sensors is comparable as expected ($T \sim 1$). As the sensor pair gets closer to the crack, L_2 = 380 mm in Fig. 25, the transmission coefficient is still close to one. In addition, sensor #2 detects the crack reflection providing a first indication of the presence of the defect. This demonstrates that reflection data can be monitored concurrently with transmission data.

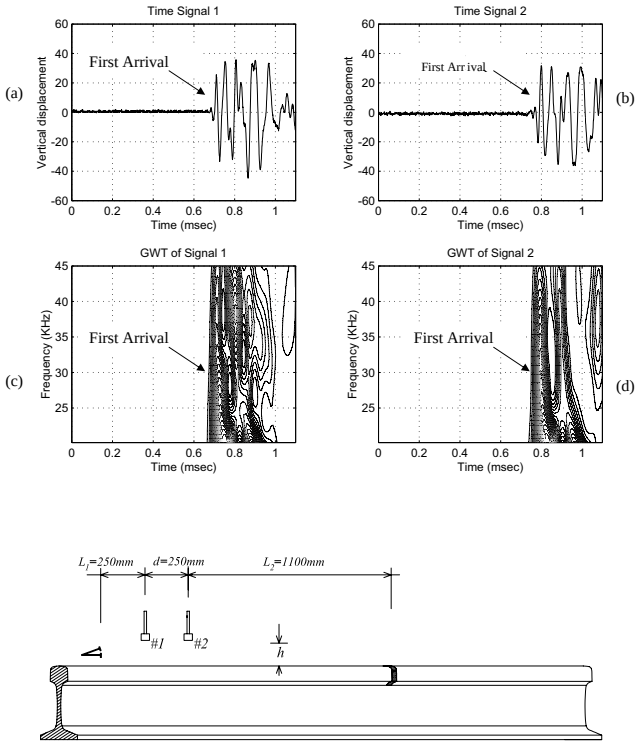


Figure 24 – Crack detection by transmission approach (both sensors prior to defect).

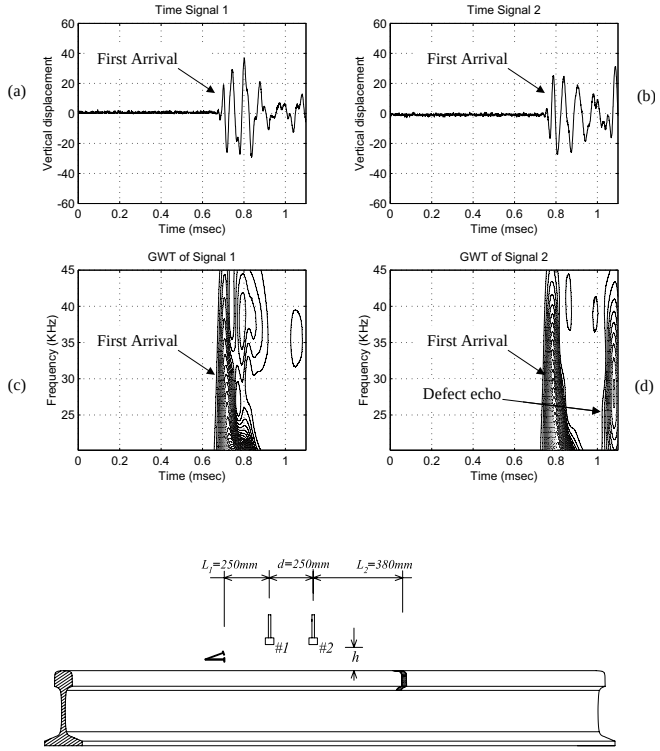


Figure 25 – Crack detection by transmission approach (both sensors prior to defect).

Fig. 26 shows what happens when sensor #2 is moved past the crack. In this case the signal at sensor #1 is clearly larger than that at sensor #2 because much of the energy is reflected, rather than transmitted past the defect ($T < 1$). The defect reflection is still visible at sensor #1 besides the first arrival.

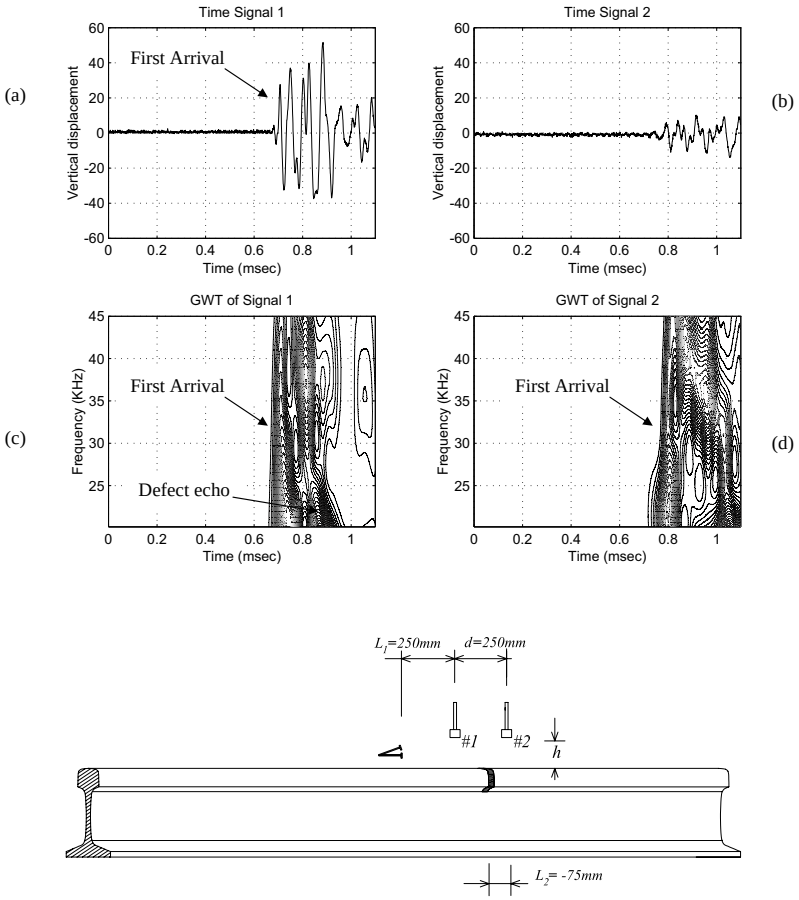


Figure 26 – Crack detection by transmission approach (sensors across defect).

Finally, as the sensor pair is moved past the crack, Fig. 27, the two signals become comparable again ($T \sim 1$).

Fig. 28 summarizes the transmission coefficients measured at 30 kHz for different positions of the sensors relative to the crack (as above, L_2 is the distance from sensor #2 to the defect). As expected, T is close to 1 when both of the sensors are on the same side of the defect. T decreases substantially when the sensors are on either side of the defect, providing a clear indication of its presence.

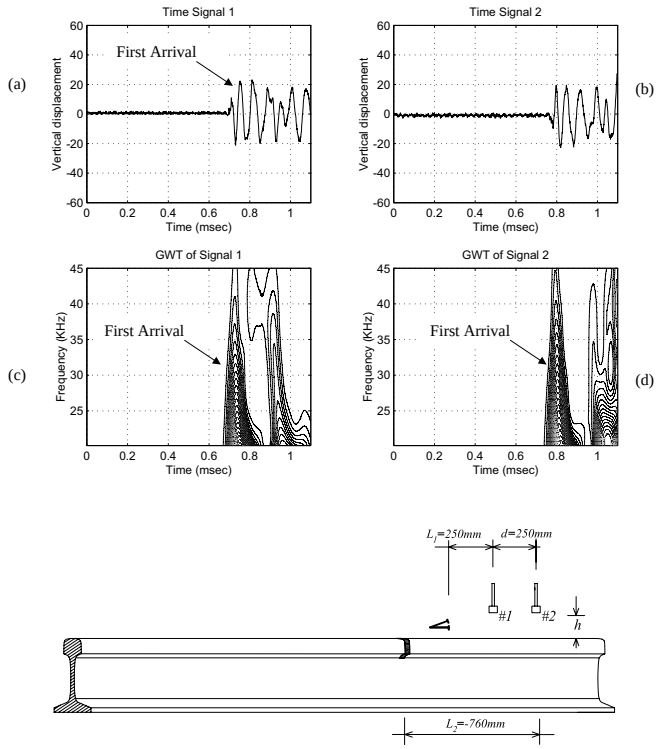


Figure 27 – Crack detection by transmission approach (both sensors past defect).

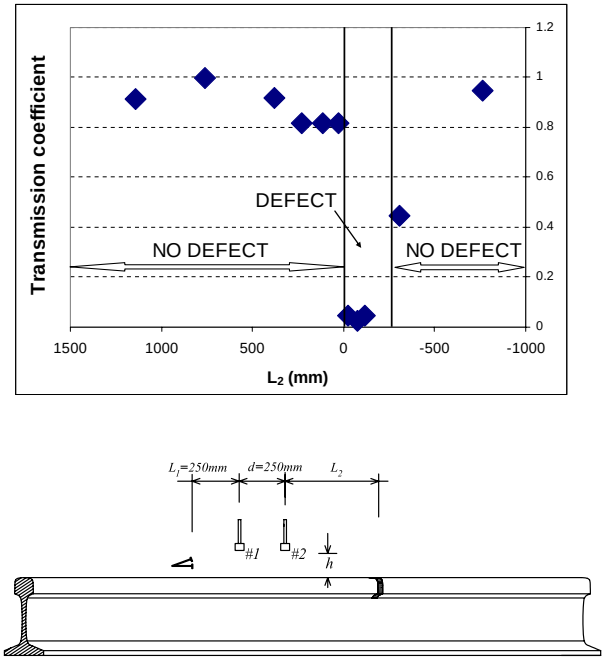


Figure 28 - Variation of the transmission coefficient for perfectly-transverse crack as a function of sensor position relative to the defect.

The same tests were performed on 100% head cracks oriented at 10º, 20º and 35º from the transverse rail direction. The results are shown in Fig. 29. It can be seen that the oblique defects produced the expected trend of *T* as a function of sensor position, although the results were more scattered than for the perfectly-transverse crack.

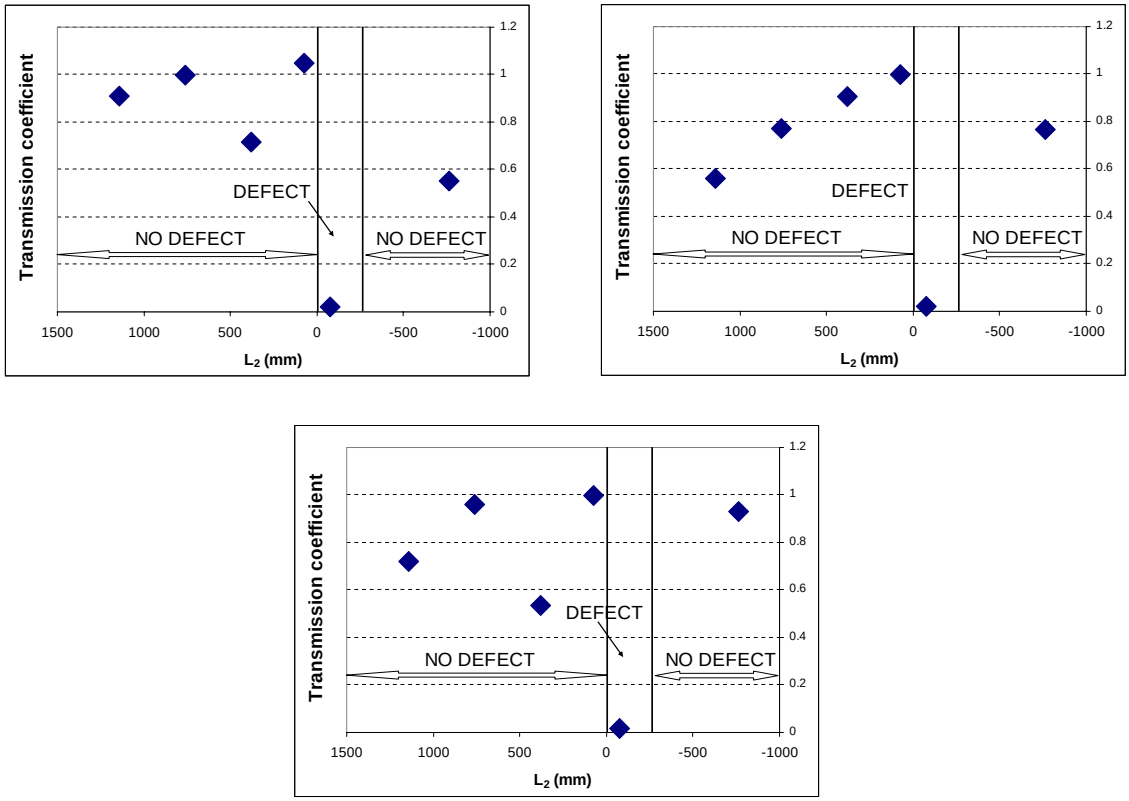


Figure 29 - Variation of the transmission coefficient for oblique cracks as a function of sensor position relative to the defect (top left: 10º orientation crack; top right: 20º orientation crack; bottom: 35º orientation crack).

The transmission results presented in this section refer to large head cracks. These can be detected with frequencies below 50 kHz generated by an impulse hammer. It should be remarked that the same approach of monitoring reflection and transmission coefficients proves successful for detecting much smaller head defects, specifically surface-breaking cracks as small as 1 mm in depth. For the smaller defects, higher frequencies, above 100 kHz, are needed. These can be excited by a pulsed laser as shown in following sections of this report.

2.2.4 Use of the Discrete Wavelet Transform to Enhance Defect Detection

The extraction of the reflection and the transmission coefficients presented so far was performed by transforming the signal in the CWT time-frequency domain. Although efficient in the laboratory, this approach is too slow for field application. The Discrete Wavelet Transform (DWT) is the fast version of the CWT and it can be implemented in real-time as required by high-speed inspections in the field. Besides increasing speed, the DWT performs an excellent de-noising of the signals thereby increasing the chances of properly identifying

the signature of a potential defect that would be otherwise missed or misclassified. This is particularly relevant in the case of small defects, discussed later in the report, whose signatures may be buried into noise.

The DWT operated here uses a series of low-pass and high-pass filters to decompose the original time signal into various "levels" representing different frequency bands (filter bank decomposition) (Mallat 1999). At each decomposition level, a low-pass filter and a high-pass filter are used to obtain approximations (*cA*) and details (*cD*) containing wavelet coefficients, as shown in Fig. 30.

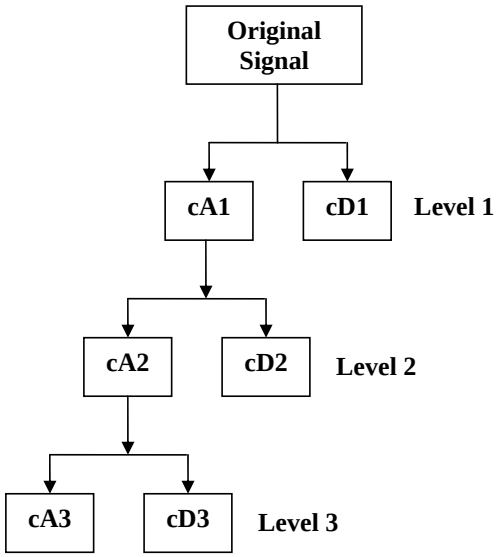


Figure 30 – Discrete Wavelet decomposition of a signal by filter bank.

A decomposition level, j, corresponds to a center frequency, $f(j)$, given by:

$$f(j) = \frac{\Delta \times F}{2^j} \tag{15}$$

where Δ is the sampling frequency of the original signal and F is the center frequency of the particular mother wavelet used. For signal reconstruction, the wavelet coefficients are first correlated to inverse high-pass and inverse low-pass filters to generate the reconstructed approximations (*A*) and reconstructed details (*D*). Finally, the original signal is reconstructed from the linear combination of reconstructed approximations and details:

$$s = A_N + \sum_{i=1}^{N} D_i \tag{16}$$

Fig. 31 demonstrates the DWT decomposition of a time signal detected by the air-coupled sensor positioned 1,120 mm (44") away from the 100% transverse head crack with a lift-off of 76 mm (3") from the rail top surface. It is clear that increasing the decomposition level implies zooming into the low frequency portions of the signal. The Daubechies wavelet of order 10 (db10) was used for this analysis.

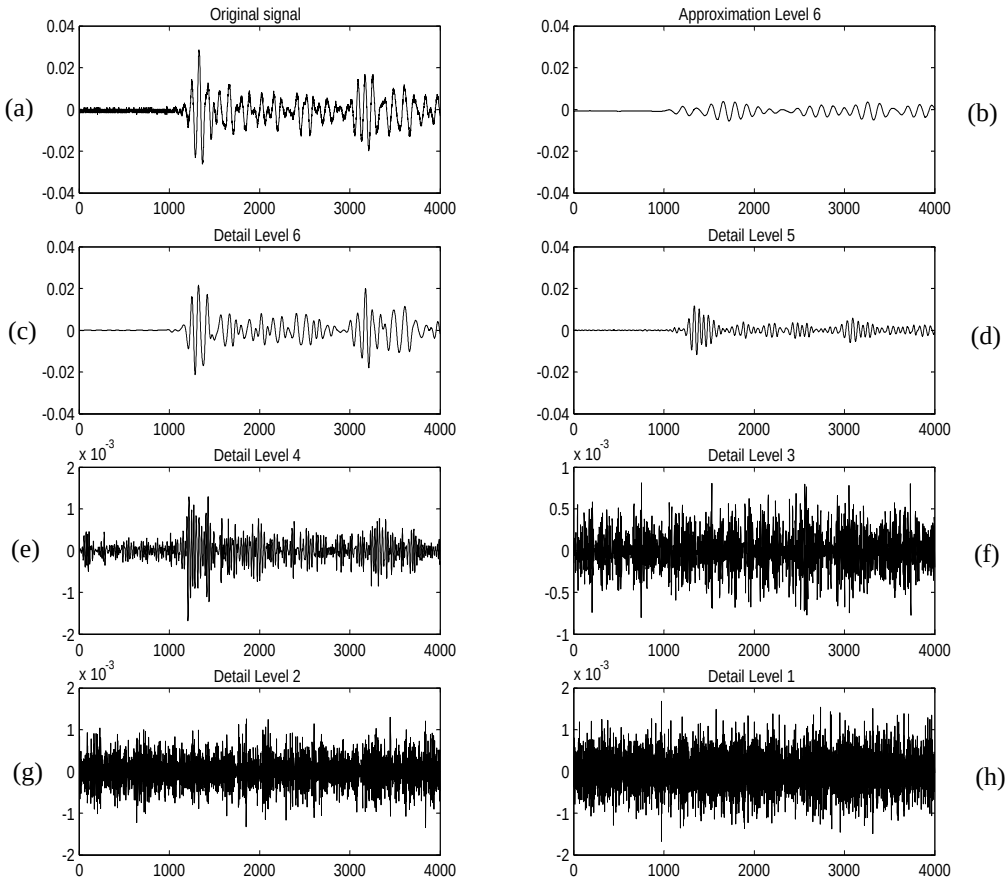


Figure 31 –The first six DWT decomposition levels of a signal detected in the rail.

Fig. 32 demonstrates the procedure of "pruning" and "thresholding" wavelet coefficients to increase the defect detection sensitivity. In particular, Fig. 32a presents the time signal reconstructed only from the level 6 DWT decomposition (pruning). From eq. (15), level 6 is centered at 26 kHz that contains most of the reflected energy. The plot in Fig. 32a resembles closely the raw time history recorded by the sensor. The defect reflection is not easily distinguishable from the first arrival due to noise and reverberations. If only the largest wavelet coefficients are retained (thresholding), the reconstructed signal is de-noised and the defect reflection now appears clearly (Fig. 32b). In Fig. 32b the threshold was set to 60% of the maximum value of the wavelet coefficients.

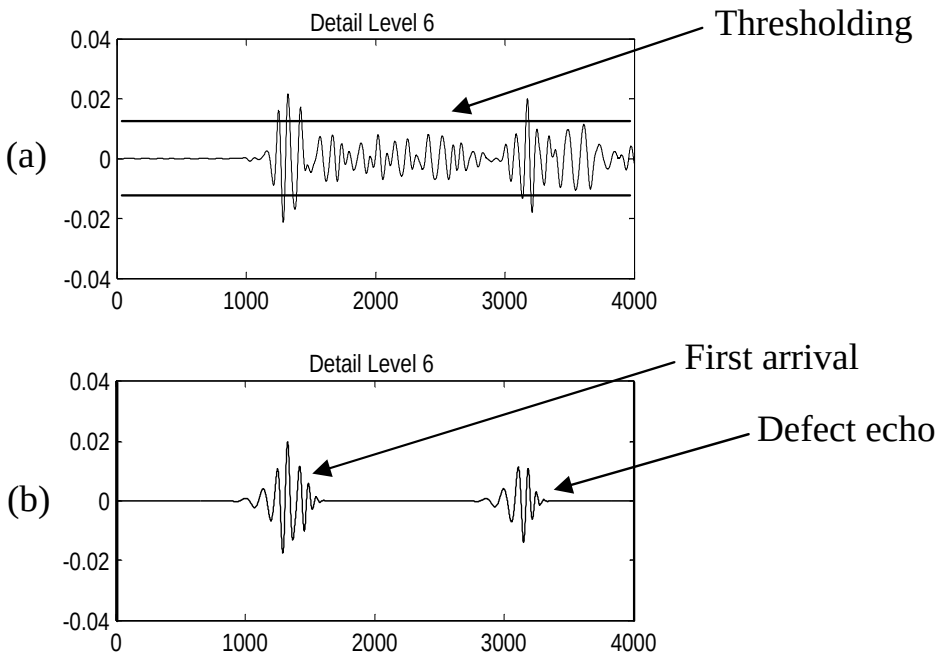


Figure 32 – Pruning and thresholding DWT decomposition level 6 to increase defect detection sensitivity.

Another important product of the DWT is signal compression. At each decomposition level, the number of points is halved (downsampling). Thus an original signal that had 4,000 points is reduced to only 60 point at level 6 and much less after thresholding. The task of extracting reflection and transmission coefficients from such compressed signals results in a very efficient process. The following discussion focuses on transmission measurements.

The role of the chosen level of DWT decomposition is important as different choices highlight different frequency components of the signal. An example is given in Fig. 33. Here decomposition levels 5 and 6 reconstruct frequency bands centered at 32 kHz and 16 kHz, respectively (db10, sampling frequency 1500000). The difference is evident by comparing the plots in the middle row to those in the bottom row of the figure. Such multi-resolution analysis allows the extraction of the transmission (and/or reflection) coefficients for various frequency bands at once thereby providing complete information on the defect being detected. Frequency-dependent transmission coefficients can be computed from eq. (14) by taking the ratio A_2/A_1 at level 6 (Fig. 33c and d) and the ratio A_2/A_1 at level 5 (Figs. 33e and f).

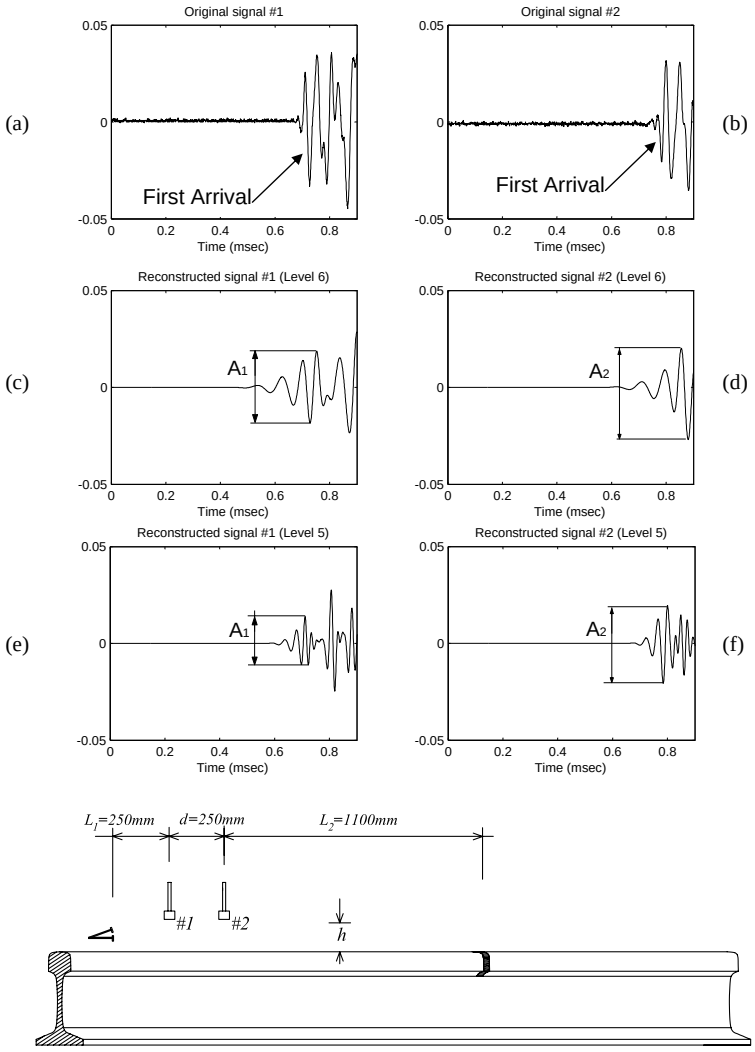


Figure 33 – Crack detection by transmission approach using either level 5 or level 6 DWT decomposition (both sensors prior to defect).

For most signals, best results were obtained by linearly combining level 5 and level 6 reconstructions following eq. (16). An example is given in Fig. 34 where both sensors are positioned prior to the 100% transverse head crack. It can be seen that such combination reconstructs the original signals very accurately. The two sensors detect roughly the same energy (T~1), with any difference being due to slight misalignments.

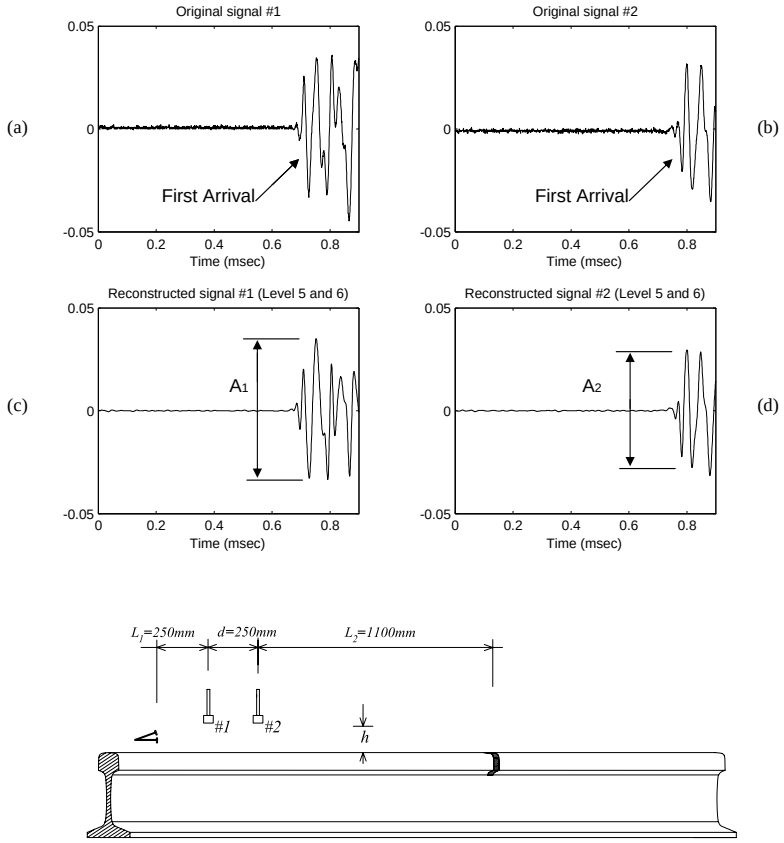


Figure 34 – Crack detection by transmission approach combining level 5 and level 6 DWT decomposition (both sensors prior to defect).

The results for sensors positioned across the crack are shown in Fig. 35 where the expected drop in transmission coefficient can be seen in both the original signals and in the DWT reconstructed signals. When both sensors are positioned past the crack, Fig. 36, the transmission coefficient is restored to its initial, unit value as the two sensors detect the same energy.

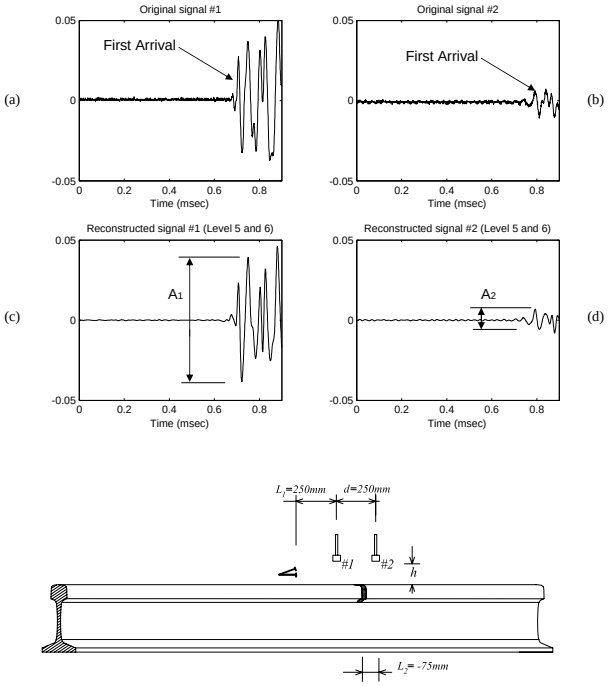


Figure 35 – Crack detection by transmission approach combining level 5 and level 6 DWT decomposition (sensors across defect).

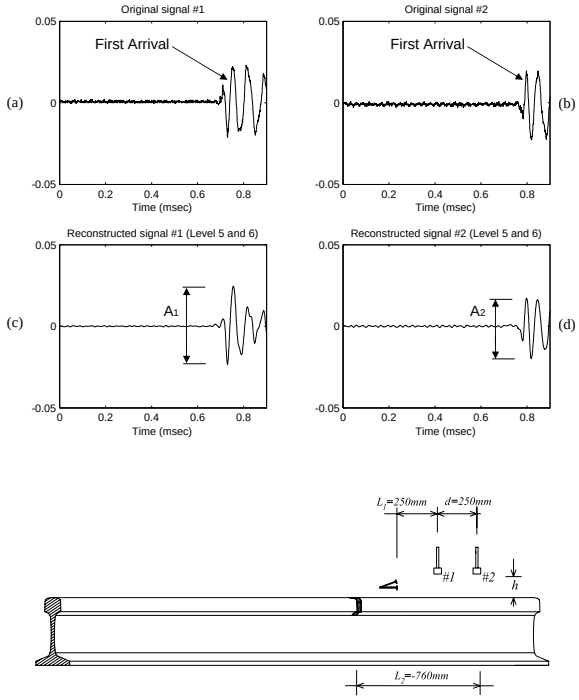


Figure 36 – Crack detection by transmission approach combining level 5 and level 6 DWT decomposition (sensors past defect).

In analogy with the results in Figs. 28 and 29, transmission coefficients were extracted using DWT analysis for the 100% head cracks at the four different orientations of perfectly-transverse, 10º oblique, 20º oblique and 35º oblique. The results are summarized in Fig. 37 considering a linear combination of DWT levels 5 and 6 for all cases. The cracks are clearly detected when the sensor pair moves across them as T drops dramatically. T then returns to its nominal value of 1 when the sensors pass the defect.

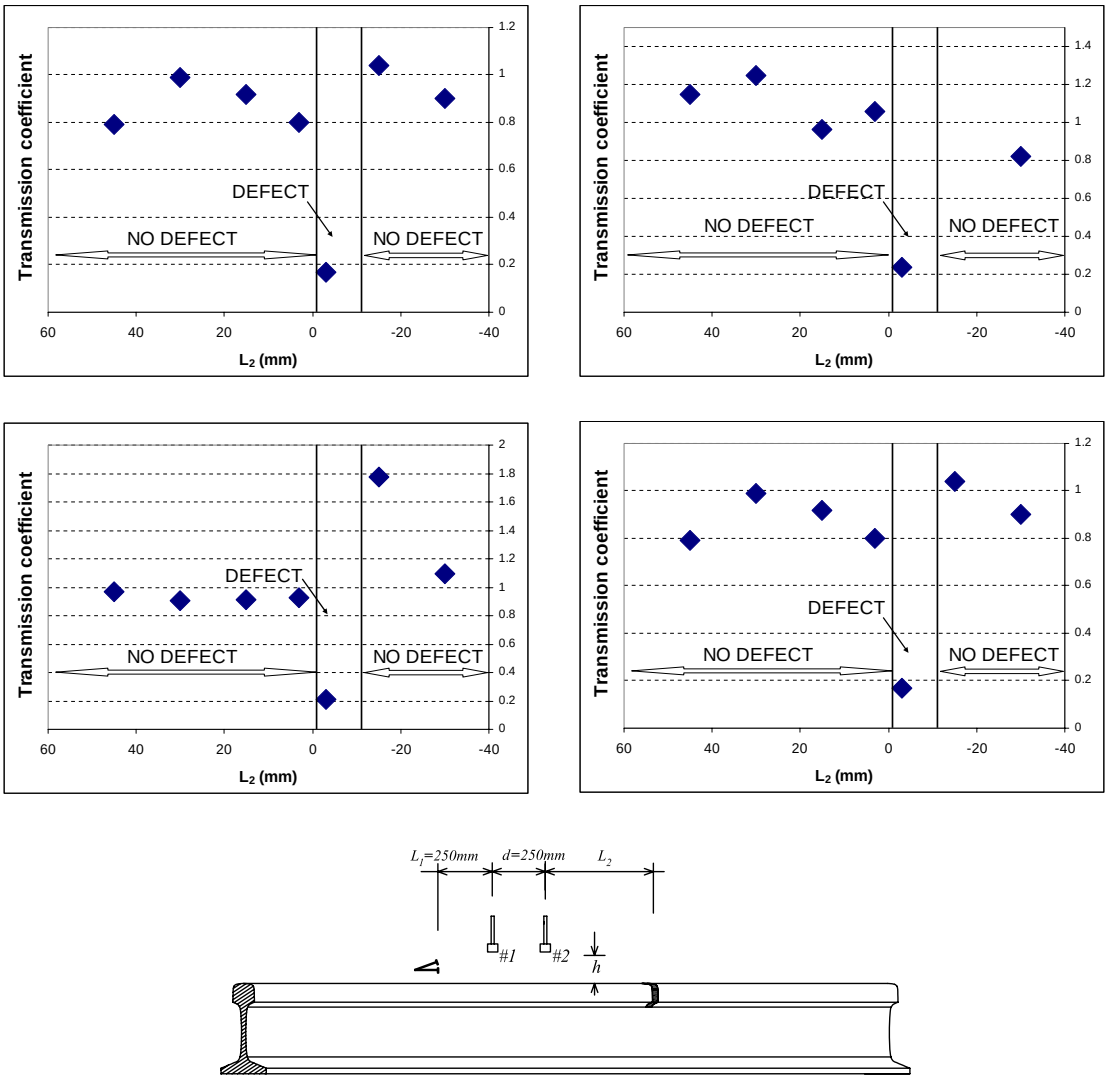


Figure 37 - Variation of the DWT-computed transmission coefficient for head cracks as a function of sensor position relative to the defect (top left: perfectly transverse crack; top right: 10º crack; bottom left: 20º crack; bottom right: 35º crack).

Although the overall trends in this figure are the same as those shown previously in Figs. 28 and 29, the main difference here is the defect detection speed. In fact, the CWT analysis resulting in Figs. 28 and 29 cannot be performed in real-time. On the contrary, the DWT analysis of Fig. 37 is done in real-time owing to its computational efficiency and data compression abilities. Moreover, while the DWT analysis is useful for detecting large cracks,

it becomes essential for detecting smaller cracks whose signatures are much more subtle and thus necessitate of heavy de-noising. This will be demonstrated in the next section.

2.2.5 Summary and Conclusions of the Hammer Generation/Air-coupled Detection Study

This section described an experimental study on the interaction of low-frequency (< 50 kHz) guided waves with large transverse cracks in the rail head. The waves were excited by an impulse hammer and were detected by air-coupled sensors that provided non-contact testing capabilities. The main advantages of the proposed inspection technology include 1) an inherent sensitivity to transverse defects despite the presence of surface shelling, 2) an increased inspection speed due to non-contact conditions and large inspection ranges, and 3) an increased defect detection reliability due to a dual detection scheme based on both reflection and transmission measurements. The reflection measurements allow the detection of a defect located far away from the sensors. As the inspection system travels along the rail and reaches the defect, the latter can be detected by a drop in transmission efficiency between the two sensors. The method provides redundancy to the inspection.

Reflection and transmission coefficients were extracted for large head cracks at four different orientations simulated in the laboratory. The effect of sensor orientation was investigated to determine the optimum angle for defect detection and the system's tolerance to vibrations. Theoretically, best results for reflection measurements should be obtained with the sensors oriented towards the defect (i.e. away from the generation hammer), whereas the opposite should be true for transmission measurements. In the frequency range of interest for the detection of large defects (DC-50 kHz), it was found that a good compromise is to orient the sensors parallel to the rail head. This configuration achieves satisfactory defect detection sensitivity in both reflection and transmission modes. The same tests concluded that variations of sensor orientation between -10^{o} and 10^{o} are well tolerated, indicating that the defect detection system would be robust to accidental sensor misalignments in the field such as those produced by strong vibrations.

The effect of sensor lift-off distance (distance between the sensors and the top of the rail head) on the defect detection sensitivity was also investigated. It was found that a lift-off between 76 mm (3") and 90 mm (3.5") should be selected for obtaining a stable reflection from the large transverse defects in the DC - 50 kHz frequency range. These distances satisfy the clearance envelope of 65 mm (2.5") generally recommended for new rail inspection sensors.

The use of the Discrete Wavelet Transform (DWT) was demonstrated to enhance the defect detection sensitivity in both reflection and transmission measurements. The enhancement results from the unmatched capabilities of the DWT for signal de-noising, signal compression, and processing speed. The result is that clear defect signatures can be obtained in real-time even in a noisy environment.

The ultimate inspection speed achievable by the proposed technology will depend on the length of rail covered in one test ("inspection range"). For reflection measurements, the inspection range depends on the size of the defect, shorter ranges for smaller defects as higher

frequencies are needed. The limited length of rails tested in the laboratory did not allow determining maximum ranges experimentally. However, from the measurement of wave attenuation values it can be concluded that minimum ranges of 10 m (32') can be easily achieved for the detection of large transverse cracks in a reflection mode. For transmission measurements, the inspection range is limited by the distance between the two sensors, *d* in Fig. 24, or, more reliably, *d*/2 in an overlap mode. Detection of a 100% head crack was shown successful for $d = 250$ mm. Larger *d* values would be possible, with the only limitation being wave attenuation which, again, is more severe at high frequencies and can thus affect the detection of small cracks. The general conclusion is that the guided wave inspection allows probing a large portion of the rail at once, thereby dramatically increasing inspection speed. This is in contrast to the conventional rail inspection that is carried out more locally on a single cross-section at a time.

3. Detection of "Small" (< 15% HA) Transverse Defects in Rail Head

This section demonstrates the detection of small, surface-breaking cracks in the rail head by high-frequency (>100 kHz) ultrasonic waves. A pulsed laser was used in conjunction with air-coupled sensors to achieve completely non-contact probing of the rail. Non-contact rail testing by lasers and air-coupled sensors is a very recent advance of the NDE community (Lanza di Scalea 2000; Kenderian et al. 2002; Lanza di Scalea and McNamara 2003; McNamara 2003; Kenderian et al. 2003). However, the drawback of any non-contact testing is a reduced signal-to-noise ratio of the defect detection procedure when compared to the conventional contact testing. The use of signal processing based on the DWT was recently demonstrated to enhance the performance of non-contact rail testing (McNamara and Lanza di Scalea 2004). However, this work presented a cross-sectional rail inspection concept rather than the long-range, guided wave inspection that is the subject of the present study.

In this report the non-contact rail testing technique is refined. First, the inspection probes stay outside the clearance envelope of 65 mm (2.5") from the top of the rail head that is recommended for new rail testing systems. Defects are detected by monitoring both reflection and transmission coefficients measured by a pair of sensors. This dual detection provides robustness and redundancy to the inspection. The proposed system utilizes the DWT processing, discussed in the previous section, to increase the defect detection reliability, the inspection range and the inspection speed.

3.1 Laser Generation / Air-coupled Detection: Experimental Setup and Procedure

The laser used was a pulsed, Nd:YAG operating at 1064 nm with a pulse duration of ~8 nsec and a maximum pulse repetition rate of 20 Hz. Through a system of lenses, the laser beam was focused to a 20 mm-long line to generate guided ultrasonic waves at frequencies >100 kHz traveling along the rail running (longitudinal) direction. At these frequencies, the wave energy is confined to the surface of the rail head (surface waves) and it thus interacts well with surface-breaking cracks oriented transversely to the rail running direction (transverse cracks). The wave detection system was similar to that discussed in the previous sections of the report. It used a pair of air-coupled sensors with a sensor spacing d = 400 mm. The experimental setup can be seen in the photos of Fig. 38 and in the schematics of Fig. 39.

The rail under examination was the usual 115-lb A.R.E.M.A. tee type. Two types of transverse cracks were artificially machined, namely horizontal cuts at the top of the rail head (Fig. 39c), and 45° oblique cuts at the gage-side corner of the rail head (Fig. 39d). The cracks were tested at increasing depths, s in Fig. 39, from 1 mm to a maximum of about 10 mm. For each damage condition the acquisition was repeated ten times. Table 2 summarizes the tests performed.

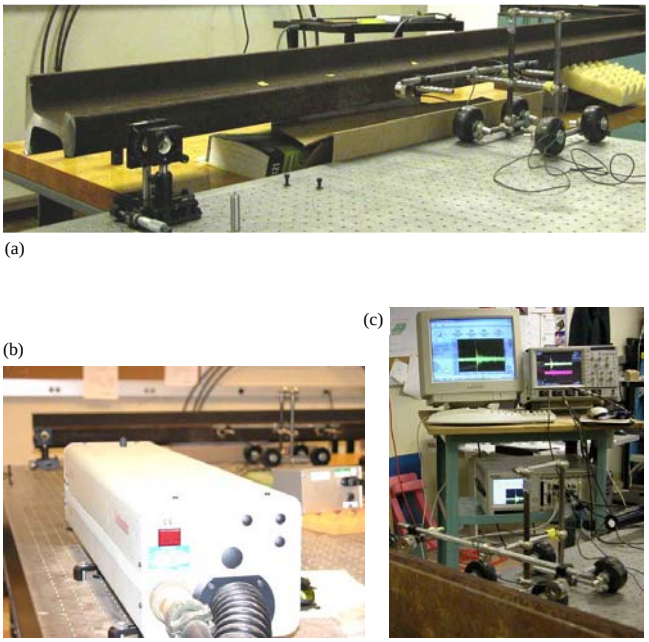


Figure 38 – Laser/air-coupled system for non-contact detection of small, transverse cracks in rail head. (a) overall experimental setup; (b) pulsed laser; (c) sensors with data acquisition system.

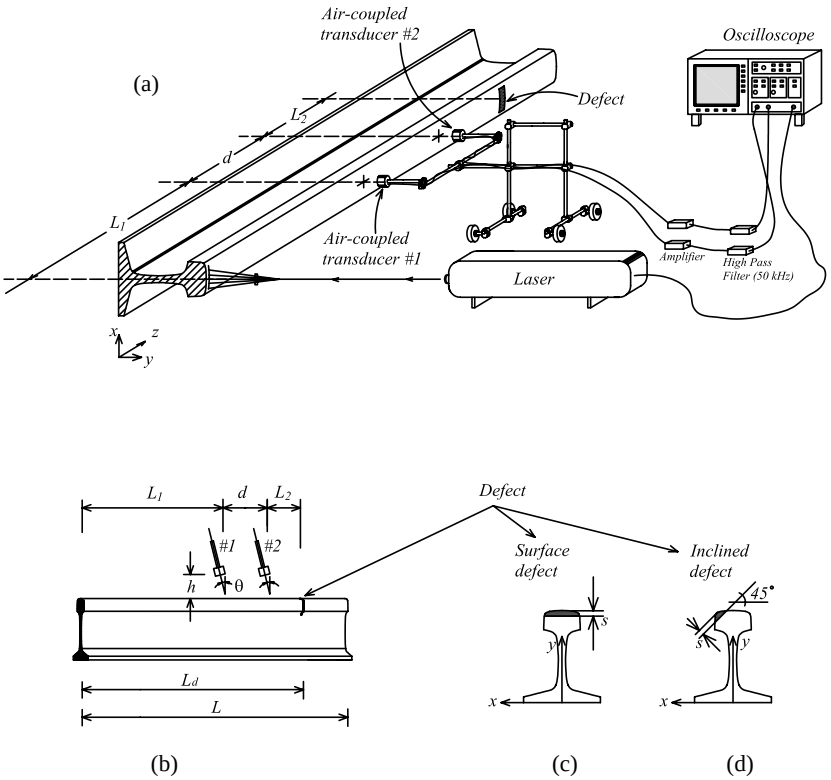


Figure 39 – (a) overall experimental setup; (b) notation of experimental parameters; (c) and (d) types of defects considered.

Table 2 – Tests performed with the laser/air-coupled crack detection system. (θ=sensor inclination angle; L_d=distance laser source to crack; L_1=distance laser source to sensor#1).

Test #	θ	L_d (mm)	L_1(mm)	Notch depth (mm)	Rail length (mm)
1	0°	1500	600	0 to 10	2100
2	"	"	800	"	"
3	"	"	1000	"	"
4	"	"	1200	"	"
5	"	"	1400	"	"
6	+5°, +10°, +15°, +20°	"	"	"	"
7	+6°	1000	500	"	"
8	+6°	"	700	"	"
9	+6°	650	350	"	"
10	-6°	1000	500	"	"
11	-6°	"	700	"	"

The first defect consisted of a straight cut 4 mm wide, with progressive depths monitored up to 6 mm. The notch was machined 1500 mm far from the line of the acoustic generation. The two air-coupled sensors were placed at different distances from the defect. Tests #1-5 probed the defect with L_1 equal to 600, 800, 1000, 1200 and 1400 mm, respectively. In the last two cases sensor #2 was positioned past the defect. The sensor inclination angle was first set to 0°, i.e. sensors parallel to the head top surface. The effect of varying sensor orientations was investigated in Test #6. The objectives of this first set of tests were to verify experimentally:

1. the smallest detectable crack (inspection sensitivity);
2. the maximum distance from the sensors at which the crack could be detected (inspection range);
3. the signal parameters most sensitive to the crack (defect-sensitive features);
4. the effects of sensor orientation on the defect detection sensitivity.

The second defect (Tests #7-8) consisted of a straight cut 4 mm wide, with progressive depth monitored up to 9 mm. The notch was machined 1000 mm far from the line of the acoustic generation. Two distances L_1 were monitored, 500 and 700 mm. The sensor inclination angle was equal to 6°. The objective of this second set of tests was to demonstrate the detection of the waves reflected from the defect (reflection measurements).

The third defect (Test #9) consisted of an inclined cut, 7 mm wide with progressive depth monitored up to 7 mm. The objective of this test was to assess the capability of the system for detecting defects in the gage-side corner of the rail head.

The last set of tests (Tests #10-11) essentially repeated Test #7-8 in a different rail specimen but with a sensor orientation of -6°. This orientation enhanced the detection of the wave incoming from the laser source and it was thus examined for detecting defects by transmission, rather than reflection, measurements.

The nominal lift-off distance used for the sensors was 75 mm (3") from the top of the rail head thus satisfying the recommended clearance envelope mentioned above.

3.2 Laser Generation / Air-coupled Detection: Experimental Results

Fig. 40 illustrates the results from Test #5 with no defect present. Fig. 40a and b show the raw time waveforms detected by the two sensors. In these plots it is difficult to identify the waves, since the high-frequency signals are buried into noise. The corresponding GWT scalograms are illustrated in Fig. 40c and d. The first arrival has a dominant frequency around 200 kHz, demonstrating that the system is indeed able to generate high-frequency waves. The lift-off echo and the echo from the opposite end of the rail section can be also seen. Fig. 40e and f present the result of DWT processing of the raw time signals. These DWT reconstructions were obtained after pruning and thresholding levels 3 and 4 and discarding all wavelet coefficients at other decomposition levels. As before, the db10 mother wavelet was used. The DWT processing effectively removes the noise and makes signal identification possible. The de-noising is performed in real-time. The results clearly demonstrate the necessity for DWT processing of the raw signals obtained by this truly non-contact rail testing system.

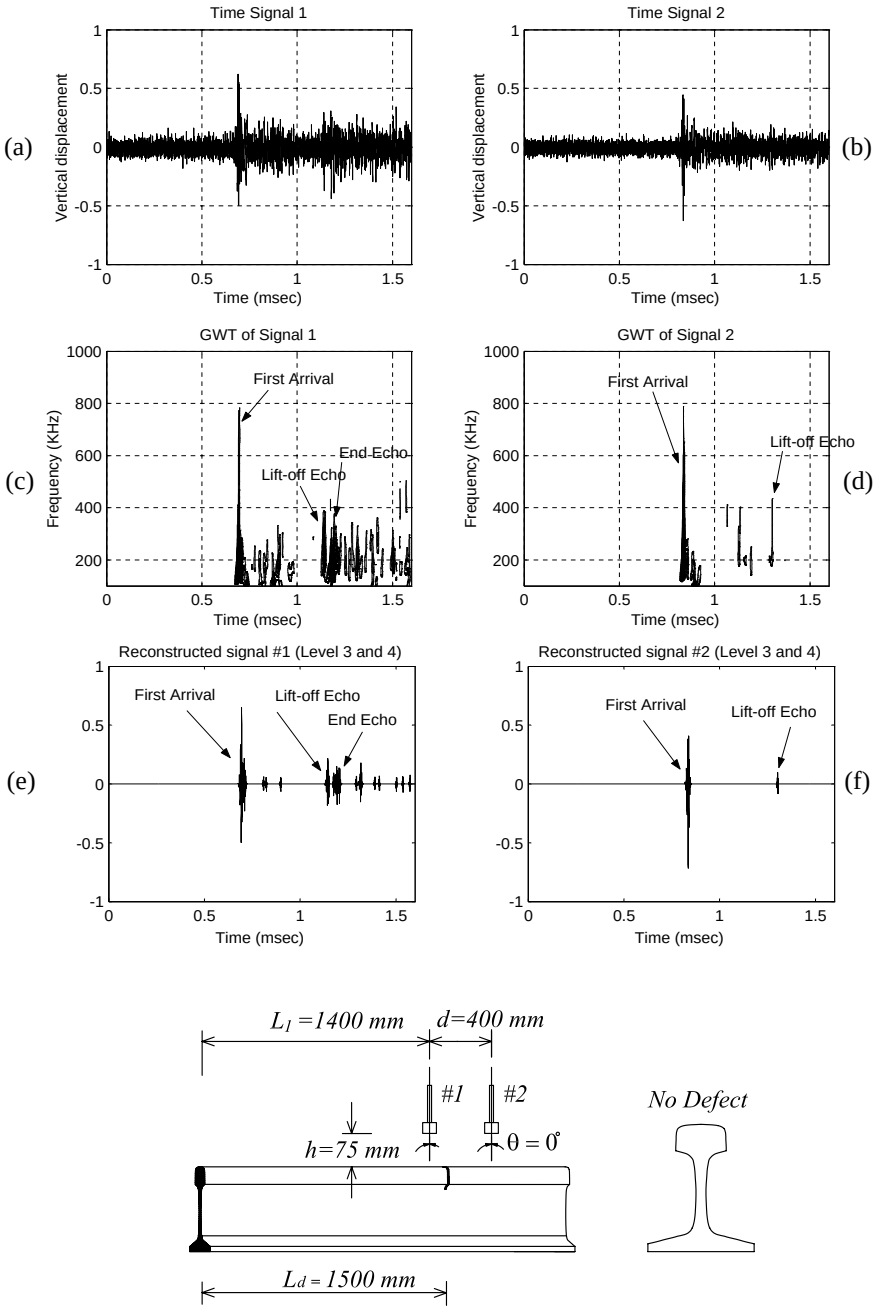


Figure 40 – Laser/air-coupled test (no defect): (a) and (b) raw time waveforms; (c) and (d) continuous wavelet scalograms; (e) and (f) discrete wavelet reconstructions.

3.2.1 Detection of Transverse Cracks at the Top of the Rail Head

Fig. 41 shows the case of a 5 mm deep horizontal cut when the sensor pair is across the defect (Test #5). Again, the DWT processing is needed to identify the signals of interest. From the first arrival and the defect echo at sensor #1 prior to the cut (Fig. 41e), the reflection coefficient can be extracted with the same procedure shown in the previous section. By

considering the first arrival at sensor #2 past the cut (Fig. 41f), also the transmission coefficient can be obtained.

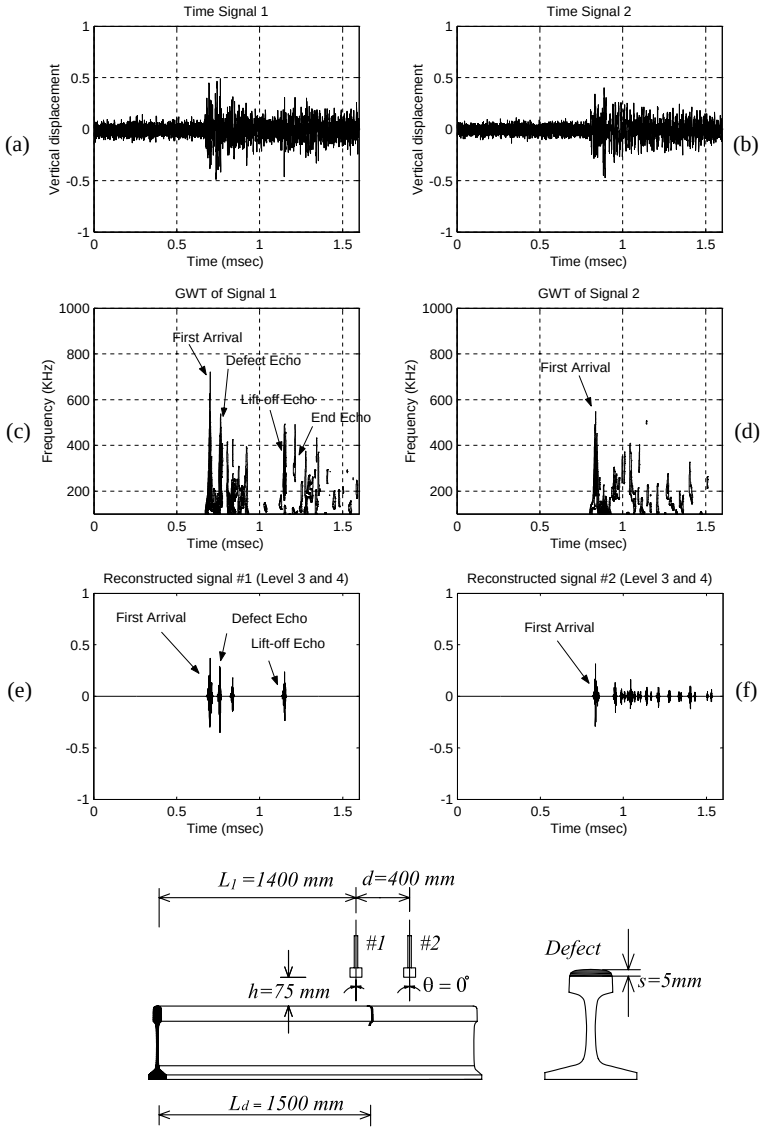


Figure 41 – Laser/air-coupled test (5mm-deep horizontal defect): (a) and (b) raw time waveforms; (c) and (d) continuous wavelet scalograms; (e) and (f) discrete wavelet reconstructions.

With the configuration setup in Test #5, the smallest cut detectable by reflection measurements was 4 mm in depth. In this case, the distance from the laser source to the sensors was the largest allowed by the setup. The results for the 4 mm deep cut are presented in Fig. 42 where only the signals from sensor #1 prior to the defect are shown.

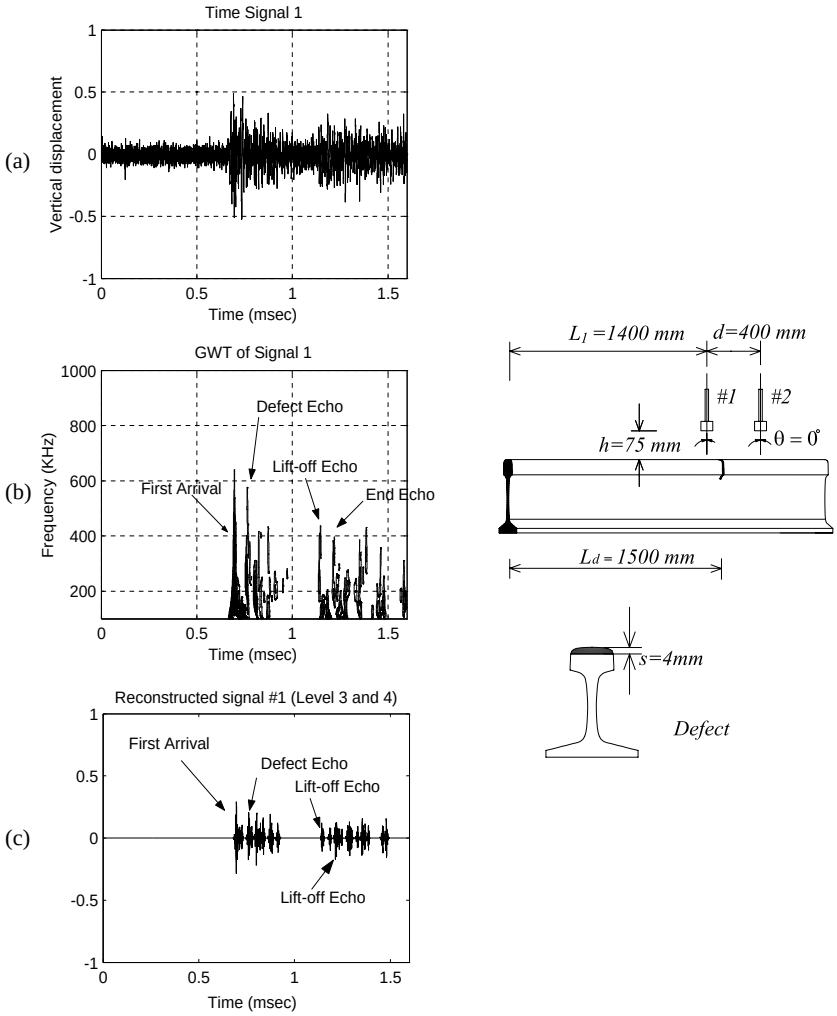


Figure 42 – Laser/air-coupled test (4mm-deep horizontal defect): (a) raw time waveform; (b) continuous wavelet scalogram; (c) discrete wavelet reconstruction.

To determine the maximum inspection range, the results from Test #3 are presented in Fig. 43 for the 5 mm deep notch. In this case the cut is 500 mm away from sensor #1. The defect echo is clearly visible in the DWT reconstructed signal at sensor #1, and reflection coefficients can thus be extracted.

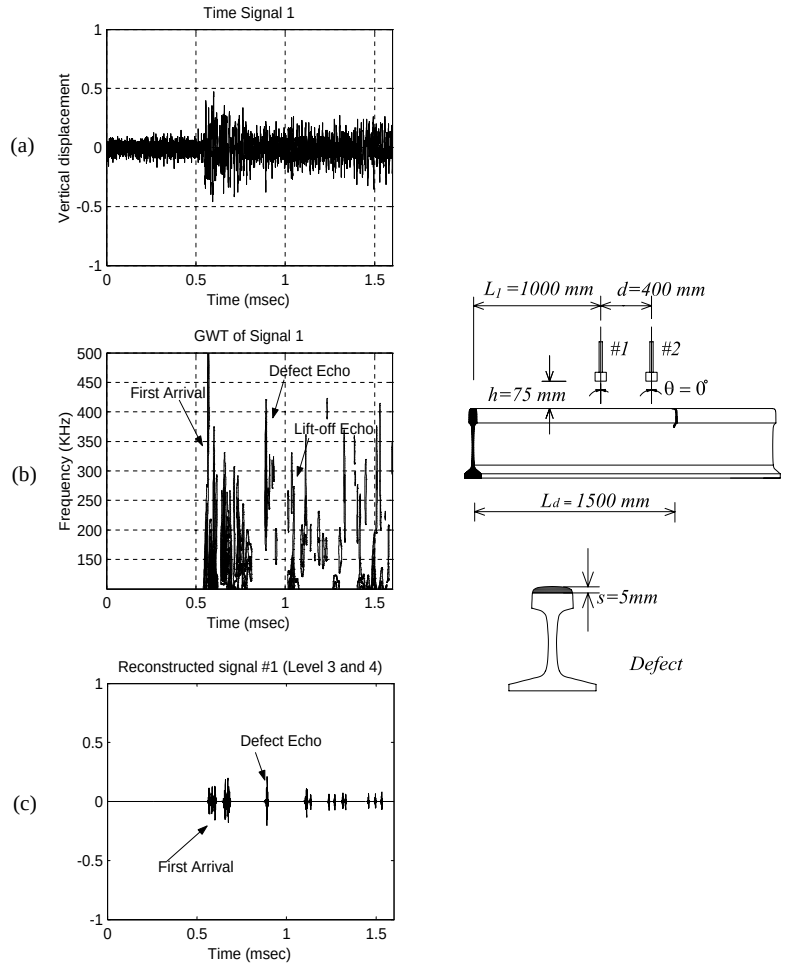


Figure 43 – Laser/air-coupled test (5mm-deep horizontal defect): (a) raw time waveform; (b) continuous wavelet scalogram; (c) discrete wavelet reconstruction.

One of the objectives of the tests was to determine those signal features that are most sensitive to the presence of the defects. These features were extracted from the DWT reconstructed signals using pruning and thresholding at level 4 only. Focusing on transmission measurements, the first feature was the transmission coefficient obtained from the ratio between the maximum amplitude recorded by sensor #2 and that recorded by sensor #1 in the time domain. The second feature considered instead the peak-to-peak amplitudes of these signals to perform the ratio. The third feature considered the areas below the Fast-Fourier Transform (FFT) amplitude spectra to perform the ratio. Fig. 44 compares the three ways of computing the transmission coefficient, *T*, as a function of defect depth in Test #4. The points connected represent the average of ten acquisitions; the vertical bars at each point are the $2\sigma_{dev}$ (standard deviation) deviations. It can be noticed that the *T* values are not lower than one as theoretically expected. A different sensitivity to higher frequencies of the sensors introduces a systematic error on the measurements that, however, does not corrupt the test. What is important is the decrease in *T* as the defect depth increases, since larger defects reflect more energy back towards the laser source. It can be seen that the maximum amplitude feature (Fig. 44b) and the peak-to-peak amplitude feature (Fig. 44b) produce essentially the same results with an almost linear decay of *T* with defect depth. Defect sizing, beyond mere

detection, appears possible by transmission measurements. The FFT area feature (Fig. 44c) overestimates the effect of the 2mm deep cut. In all cases the standard deviations are relatively small, guarantying the repeatability of the experiments. These deviations increase for smaller cut depths due to the more subtle signatures of the shallow defects. It is still remarkable that meaningful data are obtained for cuts as small as 1mm in depth.

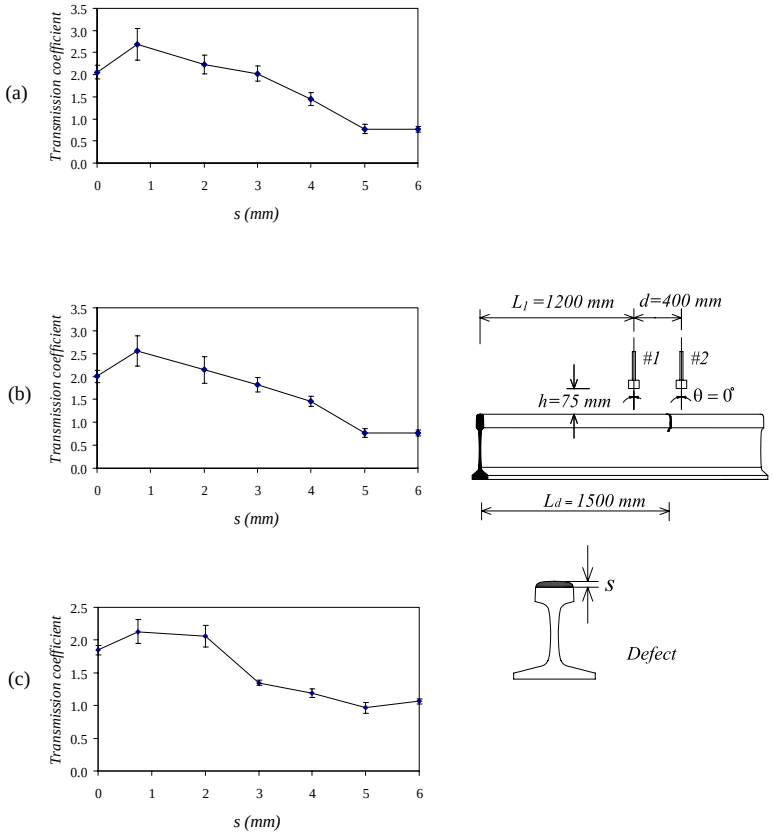


Figure 44 – Transmission coefficients as a function of defect depth computed by using the max amplitudes (a), the peak-to-peak amplitudes (b), and the FFT areas (c) of the DWT processed signals from the sensor pair (Test #4).

Similar plots, shown in Fig. 45, were obtained for Test #5 where the sensors were further away from the laser source. In this case the results are less regular than those obtained for Test #4. This is probably due to the proximity of the defect echo and the first arrival at sensor #1.

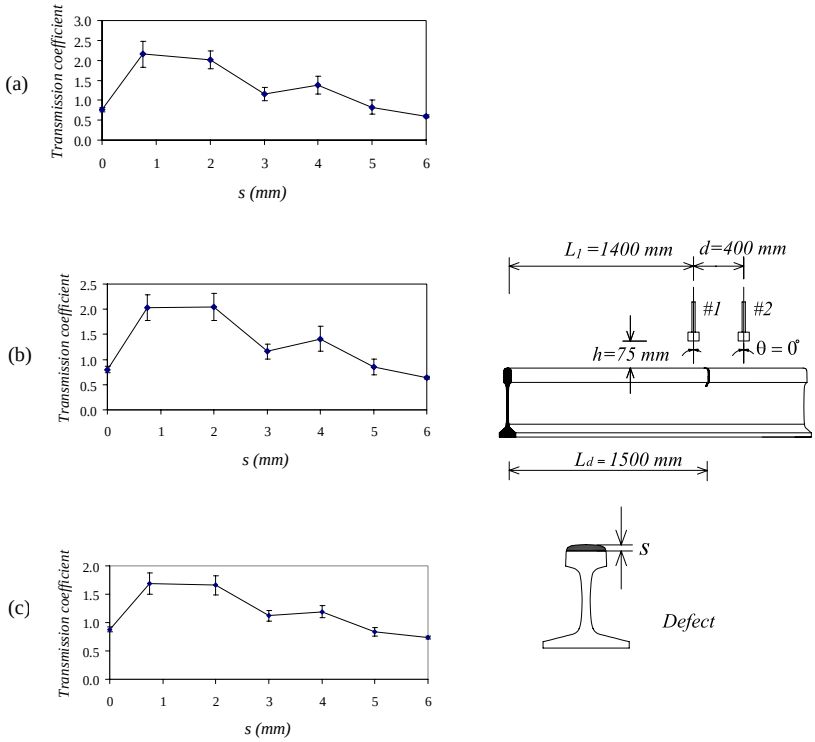


Figure 45 – Transmission coefficients as a function of defect depth computed by using the max amplitudes (a), the peak-to-peak amplitudes (b), and the FFT areas (c) of the DWT processed signals from the sensor pair (Test #5).

Figs. 44 and 45 indicate that it is possible to achieve a quantitative detection of very small, surface-breaking defects by monitoring the transmission coefficient between two non-contact sensors. The setup adopted in these figures, however, was not optimized for the simultaneous detection of the incoming signal and of the defect echo. In fact, as discussed in section 2.2.1, Snell's law dictates that the ideal air-coupled sensor orientation to maximize the sensitivity to the incoming wave is θ=6.3°. Test #6 confirmed this theoretical prediction. The sensors were placed at the same position as in Test #5, and their orientations were varied at angles 5°, 10°, 15° and 20°. The results, not shown here, demonstrated that an inclination angle of 5° provided best sensitivity to the first arrival signals. This is thus the optimum orientation for detecting the defects by transmission measurements.

The remaining set of tests (Tests #7-11) was performed with inclination angles of $\theta = \pm 6°$ where the positive and negative values are assigned to sensors facing the defect and sensors facing the laser source, respectively (same convention used in section 2.2.1). Fig. 46 shows the results from Test #7 for $\theta = +6°$ with no defect present. The first arrivals are detected with a low sensitivity since the sensor orientations do not favor the incoming signals.

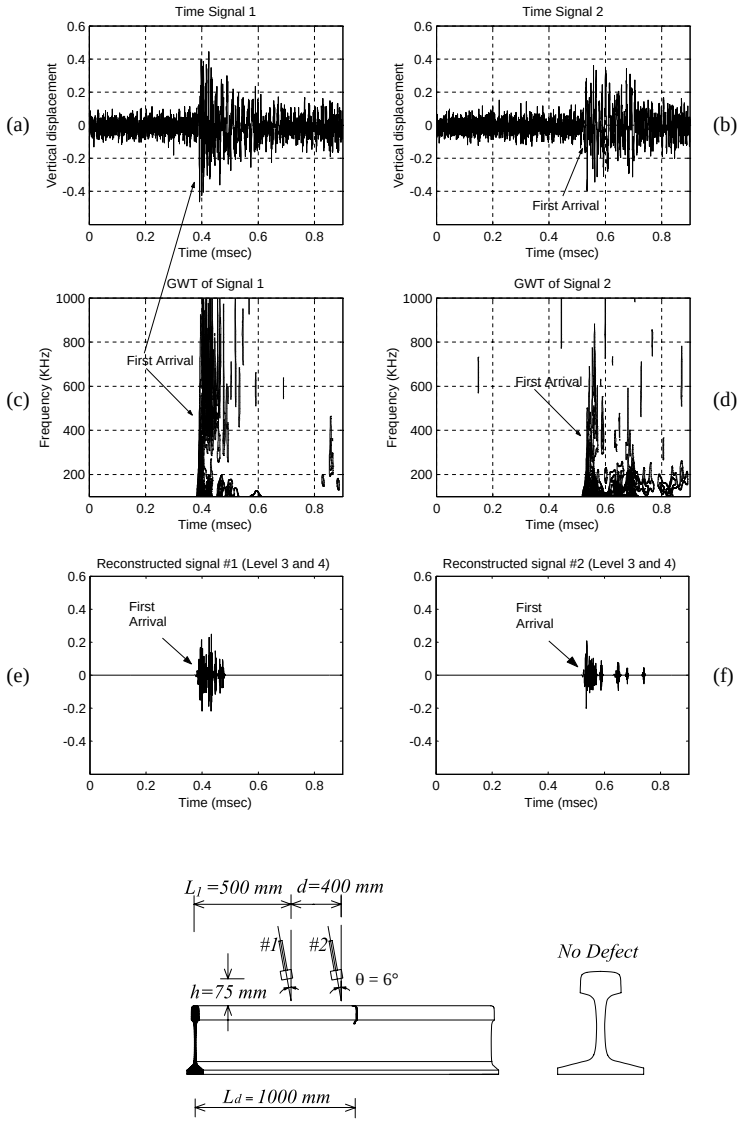


Figure 46 – Laser/air-coupled test (no defect): (a) and (b) raw time waveforms; (c) and (d) continuous wavelet scalograms; (e) and (f) discrete wavelet reconstructions (Test #7).

The case of a 1mm deep cut is shown in Fig. 47. The +6° orientation is clearly effective in detecting the reflection from such a small defect after DWT processing. The defect reflection is actually visible in both sensor #1 further away from the cut, and sensor #2 closer to the cut. The DWT pruning in this case used decomposition levels 3 and 4 that corresponded to frequencies of 427 kHz and 213 kHz, respectively. These high frequencies were the most sensitive to the shallow cut.

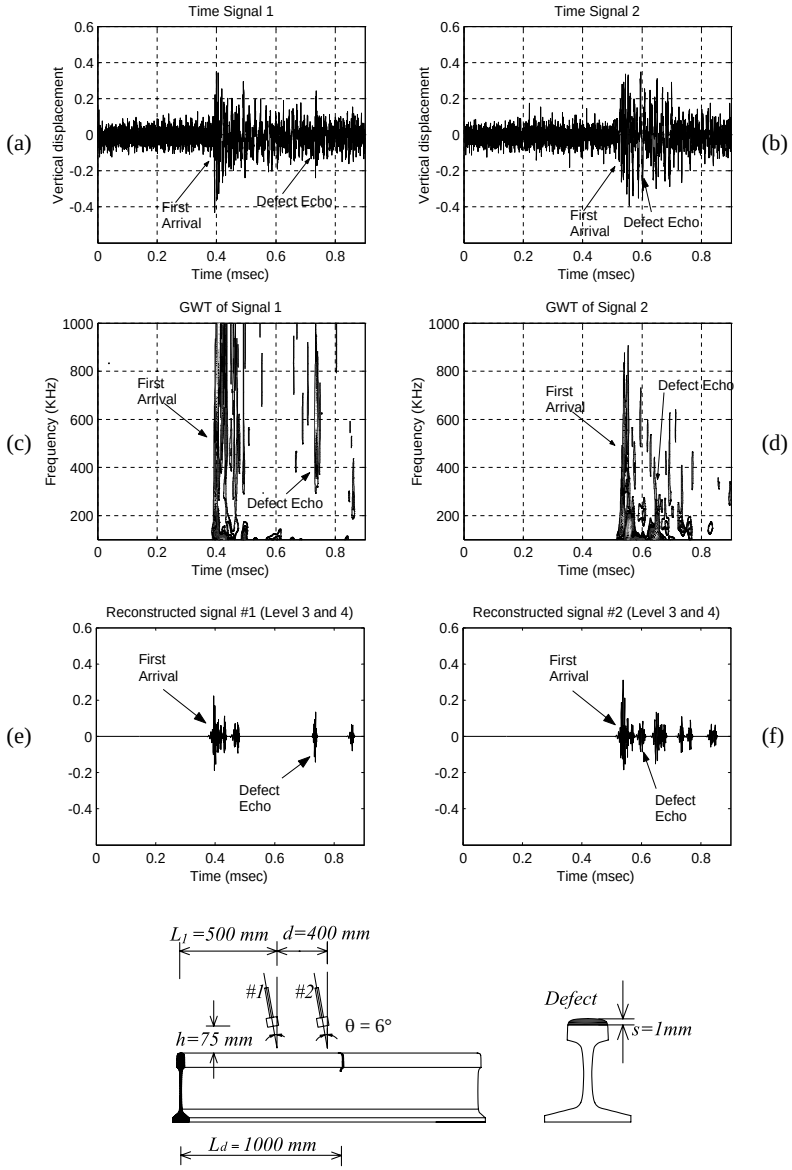


Figure 47 – Laser/air-coupled test (1mm-deep horizontal defect): (a) and (b) raw time waveforms; (c) and (d) continuous wavelet scalograms; (e) and (f) discrete wavelet reconstructions (Test #7).

The case of the 5mm deep cut is presented in Fig. 48. Here the defect reflections are extremely clear due to the larger defect size.

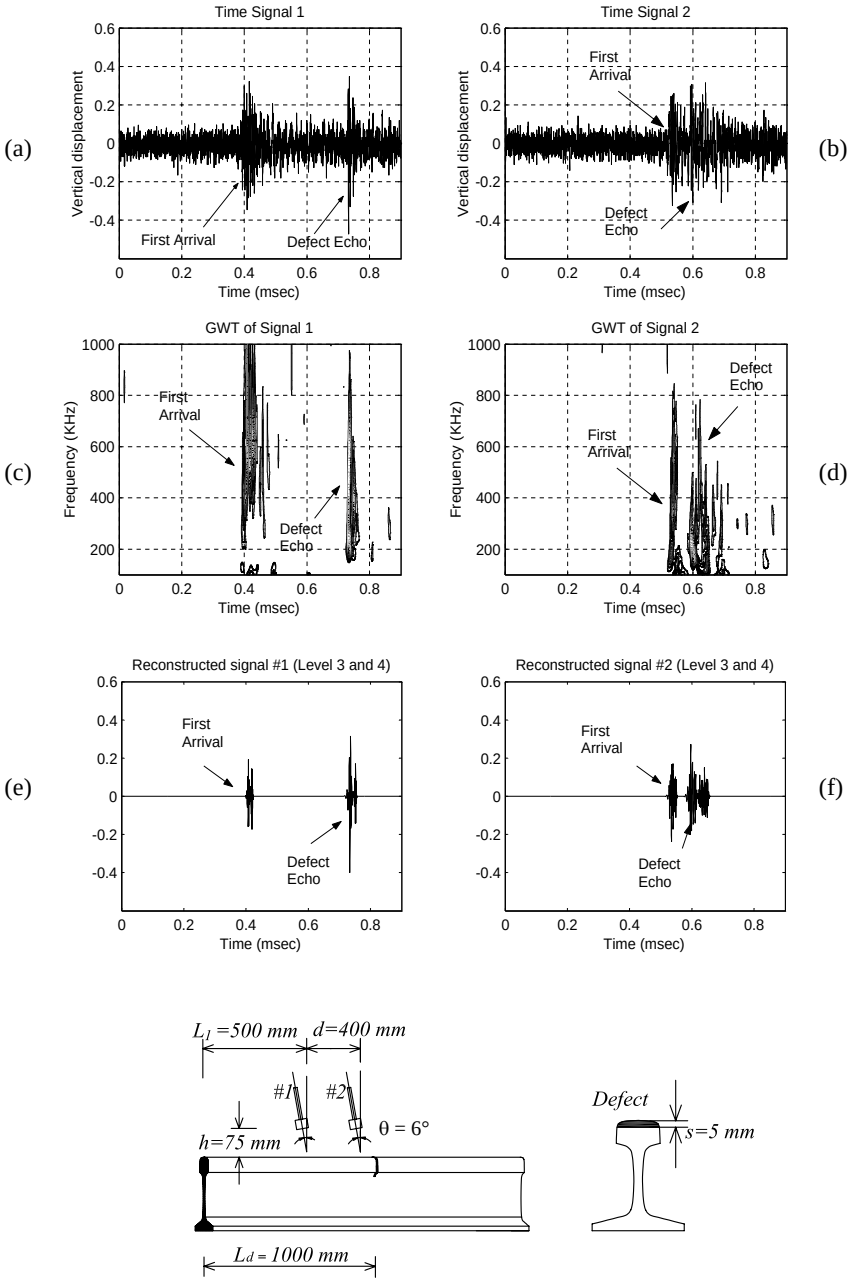


Figure 48 – Laser/air-coupled test (5mm-deep horizontal defect): (a) and (b) raw time waveforms; (c) and (d) continuous wavelet scalograms; (e) and (f) discrete wavelet reconstructions (Test #7).

Fig. 49 plots the normalized amplitude of the defect reflections recorded by sensor #1 as a function of signal frequency and defect depth. The points are average values recorded in ten measurements. It can be seen that the smaller defect depth of 2 mm reflects mostly at 600 kHz. The larger defect depths of 5 mm, 7 mm and 9 mm produce appreciable reflections also at the lower frequency value of 200 kHz. These result stress the important point that higher frequencies are needed to detect smaller defects. Hence the choice of the DWT levels to consider for the pruning and thresholding procedures must be made accordingly.

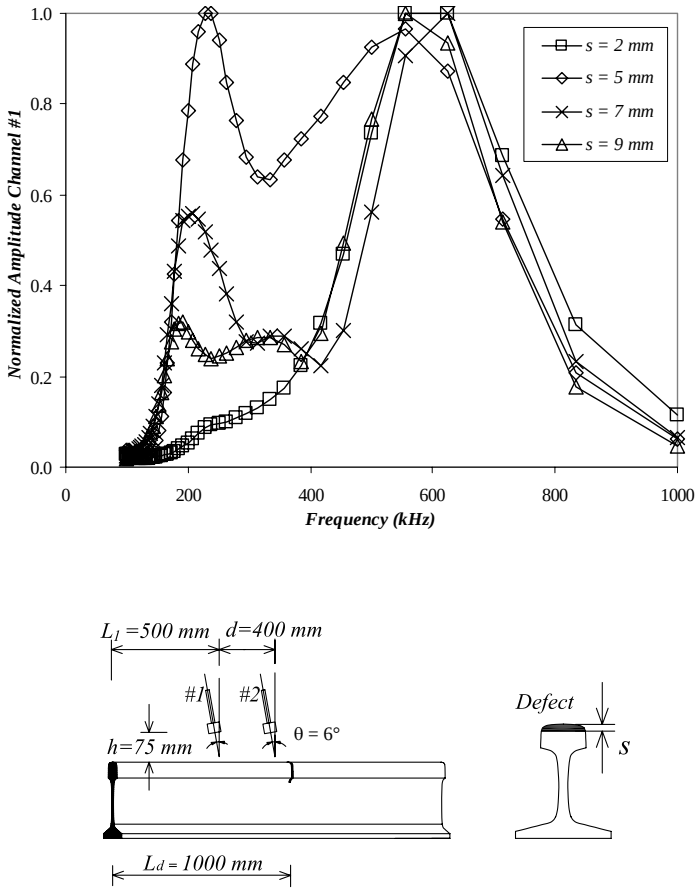


Figure 49 – Normalized amplitude of defect reflections at sensor #1 as a function of frequency for varying defect depths (Test #7).

By taking the ratio between the defect reflection and the first arrival in the DWT processed signals, the reflection coefficient, *R*, can be extracted. This value is an absolute parameter in that it is not affected by variations in the laser generation power. Fig. 50 shows the *R* values computed from sensor #1 as a function of signal frequency and defect depth. Values larger than one are a consequence of the sensor orientation chosen that favors the detection of the defect reflection over that of the first arrival. As before, the points are averages over ten acquisitions. It is confirmed that strong reflections occur at 600 kHz for the smaller defect (2 mm) and at 200 kHz for the larger defects (5 mm and 7 mm).

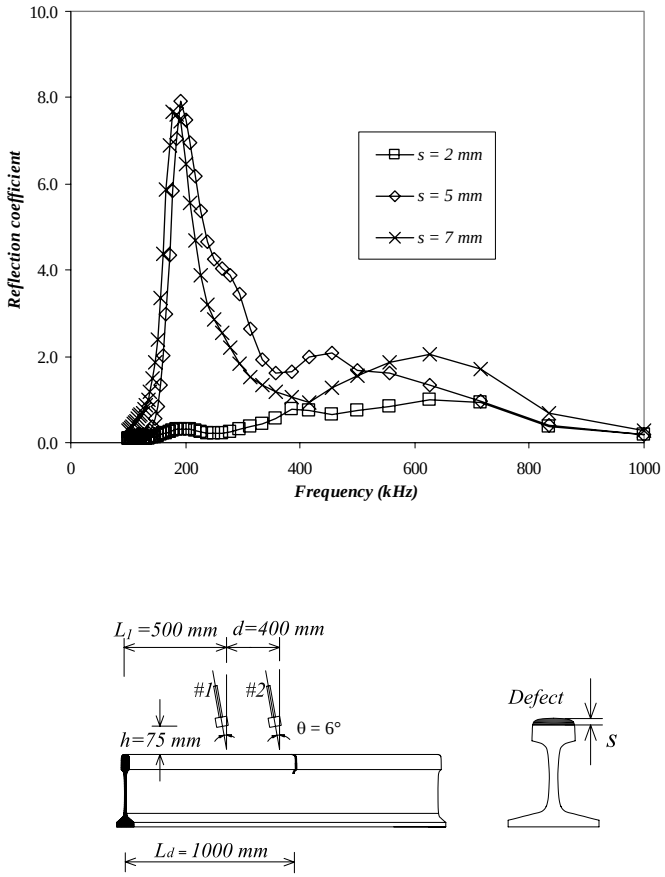


Figure 50 – Reflection coefficients at sensor #1 as a function of frequency for varying defect depths (Test #7).

A different representation of the same tests is presented in Fig. 51 where *R* is plotted against defect depth. Each curve here refers to the same frequency. This plot remarks, once more, that higher frequencies are sensitive to smaller defects.

In summary, Figs. 50 and 51 suggest that quantitative detection of small, horizontal defects can be achieved by monitoring the *R* values obtained from air-coupled sensors.

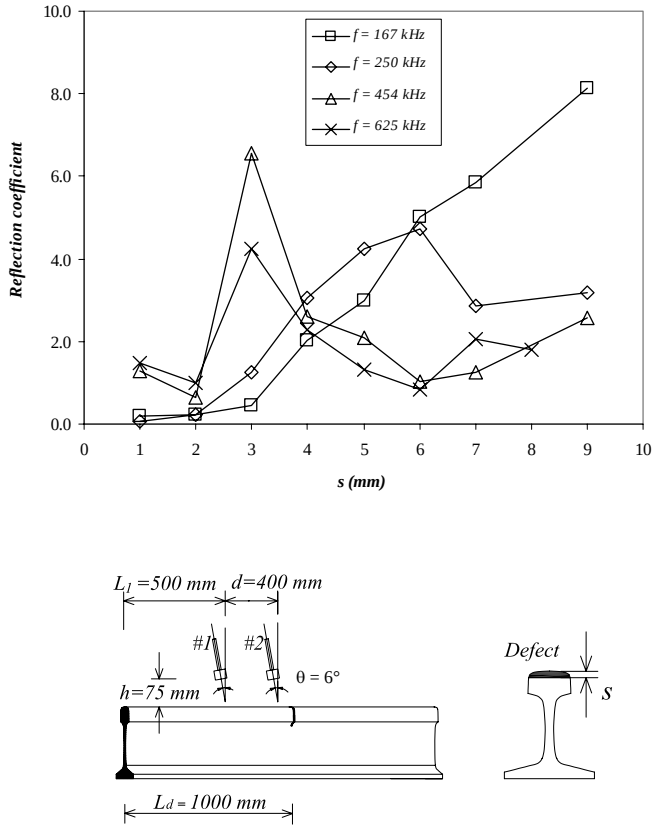


Figure 51 – Reflection coefficients at sensor #1 as a function of defect depth for varying frequencies (Test #7).

3.2.2 Detection of Transverse Cracks at the Gage Corner of the Rail Head

The case of Test #9 for the 45° oblique cut at the gage-side corner of the rail is shown in Figs. 52-56. In these figures the sensors were positioned across the cut. Sensor #1 was at a distance of 300 mm from the cut. The sensor orientation of 6° was used in these figures to maximize detection of the echo.

For a cut depth of 1 mm (Fig. 52), the defect echo is visible in the DWT reconstructed signal at sensor #1. By using the signals at both sensor #1 and sensor #2, reflection and transmission coefficients can be extracted with the usual procedure. The ability of detecting a crack of only 1mm in depth and located in corner of the rail head is a remarkable result of the proposed inspection system. Again, it should be stressed how this is possible only after DWT de-noising of the raw signals that makes signal identification possible.

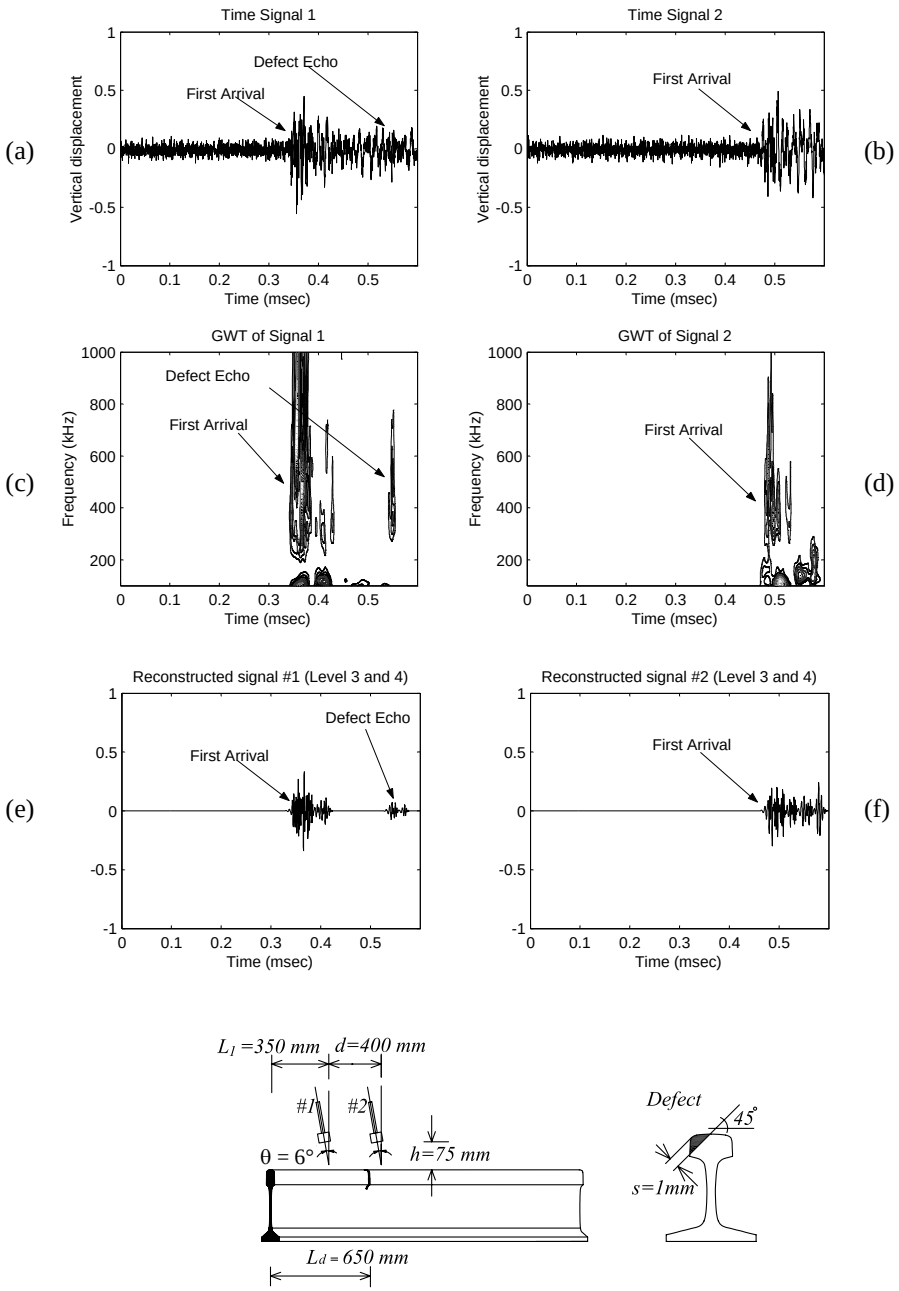


Figure 52 – Laser/air-coupled test (1mm-deep oblique defect): (a) and (b) raw time waveforms; (c) and (d) continuous wavelet scalograms; (e) and (f) discrete wavelet reconstructions (Test #9).

The cases of defect depths 4 mm and 8 mm are shown in Figs. 53 and 54, respectively. It can be seen how the defect reflections in Figs. 52e, 53e and 54e increase with defect size as expected.

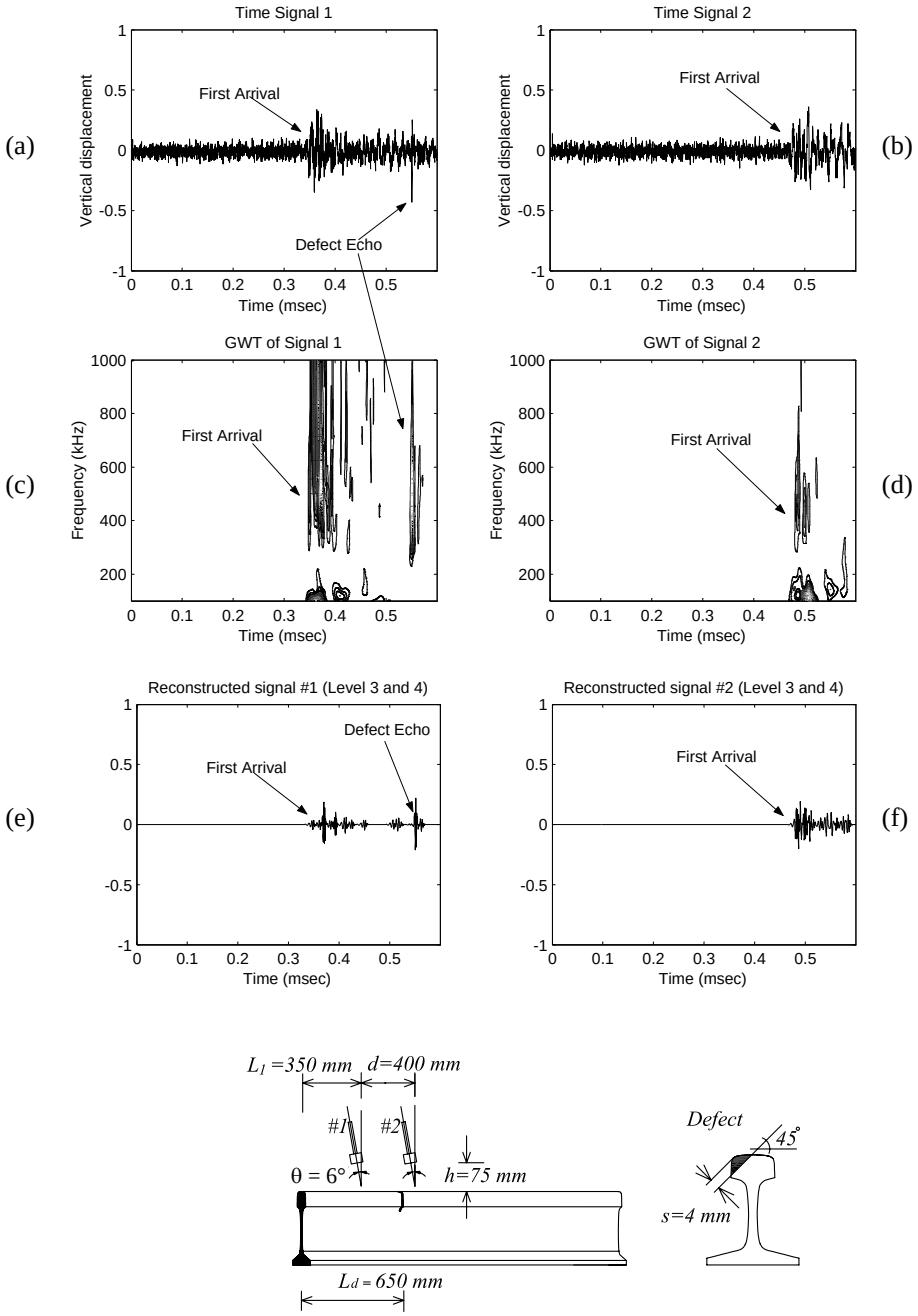


Figure 53 – Laser/air-coupled test (4mm-deep oblique defect): (a) and (b) raw time waveforms; (c) and (d) continuous wavelet scalograms; (e) and (f) discrete wavelet reconstructions (Test #9).

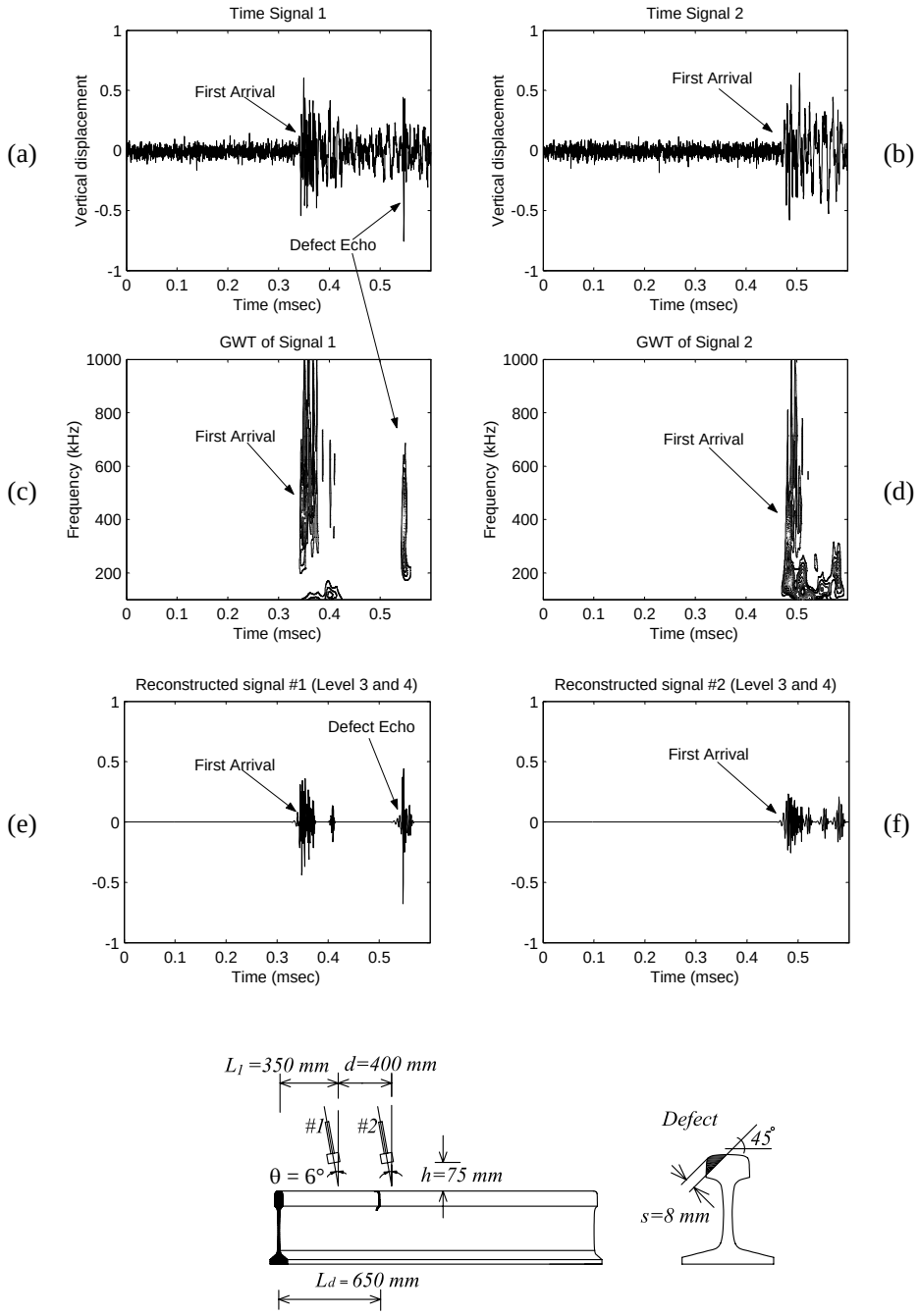


Figure 54 – Laser/air-coupled test (8mm-deep oblique defect): (a) and (b) raw time waveforms; (c) and (d) continuous wavelet scalograms; (e) and (f) discrete wavelet reconstructions (Test #9).

In analogy with Fig. 49 for the horizontal cut, Fig. 55 plots the normalized amplitude of the reflections for the oblique cut recorded by sensor #1 (sensor orientation still 6°). The points are averages over ten acquisitions. Strong reflections in this case are seen around 400 kHz for all defect depths. Also, the primary reflection peak slightly shifts towards higher frequencies with decreasing cut depth. The smaller, 1.5mm deep cut also reflects efficiently at 600 kHz.

The general trend of increasing frequencies with decreasing defect size is confirmed for the gage-side corner cut.

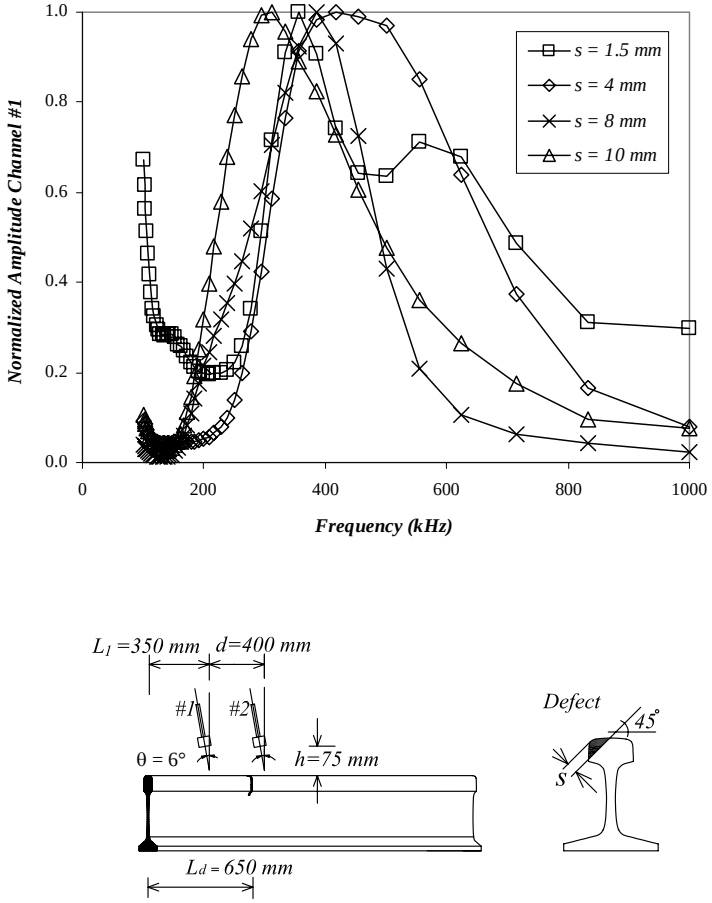
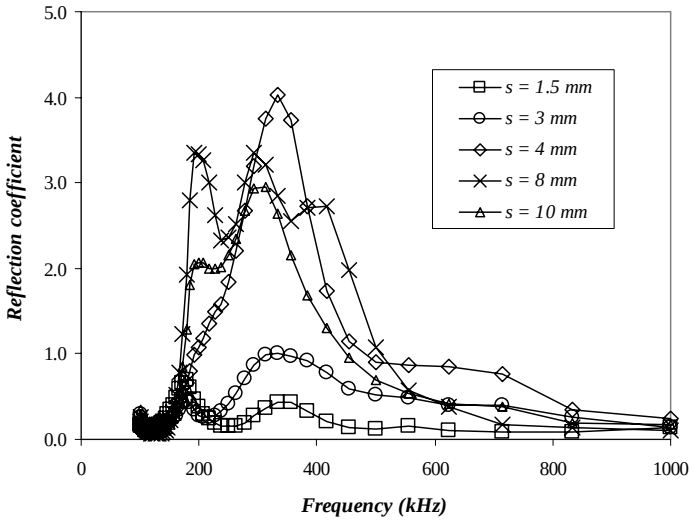
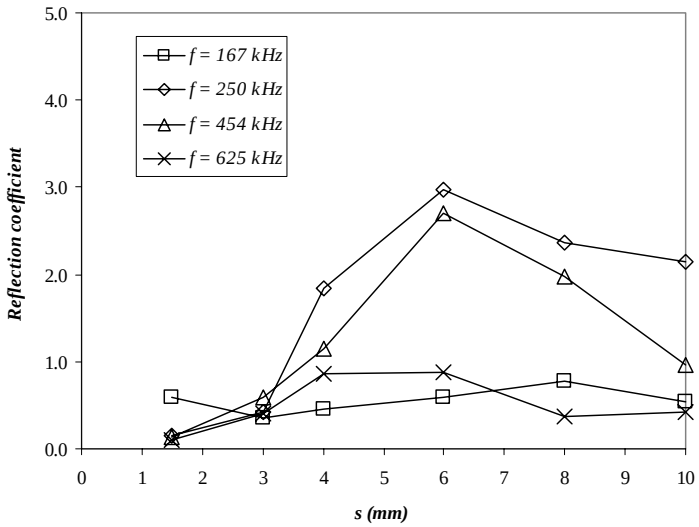


Figure 55 – Normalized amplitude of defect reflections at sensor #1 as a function of frequency for varying defect depths (Test #9).

The reflection coefficient plots as function of frequency and cut depth are presented in Fig. 56. Here the *R* values are shown in the same two representations as in Figs. 50 and 51 for the horizontal cut. It can be seen that *R* generally increases with defect depth as expected. The reflection maxima are in the 250 kHz – 450 kHz range and the larger defect sizes tend to reflect strongly at the lower frequencies.





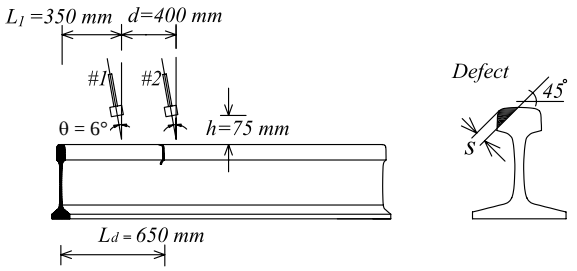


Figure 56 – Reflection coefficients at sensor #1 as a function of frequency for varying defect depths (top graph) and as a function of defect depth for varying frequencies (bottom graph) (Test #9).

As for the horizontal defect, the plots in Fig. 56 can be used to obtain quantitative information on the size of the corner defect by reflection measurements.

3.2.3 Transmission Measurements

Tests #10 and #11 were performed with a sensor orientation of -6°, implying sensors oriented away from the cut and towards the laser source. This configuration favors the detection of the first arrivals (incoming signals) at the expenses of the detection of the defect reflections. Defects can be optimally detected by transmission, rather than reflection measurements.

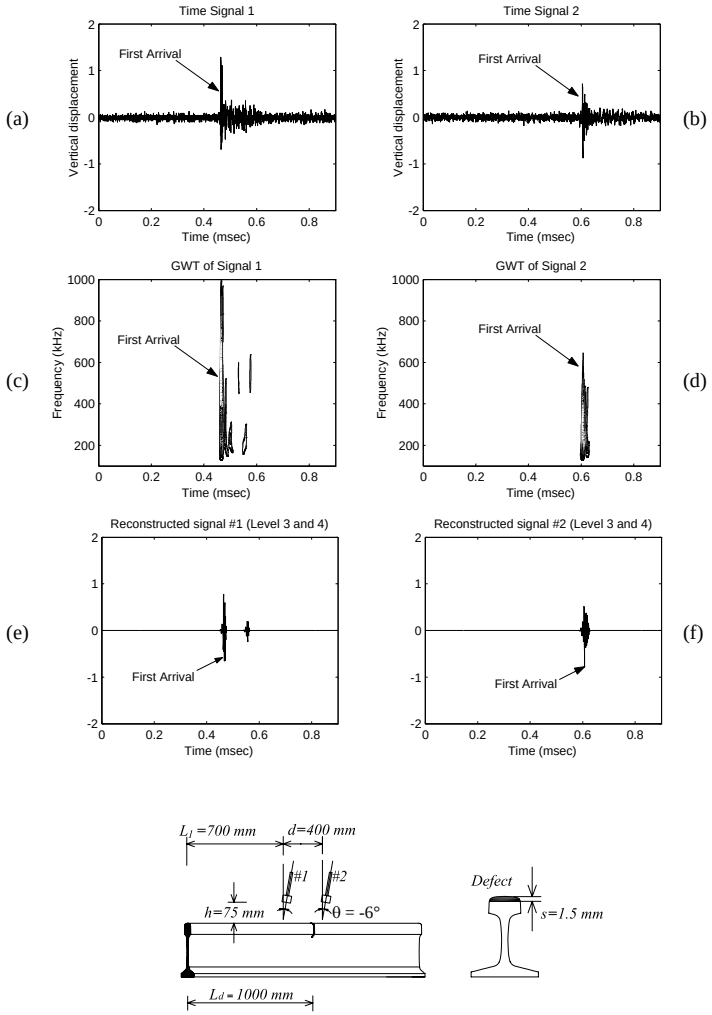


Figure 57 – Laser/air-coupled test (1.5mm-deep horizontal defect): (a) and (b) raw time waveforms; (c) and (d) continuous wavelet scalograms; (e) and (f) discrete wavelet reconstructions (Test #11).

Fig. 57 demonstrates the detection of a 1.5mm deep cut by transmission measurements. It can be seen that the first arrival at sensor #2 past the defect is smaller than that at sensor #1 prior to the defect.

Fig. 58 illustrates the case of a 4mm deep cut where the decrease in sensor #2 signal is more dramatic due to the larger defect size.

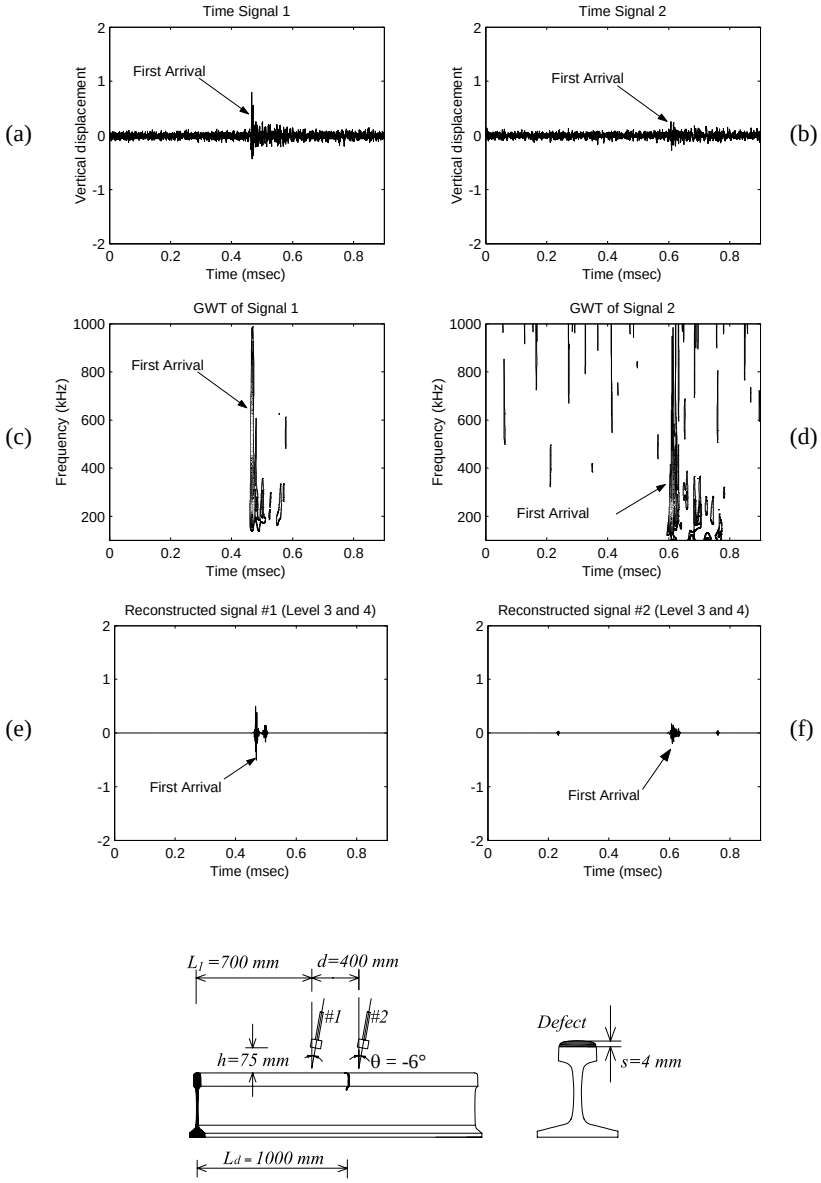


Figure 58 – Laser/air-coupled test (4mm-deep horizontal defect): (a) and (b) raw time waveforms; (c) and (d) continuous wavelet scalograms; (e) and (f) discrete wavelet reconstructions (Test #11).

The case of an 8.5mm deep cut is presented in Fig. 59. The signal at sensor #2 almost disappears.

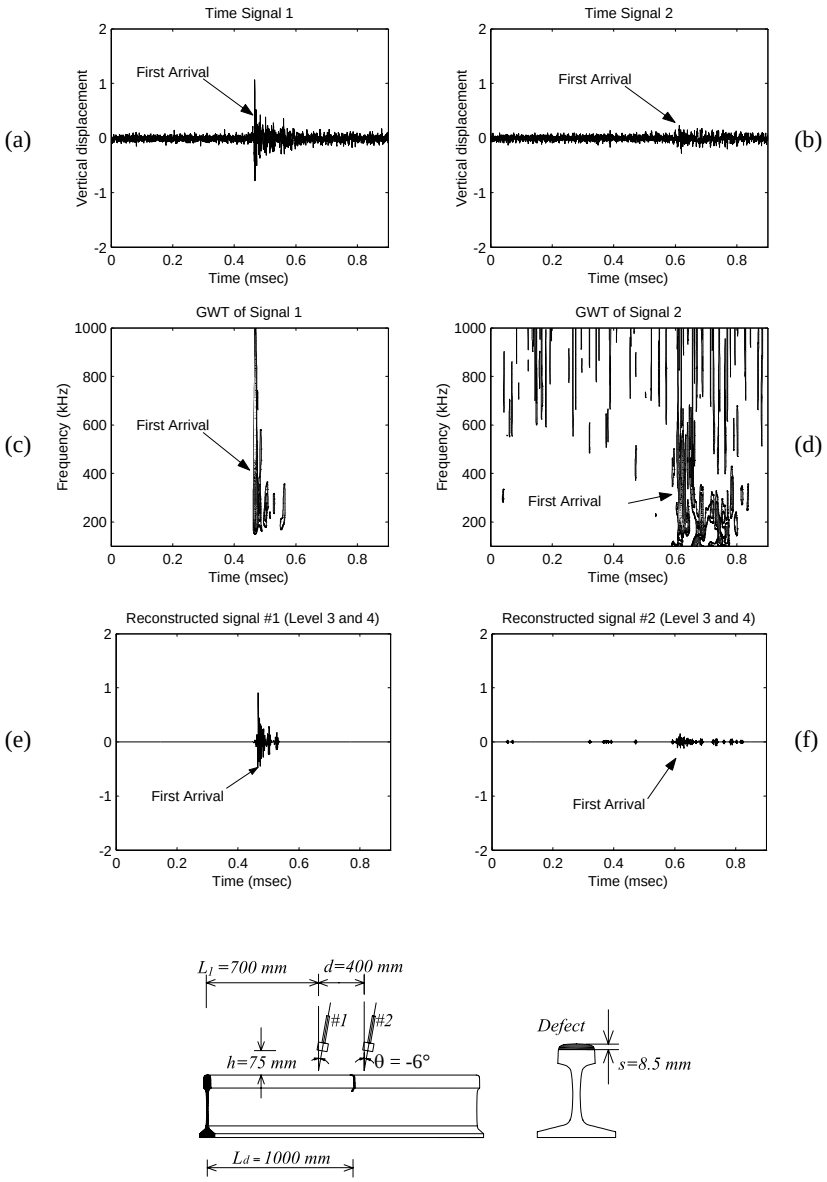


Figure 59 – Laser/air-coupled test (8.5mm-deep horizontal defect): (a) and (b) raw time waveforms; (c) and (d) continuous wavelet scalograms; (e) and (f) discrete wavelet reconstructions (Test #11).

Defect detection can be accomplished by monitoring the classical wave attenuation coefficient (linear attenuation) that is defined by:

$$\alpha[\mathrm{dB/m}] = \frac{10\log\dfrac{A_2(f)}{A_1(f)}}{d} \tag{17}$$

where A_1 and A_2 are the amplitudes of the GWT scalograms of sensor #1 and sensor #2, respectively, and d is the distance between the sensors. This representation allows to appreciate the dependence of the attenuation on the wave frequency content. Fig. 60

summarizes the attenuation coefficients extracted at various frequency values when the sensors are across cuts of varying depths up to 8.5 mm.

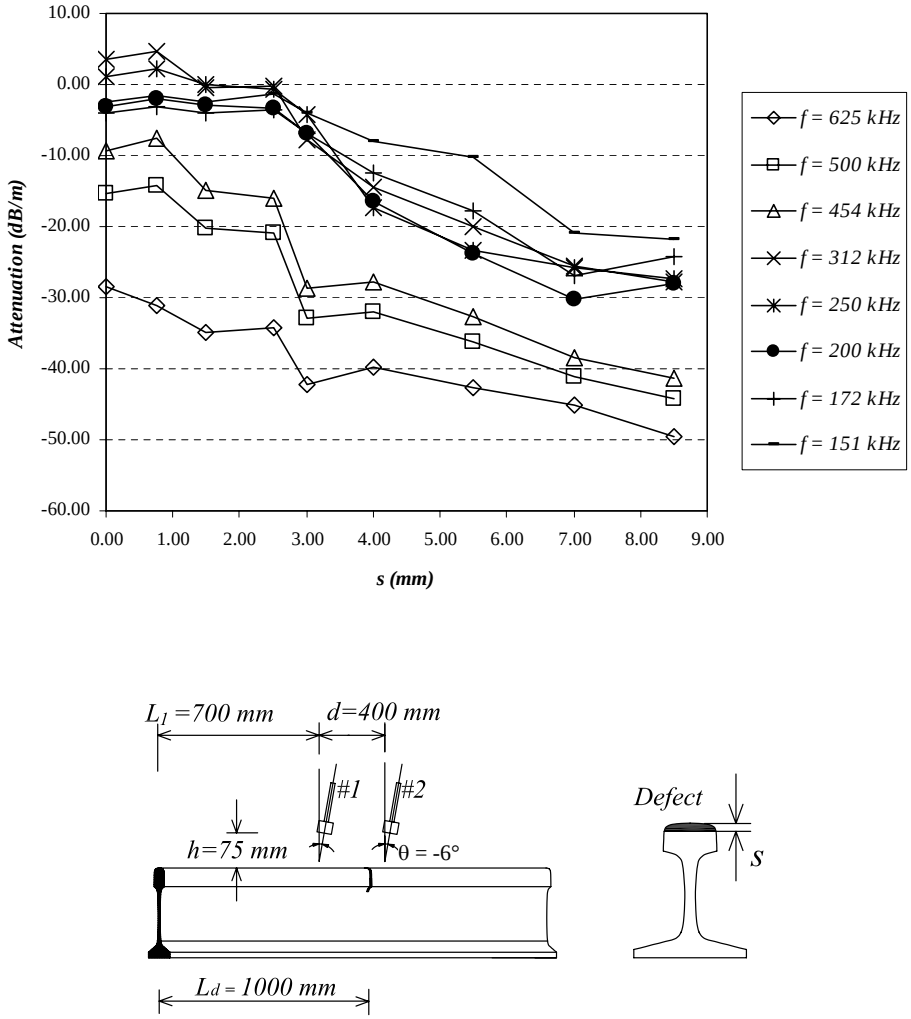


Figure 60 – Linear attenuation coefficient as a function of defect depth for varying frequencies (Test #11).

It can be seen that the attenuation generally increases (in absolute value) with increasing frequency. More importantly, the attenuation also increases with increasing defect size, and this is the phenomenon that can be used for defect detection. For smaller defects (< 2.5 mm in depth), the effect occurs only at the higher frequencies, above 450 kHz. For larger defects (> 2.5 mm in depth), the effect occurs at all frequencies between 150 kHz and 650 kHz, with roughly the same rate of change of attenuation with defect size. What was found for reflection measurements thus remains valid for transmission measurements, that is the importance of probing small defects with high frequency waves.

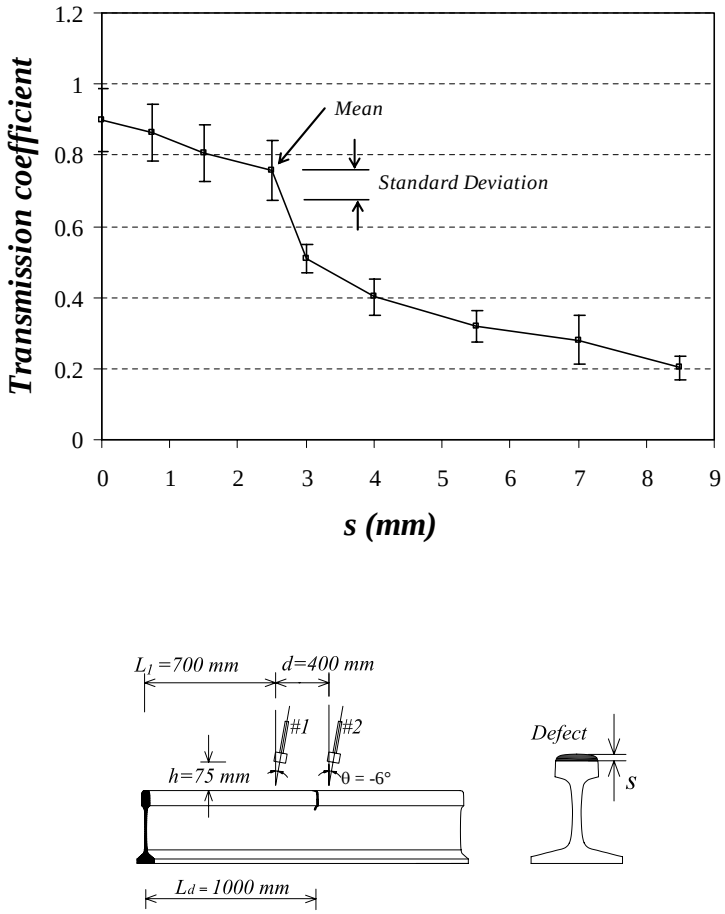


Figure 61 – Transmission coefficient as a function of defect depth obtained by pruning and thresholding DWT decomposition levels 3 and 4 of the first arrivals at the two sensors (Test #11).

The results in Fig. 60 were extracted from the continuous wavelet transforms of the signals from the two sensors. Consequently, while providing important insights on the frequency-dependence of the defect detection procedure, these data cannot be obtained in real-time. By using the DWT processed signals, the transmission coefficient, *T*, can be obtained in real-time at a given frequency band from the usual eq. (14). Fig. 61 summarizes the transmission coefficients measured for increasing cut depths by using the linear combination of DWT levels 3 and 4, corresponding to a frequency band centered at around 320 kHz. A thresholding at 50% of the maximum coefficient value was also applied before extracting *T*. For each defect depth, the plot shows the mean values and the standard deviations obtained over ten acquisitions. The decrease in transmission efficiency with increasing defect depth is evident, even for a cut depth of only 1 mm. The possibility thus exists for quantitative sizing of the surface-breaking defects in a real-time test.

3.3 Summary and Conclusions of the Laser Generation/Air-coupled Detection Study

This section demonstrated experimentally the guided wave technique for the detection of small, surface-breaking cracks in the rail head. In contrast to the detection of large cracks

described in the Section 2, for small cracks high-frequency (>100 kHz) ultrasonic waves are needed. These high-frequency waves were excited by a pulsed laser and were detected by air-coupled sensors. The inspection probes are thus completely non-contact. The sensor lift-off distance adopted for these tests was 75 mm (3") that satisfies the clearance envelope generally recommended for new rail defect detection systems.

Two types of surface-breaking cracks were investigated, namely a horizontal crack at the top of the rail head and a 45° oblique crack at the gage-side corner of the rail head. The cracks were examined at depths ranging from 1 mm to 10 mm. The same two defect detection modes of reflection and transmission measurements proposed in Section 2 for the large cracks were investigated for the small cracks. Generally, a sensor orientation of 6° proved most successful for detecting the cracks. In reflection measurements, best detection was achieved by orienting the sensors towards the defects, i.e. away from the laser generation source. The opposite orientation proved optimum for transmission measurements. Both the horizontal and the oblique cracks were successfully detected at depths as small as 1 mm once Discrete Wavelet Transform (DWT) processing was used to de-noise the defect signatures.

Reflection (R) and transmission (T) coefficients were extracted from the DWT-processed data as a function of crack depth and signal frequency. The results obtained can be used to perform quantitative, rather than qualitative defect detection since clear trends exist for both R and T coefficients as a function of defect size. It was found that frequencies in the range 100 kHz – 600 kHz are most effective to detect these cracks. Also, the optimum detection frequencies generally increase for smaller crack sizes.

The inspection range achievable by the laser/air-coupled system is at least 500 mm (20") for surface-breaking cracks as shallow as 1 mm in reflection measurements. Larger ranges were not tested due to the limitation of the experimental setup adopted. In transmission measurements the range depends on the distance between the two sensors, as remarked in the conclusions of Section 2. A sensor distance of at least 400 mm (16") was successfully used for the detection of the small cracks. These ranges are larger than those used in conventional rail inspection techniques, and would thus result in increased inspection speeds.

4. AUTOMATIC DEFECT CLASSIFICATION

An important step of any new defect detection technology is automatic defect classification. This tool was added to the long-range detection of rail defects, particularly the critical transverse cracks. The classification algorithm was based on support vector machines, the new generation of Neural Networks. This section presents a brief introduction to support vector machines, followed by a description of the procedure used to determine the best data to be used to train the automatic defect classification tool, and concluding with lessons learned during this study. These results were recently published by McNamara et al. (2004).

4.1 Experimental Setup, Procedure and Defects Considered

The experimental data used to train the pattern recognition algorithm were extracted from tests of the type discussed in Section 2 that employed low-frequency waves excited by an impulse hammer for defect detection. The specific data of interest were the reflection coefficient plots corresponding to transverse-type cracks in the rail head of varying sizes and orientations. The reflection coefficient plots were obtained in the 10 kHz – 40 kHz range after wavelet processing of the "vertical", "transverse", and "longitudinal" vibrational modes. These modes were detected by accelerometers attached to the rail head. For a more in depth description on the testing procedure please refer to Lanza di Scalea and McNamara (2003).

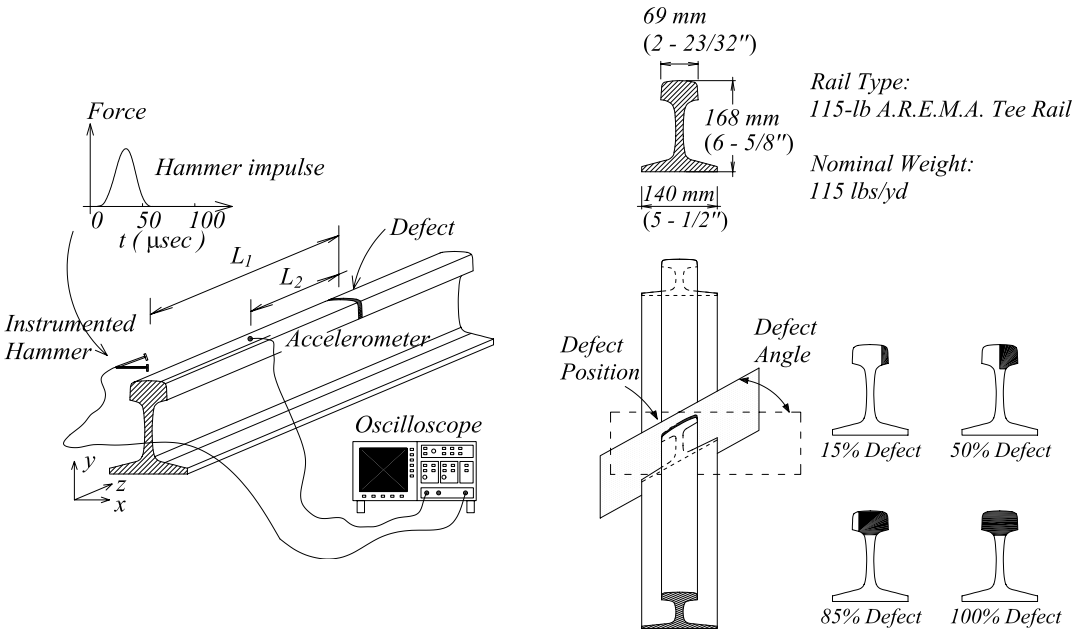


Figure 62 – Experimental setup for the automatic defect classification study.

The rail under investigation was the usual 115lb A.R.E.M.A. type schematized in Fig. 62. As shown in this figure, three types of transverse-type head defects were investigated. These were cut in the laboratory on rail sections donated by San Diego Trolley, Inc. Each type was defined by its orientation to the longitudinal axis of the rail. A perfectly-transverse defect (0°), an oblique defect oriented at 20°, and an oblique defect oriented at 35° were studied. Each of these defects was manufactured in four sizes, 1Q, 2Q, 3Q, and 4Q. The Q's stand for

one quarter of the cross sectional area of the rail head. Therefore, 1Q refers to a manufactured defect that spanned over one quarter of the total cross sectional area of the rail head. 2Q spanned over half the cross sectional area of the head, 3Q over three quarters, and 4Q completely cut through the head. Three orientations and four defect sizes resulted in twelve defects to be investigated.

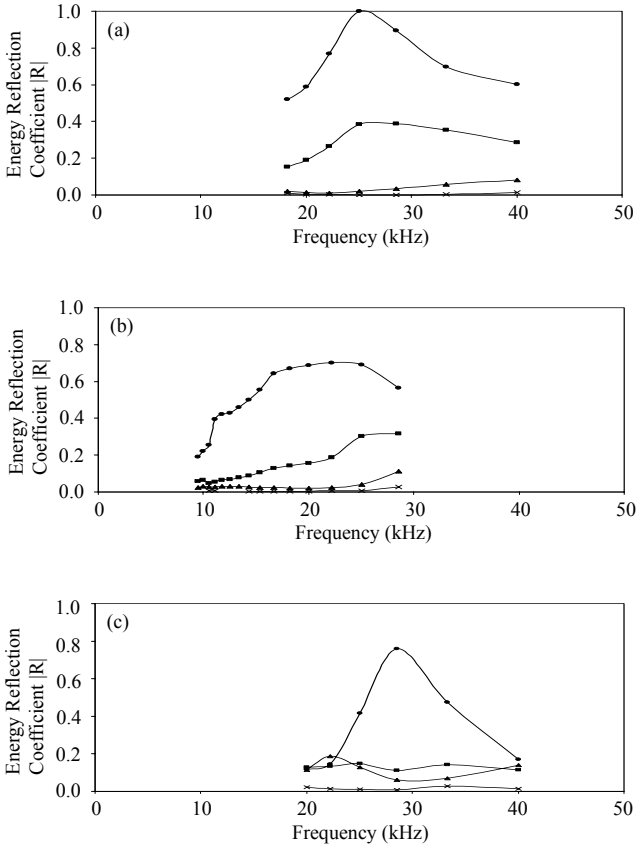


Figure 63 – Same-mode energy reflection coefficient spectra for the (a) vertical, the (b) transverse, and the (c) longitudinal modes for the 0° defect configuration (●, 4Q size defect; ■, 3Q size defect; ▲, 2Q size defect; ×, 1Q size defect).

Each of these defects were subjected to waves traveling down the rail in one of three modes of vibration, namely, vertical, transverse, and longitudinal. The energy reflected back from each of these defects was converted to reflection coefficient spectra. The spectra corresponding to results from the transverse (0°) defects subjected to the three modes of vibration are shown in Fig. 63. It is evident from these plots that the spectra do change depending on the type of defect that is present, and the type of wave that is used to inspect for defects. The question then became, what mode of vibration (i.e. test) and frequency range are best suited for defect detection. The procedure used to answer this question will be the main focus of this section. Before this procedure is discussed, soft computing paradigms and support vector machines will be introduced.

4.2 Soft Computing Paradigms

The field of pattern recognition has developed as a part of a more general field of existing soft computing paradigms which include neural networks, wavelet networks, fuzzy systems, Bayesian classifiers, fuzzy partitions, and most recently support vector machines (Ying and Licheng, 2001). In practice the differences in results depend more on the choice of activation functions, training parameters, network size, and topology than on the type of paradigm that is used. Although most of these methods have shown promise for pattern recognition applications, a detailed discussion of nondestructive evaluation applications will be limited to support vector machines.

There are two general learning paradigms: supervised and unsupervised learning (Haykin, 1999). Supervised learning can be thought of as learning with the help of a teacher. The teacher has information regarding the application environment. This information comes in the form of input-output examples. The teacher then provides the learning machine with the desired response associated with a particular example. An error signal is provided to the learning machine after it is exposed to a particular example. The error signal is defined as the difference between the desired response and the actual response of the learning machine. In unsupervised learning there is no teacher to correct the learning machine. Instead, a "task independent measure" of the performance of the learning machine is used to optimize the learning machine's free parameters. A competitive learning rule can be used to implement unsupervised learning. The "smart system" that will be discussed in this section belongs to the supervised learning paradigm.

The field of statistical learning theory (SLT), although a relatively new field has helped unify the concepts of statistical and neural pattern recognition systems, sometimes referred to as pattern classifiers. It has extended their ability to generalize from small data sets. This tends to be a great advantage for engineering, problems where data sets tend to be small by statistical standards. It could be said that one of the cornerstones of SLT is the support vector machine (SVM).

4.3 Support Vector Machines

SVMs are useful for both classification and regression (Cristianini and Scholkopf, 2002). SVMs provide additional flexibility because they are universal classifiers. Universal in the sense that they can be used to fit many different classes of discriminant functions such as linear, neural network, and radial-basis functions, without the need for major modifications of the basic learning algorithm. SVM's have been compared to other learning machines, and were found to provide a higher degree of generalization ability in classification problems (Li et al. 2000). SVMs have found success in many applications, and in fact currently hold the performance records in text categorization, handwriting recognition, and some genomics applications. But, it is their ability to generalize well given small data sets that makes SVMs an attractive alternative for many engineering applications. They have been proven effective at classifying underwater targets from acoustic backscattered signals, as well as in the

detection of damage of machine ball-bearing, and structural damage of cantilever trusses (Li et al. 2000; Zapico et al. 2001).

SVMs make use of kernel methods to simplify the task of pattern recognition. They use example (training) data to learn what the important features are for each class they are trying to identify. The fundamental idea behind kernel methods is to map data into a vector space where linear algebraic operations may be performed. One operation that can be performed is to linearly separate two types or classes of data points. The question then becomes: what space should the data be mapped to? One property that this space must have is the ability to isolate the important information from the unimportant information found in the training data. Geometrically this would mean that similar points (which could be the useful portions of data) are clumped together away from different points (which could be noise found in the data), and then separated by means of some sort of linear separator. The first obstacle encountered when trying to determine what kind of space to use is the fact that if the dimensionality of the chosen space is large enough, a trivial linear separation can be found. Although this trivial linear separation will be able to identify points from the classes found in the training examples, it is very unlikely that it would appropriately separate points from unseen data (that is data that was not used to train the kernel machine). Therefore in order to choose a space that will not provide trivial classifiers, some concepts based on statistical learning theory must be used to keep from over fitting the training data (pattern recognition). This concept is similar to using a polynomial of too great a degree to approximate experimental data (data regression). If a high enough order is chosen the polynomial curve will pass through all the data points, which in most cases will not produce a curve that can be used to predict future experimental results. Fig. 64 better illustrates these concepts.

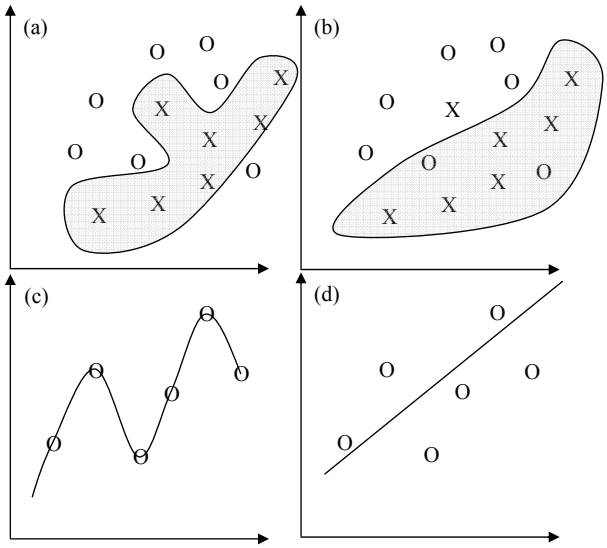


Figure 64 – Two class identification problem (a,b), and one class data regression problem (c,d). (a) classification boundary over fits training data, (b) classification boundary allows errors but better fits training data, (c) function used for data regression over fits training data, (d) regression does not over fit training data (class X and class O)

Working in high dimensional spaces can be very costly from a computational point of view. Kernel methods have a way to overcome the cost brought on by working in high dimensional spaces. Kernel methods carry out inner products in some spaces but not necessarily in the space where the training data comes from. By designing the learning process so that it only requires knowledge of the inner product between points in the embedding space, kernels can be used without explicitly performing the embedding. This is sometimes referred to as implicit mapping or implicit embedding.
The objective of pattern classification is to estimate a function such as:

$$f : \Re^n \rightarrow \{-1,+1\} \tag{18}$$

by using training data in the form of n-dimensional vectors x_i and class labels y_i so that f will correctly classify new unseen examples of the form (x, y). It would be desirable that the predicted label $f(x)$ be equal to the true label y for examples (x, y), which were generated from the same underlying probability distribution $P(x,y)$ as the training data. Statistical learning theory has shown that in order to provide good generalization (the ability to correctly classify unseen data) some restrictions on the types of functions that f can be chosen from must be made. The restriction on the types of functions that f can be chosen from has to do with the amount of available training data.

The probability of the function f mislabeling a point x whose label is y is:

$$Error = P\left(\left|x \mid f(x) \neq y\right|\right) \tag{19}$$

The goal is to upper bound this error with an expression that includes observable characteristics of the learning machine. The probability of mislabeling a new point from the same distribution, with a function that perfectly labels the training examples, is known to increase with the function complexity known as the VC (Vapnik and Chervonenkis) dimension (Vapnik and Chervonenkis, 1991). The VC dimension can be used as an observable characteristic of the learning machine, although other characteristics do exist. A measure of the error that can be expected (as a function of the number of training examples available), is the ratio of the number of training examples and the VC dimension. A ratio of 20 indicates a medium sized data set. The larger the data set the smaller the probable error (Kecman, 2001). It would seem that in order to design an effective learning algorithm, a class of functions whose capacities can be computed must be chosen. The algorithm can then be designed to keep the capacity low while fitting the training data.

Support vector machine classifiers are based on the class of hyperplanes defined as:

$$\left\langle w, x \right\rangle + b = 0 \qquad w, x \in \Re^n, b \in \Re \tag{20}$$

corresponding to the following decision functions:

$$f(x) = sign\left(\left\langle x, x \right\rangle + b\right) \tag{21}$$

where the symbol $\langle\ \rangle$ denotes the inner product of the vectors inside it. It is possible to prove that the optimal hyperplane (Fig. 65), defined as the one with the maximal margin of separation between two classes is derived from the function class with the lowest capacity. These maximal margin hyperplanes can be constructed uniquely by solving a constrained quadratic optimization problem whose solution w is defined in terms of a subset of training

examples that lie closest to the hyperplane boundary. These examples are referred to as support vectors. The support vectors associated with a particular training set contain all the relevant information for the particular classification problem.

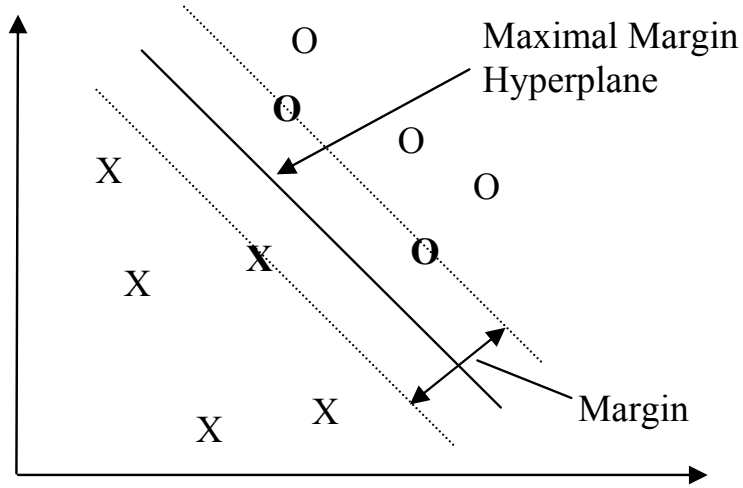


Figure 65 – A maximal margin hyperplane (support vectors are bold).

It is important to note that both the quadratic programming problem and the final decision function depend only on the inner products between examples. This is important because it allows the generalization to the nonlinear case, and it allows kernel methods to be used. The key to setting up the optimization problem (finding the maximal margin hyperplane) is to realize that for two points x^+ and x^- that lie nearest to the hyperplane the following equations holds:

$$\begin{aligned} &\langle w, x^+ \rangle + b = +1 \\ &\langle w, x^- \rangle + b = -1 \\ &\langle w, (x^+ + x^-) \rangle = 2 \\ &\left\langle \frac{w}{\|w\|}, (x^+ + x^-) \right\rangle = \frac{2}{\|w\|} \end{aligned} \tag{22}$$

where $\| \ \|$ is defined as the norm or magnitude operator. Therefore, it can be said that the margin is inversely proportional to the norm of w. So the problem becomes one of minimizing the norm of w subject to the following constraints:

$$\min \langle w, w \rangle \text{ such that } y_i [\langle w, x_i \rangle + b] \geq 1 \tag{23}$$

This equation can be solved by constructing the following Lagrangian function:

$$\frac{1}{2} \langle w, w \rangle - \sum \alpha_i [y_i (\langle w, x_i \rangle + b) - 1] \tag{24}$$

where α is the Lagrange multiplier. This can be made into a quadratic programming problem by transforming the Lagrangian function into the dual Lagrangian by imposing the optimal conditions:

$$\frac{\partial L}{\partial w} = w - \sum y_i \alpha_i x_i = 0$$
$$\frac{\partial L}{\partial b} = \sum y_i \alpha_i = 0 \tag{25}$$

The result of this transformation is a quadratic programming problem with linear constraints:

$$W(\alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle$$
$$\alpha_i \geq 0 \tag{26}$$
$$\sum \alpha_i y_i = 0$$

This formulation represents a global maximum that can always be solved efficiently. The resulting solution has the following form:

$$w = \sum y_i \alpha_i x_i \tag{27}$$

Most of the coefficients of α_i are equal to zero. The only positive coefficients correspond to the points that lie closest to the hyperplane, otherwise known as support vectors. The final decision function has the following form.

$$f(x) = \langle w, x \rangle + b = \sum y_i \alpha_i \langle x_i, x \rangle + b \tag{28}$$

where the index *i* is only carried over the training data that are support vectors. What this means is that if all the training data that are not support vectors are removed, the algorithm will come to the same conclusion. It is important to make note again that the SVM algorithm only makes use of information representing the inner products of the training data. This property makes kernel methods ideal for calculating the SVM algorithm.

Kernel methods attempt to embed data into a vector space and use linear algebra and geometry to detect hidden structure in the data. Using kernel methods it would be possible to learn classifications in the form of maximal margin hyperplanes, which was discussed previously. There are numerous reasons for embedding data into a vector space, sometimes referred to as a feature space. One important reason is that by mapping the data into an appropriate space, it is then possible to transform nonlinear data structure within the data to a linear structure.

Instead of explicitly keeping track of the individual position of each of the data points within some reference frame, the relative position of the points with reference to each other is recorded. In fact the relative quantities that are recorded are the inner products between all pairs of vectors in the embedding space. This information can be obtained in a way that is independent of the dimensionality of the embedding space. Fig. 66 illustrates the basic idea behind using kernel methods in SVM to classify data. In summary, the use of kernels makes it possible to map the data into the feature space (*F*) using a nonlinear map:

$$\phi : \Re^n \rightarrow F \tag{29}$$

then the maximal margin algorithm mentioned earlier can be applied in *F*.

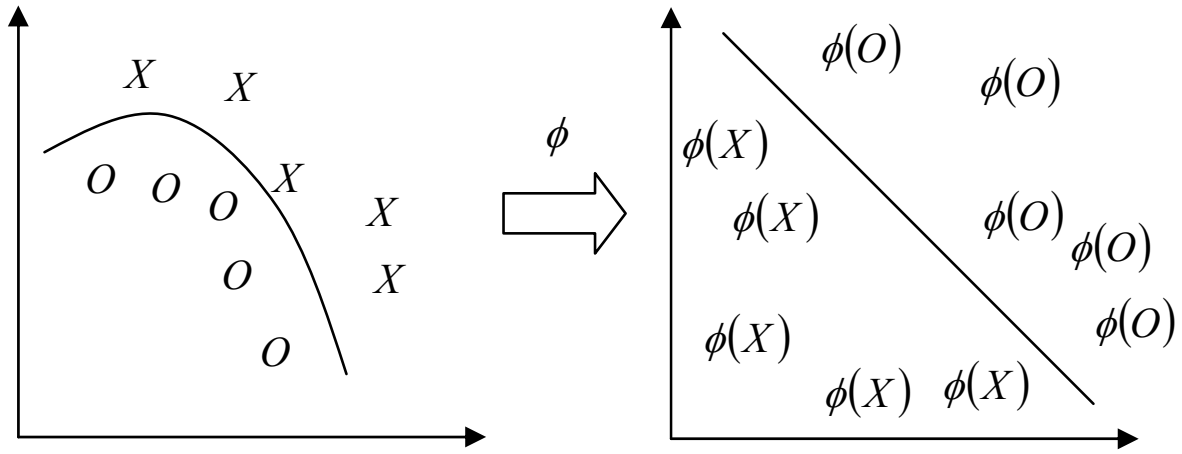


Figure 66 – Nonlinear mapping from original data space to embedded space where the two classes can be linearly separated.

4.4 Reliability Study of Automatic Classification of Rail Defects

The pilot study reported in this section was aimed at determining what combination of frequency range and mode of vibration resulted in training data that would produce a "smart system" best suited for the detection of transverse defects in rails. Each mode of vibration corresponds to a particular type of test. The defects studied corresponded to four different sizes (1Q, 2Q, 3Q, and 4Q) for three different orientations (0°, 20°, and 35°) which make a total of twelve defects. The "no defect" case is also a possible case. Due to the limited number of training examples for each case this classification problem grouped the thirteen possible defect classes into seven classes listed in Table 3.

"Leave-one-out validation" was used to train the SVM smart system. In leave-one-out validation each pattern is left out once, and the SVM performance is then measured. This provides an indication as to whether the vertical and transverse test reflection coefficients are capable of discriminating between defect sizes and orientation. A variety of kernels such as a Gaussian, Polynomial, and Linear kernels were investigated. After preliminary results, the linear kernel was chosen for this pilot study because it is the simplest kernel to use, it provides comparable results to the other two more complicated kernels investigated, and it generated the fewest number of support vectors which should increase its generalization performance.

The frequency range of interest ranged from 10 kHz to 40 kHz. The SVM algorithm was trained with four different frequency ranges. The first frequency range to be investigated was the complete range from 10 kHz to 40 kHz. In order to determine whether a particular frequency range was best suited for a particular class, three additional frequency ranges were investigated: low frequencies from 10 kHz to 20 kHz, medium frequencies from 20 kHz to 30 kHz, and high frequencies from 30 kHz to 40 kHz. The SVM smart system was trained with each of these four frequency ranges in order to determine a correlation between defect identification and the frequency range used to train the SVM.

Table 3 - Classification problem classes.

Class 1 (Large 0^o Defects)	Sizes 4Q and 3Q for the 0^o Orientation (20 Training Data Examples)
Class 2 (Small 0^o Defects)	Sizes 2Q and 1Q for the 0^o Orientation (20 Training Data Examples)
Class 3 (Large 20^o Defects)	Sizes 4Q and 3Q for the 20^o Orientation (10 Training Data Examples)
Class 4 (Small 20^o Defects)	Sizes 2Q and 1Q for the 20^o Orientation (10 Training Data Examples)
Class 5 (Large 35^o Defects)	Sizes 4Q and 3Q for the 35^o Orientation (10 Training Data Examples)
Class 6 (Small 35^o Defects)	Sizes 2Q and 1Q for the 35^o Orientation (10 Training Data Examples)
Class 7 (No Defects)	No Defect Examples (80 Training Data Examples)

Three test data combinations were used in the training data matrix. The first combination included reflection coefficients for both the vertical and the transverse long range tests (Lanza di Scalea and McNamara 2003). The longitudinal case was omitted because it can not be practically implemented in the field. The second combination only used reflection coefficients from the vertical test, and the third only used reflection coefficients from the transverse test. These three training data combinations were used to determine whether one particular test configuration or a combination would be best to detect certain defect orientations or sizes.

The results for this classification problem will be presented for all twelve possible combinations. There are two quantities of interest for each of these combinations. The first is the percent correctly classified, which means for example that a class 1 example is presented to the SVM and it categorizes it as a class 1. The second quantity of interest is the percent of false negatives. This is particularly important for the "no defect" class because a false negative in this case would mean that a defect had gone undetected. In order to do leave one out validation, one training example is left out, and the SVM is trained with the remaining examples. Then the example that was left out is presented to the trained SVM and the category it is assigned to is recorded. Then the next example is left out, and the process is carried out again. The results of this study are presented in Figs. 67 through 70 and are summarized in Table 4.

The terms "correctly classified" and "incorrectly classified", found in Figs. 67 through 70 have very particular definitions. An important error to avoid would be saying there was a defect when there wasn't (false positive), but even much more important would be saying that there wasn't a defect when there was (false negative). The term "correctly classified" means that a particular defect was classified correctly, or that a non-defect example was correctly classified as being defect free. If a defect class is said to be incorrectly classified 10% of the time, that means that 10% of the time that defect class was classified as not having a defect

(false negative). If the non-defect class is said to be incorrectly classified 10% of the time, that means that 10% of the time non-defect examples were thought to show evidence of a defect (false positive). If a particular defect class reports that it was never incorrectly classified, but it was only correctly classified 90% of the time, then it can be assumed that 10% of the time it was classified as the wrong defect class, which although being an error is not as important as the false positive or false negative errors. The most important errors are the incorrectly classified defect classes because they correspond to false negative, which would result in defects being missed.

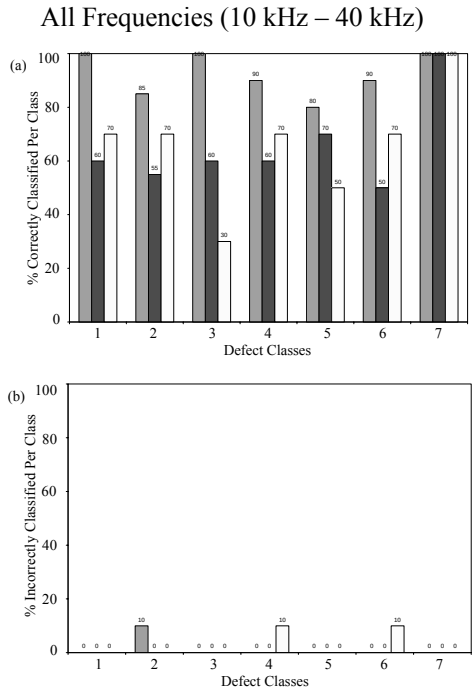


Figure 67 – (a) Percent correctly classified; (b) percent incorrectly classified (□, vertical and transverse data; ■, only vertical data; □, only transverse data) (10 kHz to 40 kHz).

High Frequencies (30 kHz – 40 kHz)

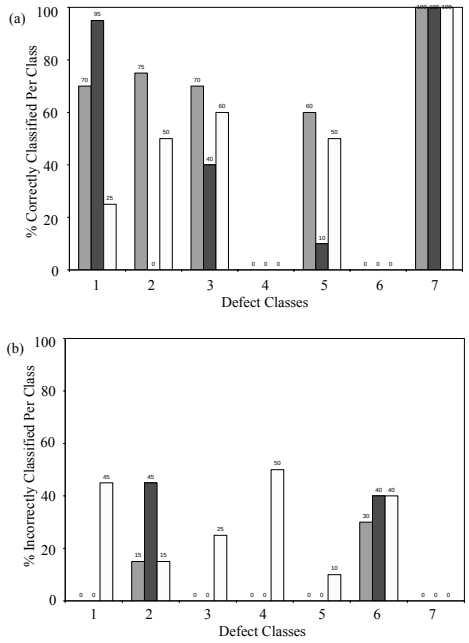


Figure 68 – (a) Percent correctly classified; (b) percent incorrectly classified (▣, vertical and transverse data; ■, only vertical data; □, only transverse data) (30 kHz to 40 kHz).

Medium Frequencies (20 kHz – 30 kHz)

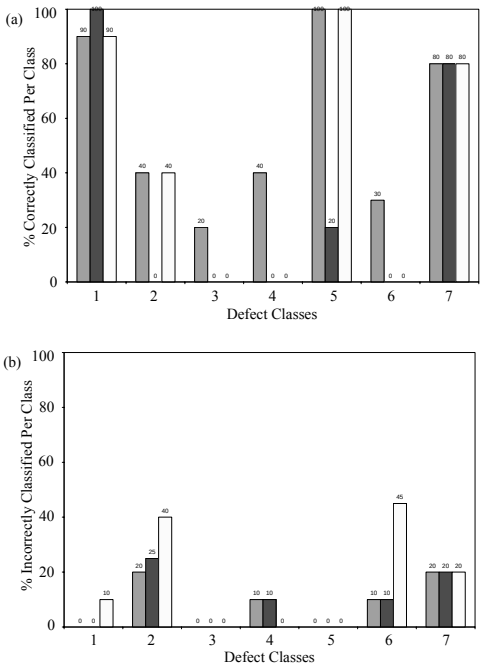


Figure 69 – (a) Percent correctly classified; (b) percent incorrectly classified (▣, vertical and transverse data; ■, only vertical data; □, only transverse data) (20 kHz to 30 kHz).

Low Frequencies (10 kHz – 20 kHz)

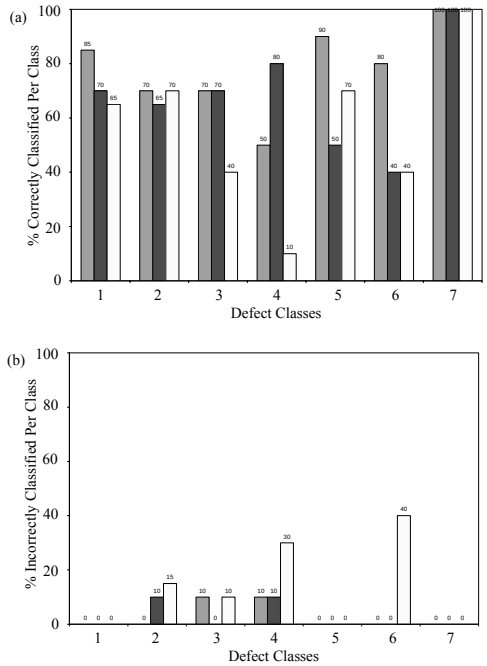


Figure 70 – (a) Percent correctly classified; (b) percent incorrectly classified (□, vertical and transverse data; ■, only vertical data; □, only transverse data) (10 kHz to 20 kHz).

Table 4 compares the results for this classification problem when the entire frequency range (10 kHz – 40 kHz) was used to train the SVM vs. the other three possible frequency ranges used to train the SVM. If using a particular frequency range produced better results for a particular class -- **better** is listed in the appropriate box. A particular frequency range could do better than the complete frequency range in one of two ways, either the percent correctly classified went up for that particular class, or the percent incorrectly classified went down for that particular class. If a particular frequency range did worse in either of these two ways than the complete frequency range, a *worse* is listed in the appropriate box. Mixed means that either the percent correctly classified may have gone up but so did the percent incorrectly classified, or they could both have gone down. If there was no change in results, then the box is marked as being the same. In summary, Table 4 provides an indication at to whether a particular frequency range is better suited to detect a particular class of defect than the entire frequency range. Although there are some improvements when other frequency ranges are used, the choice to use one of the smaller frequency ranges would depend on which defects are being looked for. If there is an equal interest in all the defect types then the entire frequency range should be used.

Table 4 Summary of defect classification from Figs. 67 through 70

	Data Used	Class1	Class2	Class3	Class4	Class5	Class6	Class7
	V + T	*Worse*	*Worse*	*Worse*	*Worse*	*Worse*	*Worse*	*Same*
High Frequencies	V Only	***Better***	*Worse*	*Worse*	*Worse*	*Worse*	*Worse*	*Same*
(30-40 kHz)	T Only	*Worse*	*Worse*	*Mixed*	*Worse*	*Worse*	*Worse*	*Same*
	V + T	*Worse*	*Worse*	*Worse*	*Worse*	***Better***	*Worse*	*Worse*
Medium Frequencies	V Only	***Better***	*Worse*	*Worse*	*Worse*	*Worse*	*Worse*	*Worse*
(20-30 kHz)	T Only	*Mixed*	*Worse*	*Worse*	*Mixed*	***Better***	*Worse*	*Worse*
	V + T	*Worse*	*Mixed*	*Worse*	*Worse*	***Better***	*Worse*	*Worse*
Low Frequencies	V Only	***Better***	*Mixed*	***Better***	*Mixed*	*Worse*	*Worse*	*Same*
(10-20 kHz)	T Only	*Worse*	*Worse*	*Worse*	*Worse*	***Better***	*Worse*	*Same*

Better = Better Than When Complete Frequency Is Used.
Same = Same As When Complete Frequency Is Used.
Worse = Worse Than When Complete Frequency Is Used.
Complete Frequency Range = 10kHz – 40kHz.
V = Vertical Test Data, T = Transverse Test Data

4.5 Summary and Conclusions of the Automatic Defect Classification Study

The results presented in this section show evidence that there is enough information within the experimental reflection coefficient data from long-range ultrasonic waves, to at least identify each of the seven classes of transverse-type head defects in the rail examined. In general using the data from both the vertical and the transverse tests as input provides the best results (Fig. 67a). The small percentages of false negatives are limited to the smaller defect sizes for all defect orientations. When only a portion of the entire frequency range is used the percent of false negatives and positives increase. Although the percent correctly classified increases for some classes (Table 4) when training with data from a certain portion of the frequency range, the increase in false negatives and positives across the board make using a portion of the frequency range a less attractive alternative.

When the complete frequency range is used the percent false negatives is very low especially if both the vertical and transverse test data are used. Also, when the complete frequency range was used there were no false positives reported. This helps increase the effective inspection speed because current standards require operators to stop and verify possible

defects. By reducing the number of false positives, the number of unnecessary stops is also reduced.

Using a portion of the frequency range shows improvements in both false positive and false negatives for the class of large defects (class 1) when the vertical test is conducted, but has a detrimental effect on the results for the remaining classes. Since the percent of false negatives (the most important error to eliminate) increases when using smaller portions of the frequency range, the drawbacks seem to outweigh the benefits.

Using both the vertical and transverse data along with the complete frequency range produces overall the best results. The greater the number of reflection coefficient spectra features used, the better the results. This could be interpreted to mean that the entire range of reflection coefficient spectra, contain information crucial to detecting and classifying rail transverse-type defects.

Although the automatic defect classification algorithm was demonstrated for large cracks, the same procedures can be applied for the detection of small cracks by high-frequency waves, subject that was discussed in Section 3 of this report.

5. RECOMMENDATIONS FOR FUTURE STUDIES

The long-range defect detection method was demonstrated for selected transverse-type cracks in the rail head that were simulated in the laboratory by saw cutting. The system needs to be validated for the detection of a larger variety of defects, including real flaws from the field. Defects in the rail web and base should be also considered, and ways to discriminate among defects located in various positions within the rail should be investigated. A literature survey of train accidents caused by track failures should be completed to determine the most critical flaws to be considered in the next study.

The automatic defect classification algorithm should be tested on this larger selection of defects. The finite element simulations could be performed again in order to extract reflection and transmission coefficients as testing data for the algorithm. Unsupervised learning methods, such as those based on outlier detection, should be considered in addition to the supervised methods that were already examined. Unsupervised learning does not need testing data and can thus potentially avoid the need for investigating many defect cases. The drawback of unsupervised learning when compared to supervised learning is a more qualitative, rather than quantitative defect detection.

The use of optical fibers for delivering the laser light should be considered for field application of the technique. This would increase the flexibility of the system. One important issue that will have to be addressed is the maximum laser power deliverable without damaging the fiber.

A prototype of the inspection system should be constructed for field validation. For this purpose hardware for ultrasound generation/detection and software for signal synchronization and processing should be developed and assembled. The use of GPS tagging could be also considered to provide information on the location of a defect that has been detected.

A final issue is the determination of the severity of a defect that has been detected, in terms of remaining life of the rail. In order to properly address this issue fracture mechanics studies should be conducted to determine the rate of growth of a given crack under service loads and its effects on the remaining useful life of the rail.

REFERENCES

Alleyne, D. and Cawley, P., "A Two-dimensional Fourier Transform Method for the Measurement of Propagating Multimode Signals," *Journal of the Acoustical Society of America*, Vol. 89, No. 3, 1991, pp. 1159-1168.

Bartoli, I., Lanza di Scalea, F., Fateh, M. and Viola, E., "Modeling Guided Wave Propagation with Application to the Long-Range Defect Detection in Railroad Tracks," *NDT&E International*, 2004, in press.

Cawley, P., Lowe, M., Alleyne, D., Pavlakovic, B. and Wilcox, P. "Practical Long Range Guided Wave Testing: Applications to Pipes and Rail," *Materials Evaluation*, Vol. 61, No. 1, 2003, pp. 66-74.

Cristianini, N. and Scholkopf, B., "Support Vector Machines and Kernel Methods: the New Generation of Learning Machines," *AI Magazine*, 2002, pp. 31- 41.

Gavrić, L., "Finite Element Computation of Dispersion Properties of Thin-walled Waveguides," *Journal of Sound and Vibration*, Vol. 173, No. 1, 1994, pp. 113-124.

Gavrić, L., "Computation of Propagative Waves in Free Rail Using a Finite Element Technique," *Journal of Sound and Vibration*, Vol. 185, No. 3, 1995, pp. 531-543.

Haykin, S., *Neural Networks – A Comprehensive Foundation*, Prentice Hall, New Jersey, 1999.

Kecman, V., *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*, MIT Press, Cambridge, MA, 2001.

Kenderian, S., Djordjevic, B.B. and Green, R.E. jr., "Laser Based and Air Coupled Ultrasound as Noncontact and Remote Techniques for Testing of Railroad Tracks," *Materials Evaluation*, Vol. 60, No. 1, 2002, pp. 65-70.

Kenderian, S., Cerniglia, D., Djordjevic, B.B., Garcia, G., Sun, J. and Snell, M., "Rail Track Field Testing Using Laser/Air Hybrid Ultrasonic Technique," *Materials Evaluation*, Vol. 61, No. 10, 2003, pp. 1129-1133.

Lanza di Scalea, F., "Advances in Non-contact Ultrasonic Inspection of Railroad Tracks," *Experimental Techniques*, Vol. 24, No. 5, 2000, pp. 23-26.

Lanza di Scalea, F., *On-line High-speed Rail Defect Detection*, Phase I Draft Report to FRA and UCSD Report SSRP-2003/07, 2003.

Lanza di Scalea, F. and McNamara, J., "Ultrasonic NDE of Railroad Tracks: Air-coupled Cross-sectional Inspection and Long-range Inspection," *Insight-NDT and Condition Monitoring*, Vol. 45, No. 6, 2003, pp. 394-401.

Lanza di Scalea, F. and McNamara, J., "Measuring High-frequency Wave Propagation in Railroad Tracks by Joint Time–frequency Analysis," *Journal of Sound and Vibration*, Vol. 273, No. 3, 2004, pp. 637-651.

Li, D., Azimi-Sadjadi, M. and Dobeck, G., "Comparing Different Neural Network Classification Paradigms for Underwater Target Discrimination," *Proceedings SPIE*, Vol. 4038, 2000, pp. 334 – 345.

Mallat, S.G., *A Wavelet Tour of Signal Processing*, 2nd ed., Academic Press, New York, 1999.

McNamara, J., *Health Monitoring of Railroad Tracks by Elastic-wave Based Non-destructive Testing*, Ph.D. Dissertation, University of California, San Diego, 2003.

McNamara, J. and Lanza di Scalea, F., "Improvements in Non-contact Ultrasonic Testing of Rails by the Discrete Wavelet Transform," *Materials Evaluation*, Vol. 62, No. 3, 2004, pp. 365-372.

McNamara, J., Lanza di Scalea, F. and Fateh, M., "Automatic Defect Classification in Long-range Ultrasonic Rail Inspection Using a Support Vector Machine-based Smart System," *Insight-NDT and Condition Monitoring*, Vol. 46, No. 6, 2004, pp. 331-337.

Moser, F., Jacobs, L.J. and Qu, J., "Modeling Elastic Wave Propagation in Waveguides with the Finite Element Method," *NDT&E International*, Vol. 32, No. 4, 1999; pp. 225-234.

Thompson, D.J., "Wheel-Rail Noise Generation, Part III: Rail Vibration," *Journal of Sound and Vibration*, Vol. 161, No. 3, 1993, pp. 421–446.

Thompson, D.J., "Experimental Analysis of Wave Propagation in Railway Tracks," *Journal of Sound and Vibration*, Vol. 203, No. 5, 1997, pp. 867-888.

Torvik, P.J., "Reflection of Wave Trains in Semi Infinite Plates," *Journal of the Acoustical Society of America*, Vol. 41, No. 2, 1967, pp. 346–353.

Trivedi, A., Krysl, P., McNamara, J. and Lanza di Scalea, F., "Adaptive Simulation of Ultrasonic Guided Waves in Railroad Tracks Using CHARMS," *Proceedings of the 6th World Congress of Computational Mechanics*, Beijing, China, 2004, in press.

Vapnik, V. and Chervonenkis, A., "Necessary and Sufficient Conditions for Consistency in Empirical Risk Minimization Method," *Pattern Recognition and Image Analysis*, Vol. 1, No. 3, 1991, pp. 283–305.

Wilcox, P., Evans, M., Diligent, O., Lowe, M. and Cawley, P., "Dispersion and Excitability of Guided Acoustic Waves in Isotropic Beams with Arbitrary Cross Section," *Review of Progress in QNDE,* Vol 21, ed. Thompson D.O. and Chimenti D.E., 2002, pp. 203-210.

Wilcox, P., Evans, M., Pavlakovic, B., Alleyne, D., Vine, K., Cawley, P. and Lowe, M., "Guided Wave Testing of Rail," *Insight-NDT and Condition Monitoring*, Vol. 45, No. 6, 2003, pp. 413-420.

Wu, T.X. and Thompson D.J., "A Double Timoshenko Beam Model for Vertical Vibration Analysis of Railway Track at High Frequencies," *Journal of Sound and Vibration*, Vol. 224, No. 2, 1999, pp. 329-348.

Ying, L. and Licheng, J., "Unification of Support Vector Machines and Soft Computing Paradigms for Pattern Recognition," *Proceedings SPIE on Neural Networks and Distributed Processing*, Vol. 4555, 2001, pp. 154-159.

Zapico, J., Worden, K. and Molina, F., "Vibration-based Damage Assessment in Steel Frames Using Neural Networks," *Smart Materials and Structures*, Vol. 10, 2001, pp. 553-559.