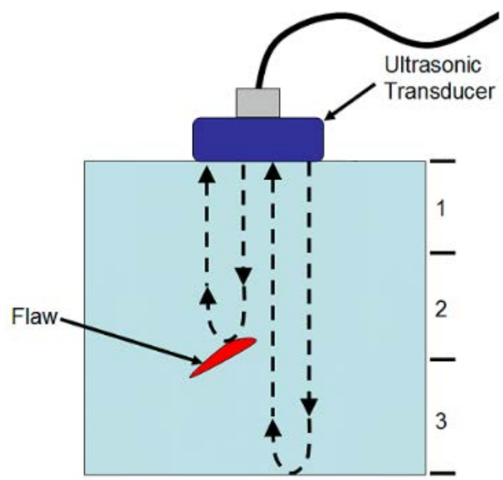
U.S. Department of Transportation

Federal Railroad Administration

Quantification of the Effectiveness of Handheld Equipment for Ground Verification of Detected Rail Internal Defects

Office of Research and Development
Washington, DC 20590



DOT/FRA/ORD-14/07

Final Report
April 2014

REPORT DOCUMENTATION PAGE			*Form Approved* *OMB No. 0704-0188*

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE April 2014	3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE

Rail Integrity Application of Ultrasonic Phased Arrays for Rail Flaw Sizing

5. FUNDING NUMBERS

DTFR-00-C-00012

Task Order 250

6. AUTHOR(S)

Leith Al-Nazer (Federal Railroad Administration), Dr. Thomas Raslear (Federal Railroad Administration), Lucas R. Welander (Transportation Technology Center, Inc.)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Data Collection: Transportation Technology Center, Inc. (TTCI)
55500 DOT Road
Pueblo, CO 81001

Data Analysis: U.S. Department of Transportation
Federal Railroad Administration
Office of Research and Development
1200 New Jersey Ave., SE
Washington, DC 20590

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Department of Transportation
Federal Railroad Administration
Office of Research and Development
1200 New Jersey Ave., SE
Washington, DC 20590

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

DOT/FRA/ORD-14/07

11. SUPPLEMENTARY NOTES

COR: Leith Al-Nazer

12a. DISTRIBUTION/AVAILABILITY STATEMENT

This document is available to the public through the FRA Web site at http://www.fra.dot.gov.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The objective of this project was to quantify the effectiveness of the rail inspection ground verification process. More specifically, the project focused on comparing the effectiveness of conventional versus phased array probes to manually detect and size internal rail defects. Signal detection theory was used to quantify the effectiveness parameter. The results tentatively indicate that phased array technology has approximately the same sensitivity as conventional ultrasonic equipment, but there are preliminary indications that phased array technology may be more effective at sizing internal defects. However, follow-on studies with additional data collection are necessary in order to draw a more definitive conclusion.

14. SUBJECT TERMS

Ultrasonic phased arrays, conventional ultrasonic, rail flaw sizing, nondestructive testing

15. NUMBER OF PAGES

40

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

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

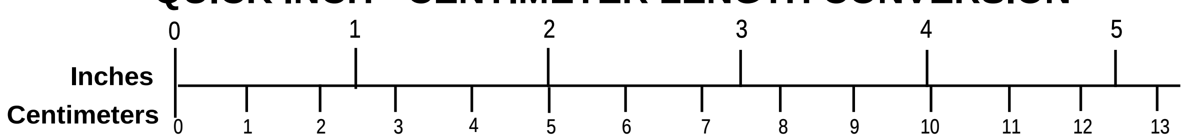
VOLUME (APPROXIMATE)

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

°F	-40°	-22°	-4°	14°	32°	50°	68°	86°	104°	122°	140°	158°	176°	194°	212°
°C	-40°	-30°	-20°	-10°	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	100°

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

Contents

Figures

Tables

Executive Summary

In the railroad industry, inspection of rail for internal defects is performed on a regular basis. This inspection is performed using hi-rail vehicles (known as rail detector cars) that are outfitted with advanced sensors and data processing units. When there is an indication of an exception (the potential presence of a defect), manual ground verification of the exception location is performed using handheld ultrasonic units. The ground verification process is conducted in order to determine more definitively whether or not a defect is actually present. Quantifying the effectiveness of both the hi-rail inspection systems and the manual ground verification process is important, as they are both integral parts of the rail inspection process. Thus, knowing the relative effectiveness parameters for each of these systems will allow for optimizing inspection frequencies by railroads, thereby minimizing risk.

Traditionally, the handheld units used in the ground verification process consisted of a probe with a single transducer fixed at a specific angle. Recently, however, probes with multiple transducers, which are able to "sweep" through a range of angles, have become more prevalent and are being adopted by other industries. Such probes are often referred to as phased array probes.

This project focused on quantifying the effectiveness of the ground verification process. More specifically, the project focused on comparing the effectiveness of using conventional versus phased array probes to manually detect and size defects. Signal detection theory was used to quantify the effectiveness parameter. The results tentatively indicate that using phased array probes does not improve the ground verification process. However, follow-on studies with additional data collection are necessary in order to draw a more definitive conclusion.

In order to perform this comparison study, it was first necessary to gather and organize a "library" of defects. This sample collection will also be useful for future studies directed towards quantifying the effectiveness of new rail inspection technologies. In order to form the library, segments of rail which had developed rail internal defects were acquired, and their defects were characterized as accurately as possible. In addition to these segments of rail with natural defects (also known as field defects), a set of rails with artificial defects was manufactured. These rails with artificial defects are referred to as master gauges. Unlike with the field defects, the exact size and orientation of the master gauge defects are known.

Nine test equipment operators participated in the study. Two of the participants used phased array equipment, and the remaining seven used conventional ultrasonic equipment. Participants came from various industries, including ultrasonic equipment manufacturers, rail inspection service providers, and railroad personnel. Initial findings indicate that handheld phased array ultrasonic units used for ground verification have the same sensitivity as conventional ultrasonic equipment and can size defects more accurately. However, a couple of points warrant clarification. First, there were only two participants using phased array equipment, so the statistical tests performed (Section 5) are not ideal; a larger sample size would allow for more definitive conclusions. Second, the two phased array participants did not have substantial experience performing ultrasonic testing on rails because most of their experience involved other types of structures (e.g., pipelines). By contrast, the participants using conventional ultrasonic testing equipment had extensive experience testing for rail internal defects.

1. Introduction

In the railroad industry, inspection of rail for internal defects is performed on a regular basis using hi-rail vehicles (known as rail detector cars) that are outfitted with advanced sensors and data processing units. When there is an indication of an exception (the potential presence of a defect), manual ground verification of the exception location is performed using handheld ultrasonic units. The ground verification process is conducted in order to determine more definitively whether or not a defect is actually present.

The first objective of this project is to quantify the effectiveness of the rail inspection ground verification process. More specifically, this project focused on comparing the effectiveness of using conventional versus phased array probes to manually detect and size defects. Signal detection theory was used to quantify the effectiveness parameter.

Knowing the effectiveness of the ground verification process is important for calculating the risk of a broken rail derailment. In addition, knowing this effectiveness parameter will help in setting optimal inspection frequencies, thereby increasing the safety of the rail industry. Finally, quantifying the effectiveness of current technology provides a baseline against which future technologies can be judged.

A second objective of this project is to form a collection or “library” of rail samples with internal defects. This library of defects will also be valuable for future research and development projects. In addition, the library may help with the process of evaluating the performance of future rail inspection systems or the performance of the human operators of the rail inspection equipment.

2. Background

The purpose of internal rail inspection is to detect flaws which develop inside the rail. Such flaws are usually not detectable through a visual inspection of the rail surface. Internal flaws develop for various reasons, and they often initiate at high-stress areas, such as the rolling contact interface between the wheel and rail. As the rail experiences repeated cyclical loading, these small initiation cracks grow in size and can eventually result in a complete rail break (also known as a service failure) if not detected and removed from the track.

Undetected internal flaws pose a serious risk to the railroad. An undetected defect may result in rail failure, causing disruption of service and the potential risk of catastrophic consequences in the event of a derailment. According to the Federal Railroad Administration (FRA) Office of Safety Analysis Web site, in 2010 and 2011, there were 103 and 104 derailments, respectively, due to undetected transverse defects, including detail fractures, compound fissures, and transverse fissures. In 2010, the costs of infrastructure and equipment damage due to these derailments were approximately $27 million, and in 2011 the costs were approximately $23 million.

1. Types of Rail Internal Defects

There are many types of rail internal defects. This project focuses on the detection of transverse defects since they are the most common type of rail internal defect. A transverse defect is a type of fatigue that develops in a plane transverse to the cross-sectional area of the railhead. Development can be normal or in multiple stages prior to failure, as discussed below. The transverse defect is only identified by the nondestructive inspection process, unless the defect has progressed to the rail running surface.

The first edition of the *FRA Track Inspector Rail Defect Reference Manual* identifies detail fractures, compound fissures, transverse fissures, and engine burn fractures as typical transverse defects found in the head of rail in North American rail service operations. The following brief descriptions are taken directly from the *FRA Track Inspector Rail Defect Reference Manual*:

- **Detail Fracture** (Figure 1): A progressive fracture originating at or near the surface of the rail head. These fractures should not be confused with transverse fissures, compound fissures, or other defects, which have internal origins. Detail fractures may originate from shelly spots, head checks, or flaking.

Figure 1. Detail Fracture

- **Compound Fissure** (Figure 2): A progressive fracture originating in a horizontal split head, which turns up or down in the head of the rail as a smooth, bright, or dark surface, progressing until it is substantially at a right angle to the length of the rail.

Figure 2. Compound Fissure

- **Transverse Fissure** (Figure 3): A progressive crosswise fracture starting from a crystalline center or nucleus inside the head from which it spreads outward as a smooth, bright or dark, round or oval surface substantially at a right angle to the length of the rail. The distinguishing features of a transverse fissure compared with other types of fractures or defects are the crystalline center or nucleus and the nearly smooth surface of the development that surrounds it.

Figure 3. Transverse Fissure

- **Engine Burn Fracture** (Figure 4): A progressive fracture originating in spots where driving wheels have slipped on top of the rail head. In developing downward, they frequently resemble the compound or even transverse fissures with which they should not be confused or classified.

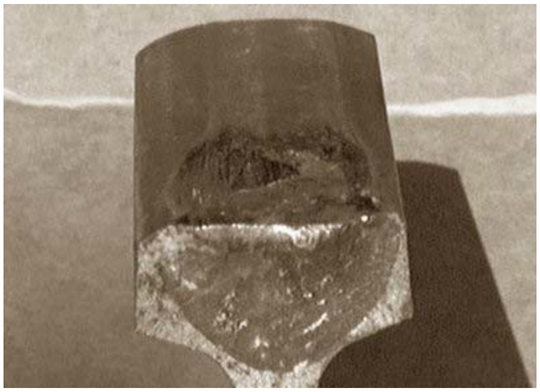


Figure 4. Engine Burn Fracture

2. FRA Regulations for Rail Inspections

Title 49, Part 213 of the Code of Federal Regulations (CFR), commonly known as the Track Safety Standards (TSS), contains requirements that FRA believes necessary to maintain safe track and a stable, viable rail network. Requirements for rail internal inspections are described in Section 213.237 of the TSS for track classes 1–5. Paragraph (a) of 49 CFR 213.237 is restated here:

(a) In addition to the track inspections required by 213.233, a continuous search for internal defects shall be made of all rail in Classes 4 through 5 track, and Class 3 track over which passenger trains operate, at least once every 40 million gross tons (mgt) or once a year, whichever interval is shorter. On Class 3 track over which passenger trains do not operate such a search shall be made at least once every 30 mgt or once a year, whichever interval is longer.

Section 213.339 of the TSS describes internal rail inspection requirements for track classes 6–9. Paragraph (a) of 49 CFR 213.339 is restated here:

(a) A continuous search for internal defects shall be made of all rail in track at least twice annually with not less than 120 days between inspections.

Some FRA regulations are the end result of recommendations made by the National Transportation Safety Board (NTSB). Appendix A provides an example of recommendations that the NTSB made after a major rail-caused derailment that occurred in 1992 near Superior, WI.

3. Current Industry Rail Inspection Practices

In order to conform to the internal rail inspection requirements outlined in Sections 237 and 339 of the TSS, railroads perform periodic, automated inspections for rail internal defects. To maximize rail life, railroads use strategic processes that minimize service failure occurrences, relying on periodic ultrasonic or induction rail testing and renewal of rail that shows obvious evidence of fatigue.

Traditionally, ultrasonic techniques are predominantly used to detect internal rail defects, although devices operating on induction principles serve as supplemental, add-on systems to detect defects that may be missed by ultrasonic-based systems. Ultrasonic techniques consist of a mechanical means for striking the rail, which introduces sound waves that travel through the rail. Historically, the mechanical means constitutes a transducer containing a piezoelectric element, but, recently, systems relying on laser technologies have been developed. Receivers are used to capture reflected sound waves, and the received sound waves are analyzed to determine whether or not a defect is present. With sufficient transducers, processing power, and operator expertise, the size and orientation of the defect may also be determined.

Initial inspections are performed with a hi-rail vehicle outfitted with advanced sensors and data processing units. When there is an indication of an exception (the potential presence of a defect), manual ground verification of the exception location is performed using handheld ultrasonic testing units. The ground verification process is conducted in order to determine more definitively whether or not a defect is actually present.

4. Phased Array Technology

Conventional handheld units used in the ground verification process consist of a probe with a single transducer fixed at a specific angle. Recently, however, probes with multiple transducers, which are able to “sweep” through a range of angles, have become more prevalent and are finding applications in other industries. In theory, using phased array technology to electronically steer ultrasonic waves increases the probability of detecting and accurately sizing defects in the rail. The wave is steered by applying a delay to a specific group of elements in order to shape the sound wave to create the desired angle and focal spot.

Phased array technology allows the sound wave characteristics to be modified for the specific inspection. These characteristics include focusing and electronic steering to provide the inspector with multiple angles of inspection in a single scan. In theory, the probability of detecting flaws at different angles of orientation increases by applying focal laws that will steer through a sequence of angles. Additional information regarding the utilization of phased array technology for ultrasonic testing of rail can be found in DOT/FRA/ORD-06/17: *Application of Ultrasonic Phased Array for Rail Flaw Sizing* (Garcia 2006).

3. Library of Defects

1. Master Gauge Development

A matrix of artificial defects, known as master gauges, was developed to provide various attributes that represent transverse defects with different flaw orientation and size. Rail samples were obtained and artificial flaws were manufactured to represent defects with differing cross sectional head area, angle of inclination, location relative to the railhead center point, and surface condition. Several of the master gauge samples were drilled with flat bottom holes into the end of the rail. The holes were drilled in various sizes, depths, and angles. The holes drilled into the railhead were at locations that would be representative of transverse defects. These holes were then plugged and the rail ends machined to mask the hole locations.

All master gauges have been documented with head geometry, surface condition, and location of defect(s). The rail samples were profiled using a laser profiler and converted to an AutoCad format, as seen in Figure 5, to document the exact geometry of the rail.

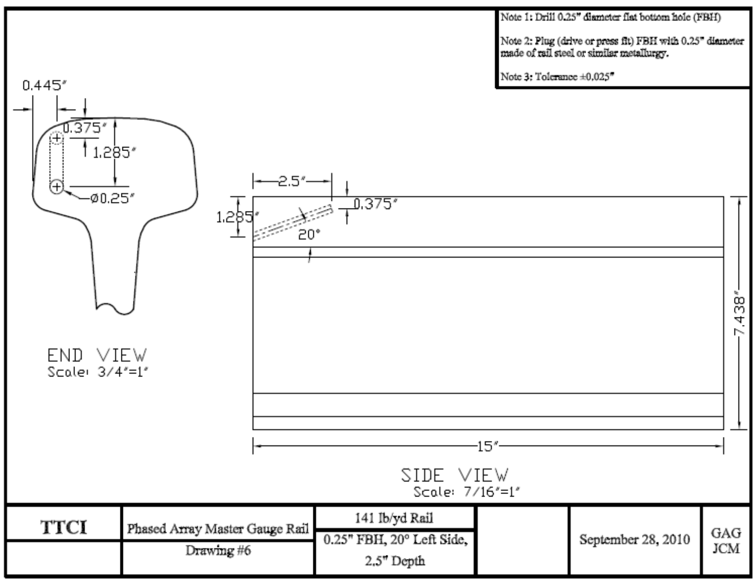


Figure 5. Master Gauge Blueprint

Table 1 contains the matrix of master gauges identifying various attributes of defects manufactured into the railhead. The numbers 1 and 2 represent the gauges developed in support of Phases I and II of this research. Number 3 identifies the suggested gauges to be manufactured during follow-on efforts in subsequent years.

Table 1. Matrix of Master Gauges

Angle Of Inclination Relative to Vertical (degree)	Position of Defect Center Relative to Center of Railhead	Wear or Surface Condition	Defect Size 5%	Defect Size 10%	Defect Size 20%	Defect Size 30%
0	Right	None	1	1	1	1
5	Right	None	3	3	3	3
10	Right	None	2	2	2	2
20	Right	None	1	1	1	1
30	Right	None	2	2	2	2
35	Right	None	3	3	3	3
0	Center	None	1	1	1	1
5	Center	None	3	3	3	3
10	Center	None	2	2	2	2
20	Center	None	1	1	1	1
30	Center	None	2	2	2	2
35	Center	None	3	3	3	3
0	Left	Surface Damage	1	1	1	1
5	Left	Surface Damage	3	3	3	3
10	Left	Surface Damage	2	2	2	2
20	Left	Surface Damage	1	1	1	1
30	Left	Surface Damage	2	2	2	2
35	Left	Surface Damage	3	3	3	3
0	Left	Severe Wear	1	1	1	1
5	Left	Severe Wear	3	3	3	3
10	Left	Severe Wear	2	2	2	2
20	Left	Severe Wear	1	1	1	1
30	Left	Severe Wear	2	2	2	2
35	Left	Severe Wear	3	3	3	3

2. Previously In-Service Rail Samples

Samples containing service-induced flaws were added to the library of defects in order to complement the master gauges. These samples are similar in size (to the extent practical) to the master gauges. Sizing of service-induced flaws was performed using a 6-decibel drop method with conventional ultrasonic testing technology and a sectorial scan (known as an S-scan) with phased array technology. Any rail sample that contained a service-induced defect open to the surface was subsequently welded and machined to mask the flaw locations.

Along with the service-induced rail flaws, worn rail has been used as a basis for some of the master gauges (Figure 6). The rail used represents varying surface conditions and railhead geometry. The compilation of varying rail flaws and rail conditions was used to represent characteristics and variables that are seen during field-testing.

Figure 6. Rail with Worn Geometry Characteristics and Surface Flaws

4. Baseline Characterization and Data Collection Procedures

During this research effort, detection and sizing were performed by comparing results from phased array ultrasonic testing with those from conventional ultrasonic testing. All inspections were performed using the pulse-echo technique (Figure 7).

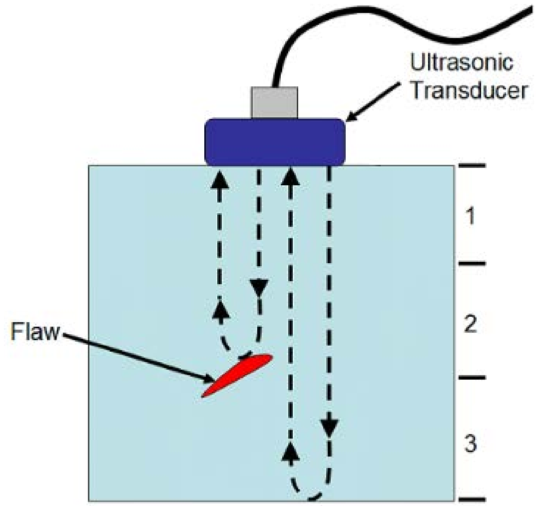


Figure 7. Pulse-Echo Technique

The pulse-echo technique measures the time required for an ultrasonic pulse to travel through a thickness of material, reflect from a surface with an impedance mismatch, and return to the transducer. The sound path distance (material thickness) can be calculated using the formula:

$$thickness = \frac{(velocity)(time)}{2}$$

The size of a flaw, when reported as a percentage of cross-sectional head area, is directly influenced by the cross-sectional area of the railhead. For example, the same size flaw has a higher percentage cross sectional head area in service-worn rail than in a new section of rail.

Since the library of defects includes service-worn rail, the head area varies between specimens. Therefore, the head area of each rail specimen was provided for the participants, who were then left to determine the area of any flaw present and the percent of cross sectional head area that the flaw occupied. Table 2 shows an example of the master key with the master gauge identification numbers removed in order to maintain anonymity for future evaluations.

Table 2. Master Key Example

Rail Number	Master Gage	Rail Length (in)	Flaw Size (dia. - in)	Flaw Size (orientation - °)	Flaw Length (in)	Flaw Location	Distance to Gauge Side (in)	Distance to Field Side (in)	Distance to H.P (Insert -in)	Distance to H.P (end - in)
203-3	132	36	0.25	5	4	G	0.4177	2.4033	1.1255	0.777
204-1	126	36	0.25	15	5.7	C	1.4894	1.5169	1.8644	0.3305
206-4	93	36	0.25	10	4	F		0.3312	1.1879	0.4788
210-1	130	36	1	10	3.1	C	1.535	1.4925	1.2377	0.6934
211-1	90	36	0.25	5	4.25	F		0.3967	0.7105	0.3377
212-3A	124	36	0.25	10	3.75	G	0.3708		1.1673	0.5022
222-1	129	36	0.5	5	5.9	G	0.6204	2.3391	1.109	0.5949
228-2	122	36	0.25	15	2.5	G	0.3973		1.2414	0.5743
232-1	127	36	0.25	15	4.8	C	1.2714	1.3169	1.6155	0.3341
233-1	94	36	1	15	1.9	C	1.2918	0.4767	1.1987	0.6951
235-2	95	36	1	5	1.4	G	0.7695	2.0266	0.9146	0.7916
236-3	131	36	1	5	2.2	F	2.0551	0.8156	0.8929	0.7011
237-1	92	36	1	10	0.83	G	0.7511	2.061	0.9432	0.7965
240-1	91	36	0.5	5	4.5	C	1.3732	1.4419	0.8619	0.4663
241-2	123	36	0.25	10	5	C	1.4018	1.4097	1.2089	0.3235
249-1	125	36	0.25	5	2.8	G	0.3775		0.7988	0.5573

1. Baseline Characterization

To establish baseline sizes of the various internal defects included in the library of defects, multiple inspection techniques were performed on each sample. This step is especially relevant to the service-induced flaws. Evaluations were performed using phased array technology with two probes, 2.25 MHz and 5 MHz, with 16 active elements. Figure 8 represents half of the active elements used during the inspections.

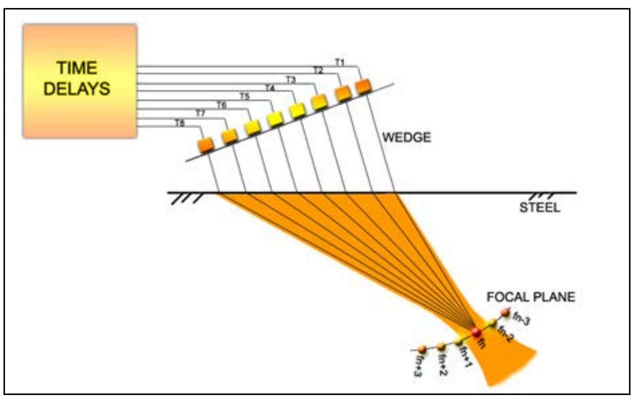


Figure 8. Diagram of Eight Element Phased Array Probe

Phased array probes produce sectorial scans, known as S-scans (Figure 9). These scans cover a wide range of angles and theoretically provide improved flaw detection capabilities.

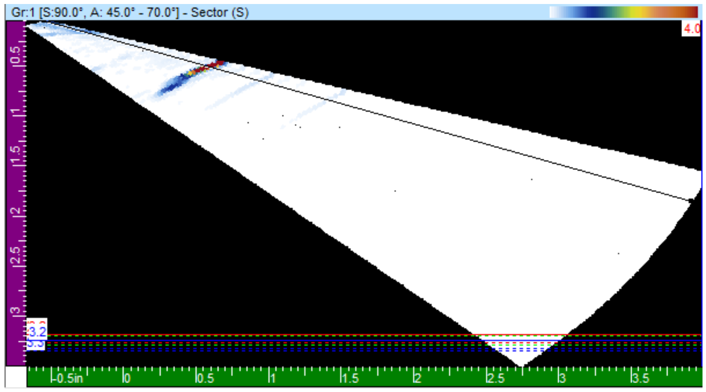


Figure 9. Sectorial Representation of a Half-Inch Flaw

In addition to phased array probes, conventional ultrasonic transducers were also used in the baseline characterization process. Conventional probes produce an amplitude scan, known as an A-Scan (Figure 10). The conventional transducers used were 2.25 MHz transducers and were half an inch in diameter. Samples were characterized and compared by using Lucite wedges to introduce refracted angles of 45 and 70 degrees. In addition, the samples were scanned using a 0-degree incident angle.

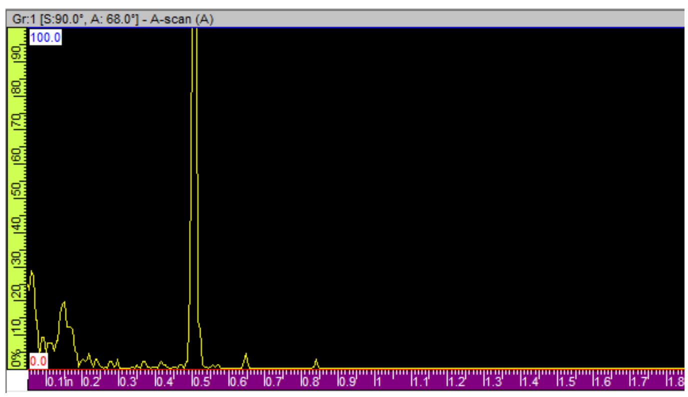


Figure 10. A-Scan Representation of a Half-Inch Flaw

2. Participant Evaluations

Following the development and baseline characterization of the library of defects, a data recording and inspection protocol was established. Participants throughout the rail inspection industry and non-destructive testing industry were invited to participate in flaw detection and sizing evaluations. Participants were asked to provide their own inspection equipment (either conventional or phased array), procedures, and necessary accessories to perform the evaluations on all specimens in the library of defects.

Participant were required to set up and calibrate their equipment to meet their companies' procedures. Once this step was complete, each participant inspected specimens from the library of defects to determine if a flaw was present and to identify the location and the size of the flaw (Figure 11). Sixty specimens were provided for each participant to inspect. Results were recorded as either a hit (true positive), a false alarm (false positive), a miss (false negative), or a correct rejection (true negative) and were tabulated for each inspection technique. Data was analyzed as individual inspection results and then combined into a composite analysis using signal detection theory.

Figure 11. Inspection Participant Using Conventional Ultrasonic Testing Equipment

The following operational protocol was used for each inspector or inspection sequence. Appendix B contains an example of an operator profile sheet.

- Evaluations scheduled Monday–Friday
  - All evaluations were performed during normal business hours
- A pretest meeting was conducted prior to inspection. The meeting:
  - Included a Safety Manager, Engineering Services Director, Project Engineer, FRA representative, and the industry participant(s);
  - Addressed schedule and objectives for the evaluations;
  - Provided time to conduct an operator profile on the inspector; and
  - Offered an open forum to voice concerns or questions prior to testing.

The following was the minimum each operator was asked to provide prior to inspection:

- Company name
- Date
- Inspection method (conventional or phased array)
- Operator name
- Level of certification
- Equipment used
- Transducer type, size, and frequency
- Type and manufacturer of couplant

- Calibration artifact (e.g., International Institute of Welding, distance sensitivity calibration block, or rail)
- Calibration procedure
- Scanning results
  - Map and/or description of scan locations
  - Flaw location (referenced from rail end)
  - Flaw orientation (horizontal, transverse, or vertical)
  - Flaw width (from gauge to field side of railhead)
  - Flaw height (from top to bottom of railhead)
  - Flaw depth (from top of railhead to top of flaw)

5. Evaluation Methodology and Results

Both conventional ultrasonic testing (UT) and phased array UT techniques were used for detection and sizing of the transverse defects. During this research effort, there were a total of seven participants that performed inspections using conventional ultrasonic equipment and two participants that used phased array ultrasonic equipment. The nine participants were labelled A through I. Participants A and B used phased array UT equipment, and participants C through I used conventional UT equipment. The test specimens contained 60 samples with 29 flawed and 31 unflawed specimens. 24 out of the 29 flawed defects were artificial defects where the exact size and angle were known. The remaining five defects were defects obtained from the field.

1. Brief Background on Signal Detection Theory

Signal detection theory was used to quantify the performance of each of the participants in the study.[1] In signal detection theory, the outcome of a "yes-no" experiment falls into one of four categories:

- Hit (or true positive)
- False alarm (or false positive)
- Miss (or false negative)
- Correct rejection (or true negative)

These four categories are summarized in Table 3.

Table 3. Signal Detection Theory Categories

		Response of Observer	
		"Yes"	"No"
State of World	Defect	Hit (True Positive)	Miss (False Negative)
	No Defect	False Alarm (False Positive)	Correct Rejection (True Negative)

The following formulae relate the probability of a hit to the probability of a miss and the probability of a false alarm to the probability of a correct rejection:

$$p(\text{Hit}) + p(\text{Miss}) = 1$$
$$p(\text{False Alarm}) + p(\text{Correct Rejection}) = 1$$

The sensitivity (commonly denoted by the symbol *d'* and pronounced "dee-prime") of each of the participants was calculated. In signal detection theory, the sensitivity is traditionally defined using the following formula:

[1] There are several references that provide a good introduction to signal detection theory, including Macmillan (2005) and Wickens (2002).

$$d' = z(\text{Hit Rate}) + z(\text{False Alarm Rate})$$

Equation 1

where *z*(Hit Rate) is the z-value associated with the specified hit rate, and *z*(False Alarm Rate) is the z-value associated with the specified false alarm rate. Given a specified rate or probability, the z-value is the corresponding x-axis value of the standard normal distribution. Another parameter, commonly known as the percent correct or accuracy, was also calculated for each participant. The accuracy parameter is a less rigorous but more intuitive measure of sensitivity than *d'*. The percent correct or accuracy is calculated using the following formula:

$$\text{Accuracy} = \frac{(\text{Hits}) + (\text{Correct Rejections})}{(\text{Hits}) + (\text{Correct Rejections}) + (\text{Misses}) + (\text{False Alarms})}$$

2. Flaw Detection Sensitivities and Bias

The data in Table 4 includes the sensitivity (*d'*) as well as the accuracy[2] for each of the participants. Participant F discovered (after concluding the test) that the settings on the UT equipment were not proper. For this reason, in most of the following analyses, Participant F is excluded in the calculation of various metrics. Whenever Participant F is included in the calculation of a metric, mention is made of the inclusion.

[2] See Section 7.1 for an explanation of the terms sensitivity and accuracy.

Table 4. Sensitivity Values for Individual Participants

Participant	Sensitivity (*d'*)	Accuracy
A	1.74	0.80
B	2.81	0.90
C	2.36[3]	0.80
D	3.54[4]	0.97
E	2.12	0.85
F	0.83	0.55
G	2.34	0.87
H	1.06	0.7
I	2.44	0.85

A Student's t-test[5] was performed in order to determine whether the sensitivities of the participants using phased array UT technology (Participants A and B) were reliably different from the sensitivities of the participants using conventional UT technology (Participants C, D, E, G, H, and I). Basically, the values in the solid blue cells in Table 4 were compared with the values in the solid orange cells. A p-value of 0.48 was obtained for a two-tailed test, which indicates that the two groups are not reliably different.[6]

Figure 12 shows the sensitivity values for each of the participants on a receiver operating characteristic (ROC) plot.[7] The blue data points represent the two participants using phased array UT equipment, and the red data points represent the participants using conventional UT equipment. Four curves with positive slope are also shown on the plot. They represent curves of constant sensitivity, commonly known as isosensitivity curves. The four isosensitivity curves have *d'* values of 0, 1, 2, and 3, as indicated in the plot's legend. Figure 12 provides a good visual justification for the exclusion of Participant F in any statistical analyses conducted.

---

[3] Participant C had no false alarms. Having a value of 0 or 1 for the hit rate or false alarm rate prevents the use of Equation 1. Therefore, a value of 0.5 was added to the number of hits, correct rejections, misses, and false alarms before calculating this *d'* sensitivity value.

[4] Participant D had no false alarms. Having a value of 0 or 1 for the hit rate or false alarm rate prevents the use of Equation 1. Therefore, a value of 0.5 was added to the number of hits, correct rejections, misses, and false alarms before calculating this *d'* sensitivity value.

[5] Information on statistical tests such as t-tests can be found in Spiegel (2013), as well as in most introductory books on statistics.

[6] P-values are described in introductory books on statistics (Spiegel 2013). P-values below 0.05 typically indicate that the two samples being compared are reliably different, and, therefore, most likely came from different populations.

[7] More information on ROC plots can be found in references on signal detection theory, including Macmillan (2005) and Wickens (2002).

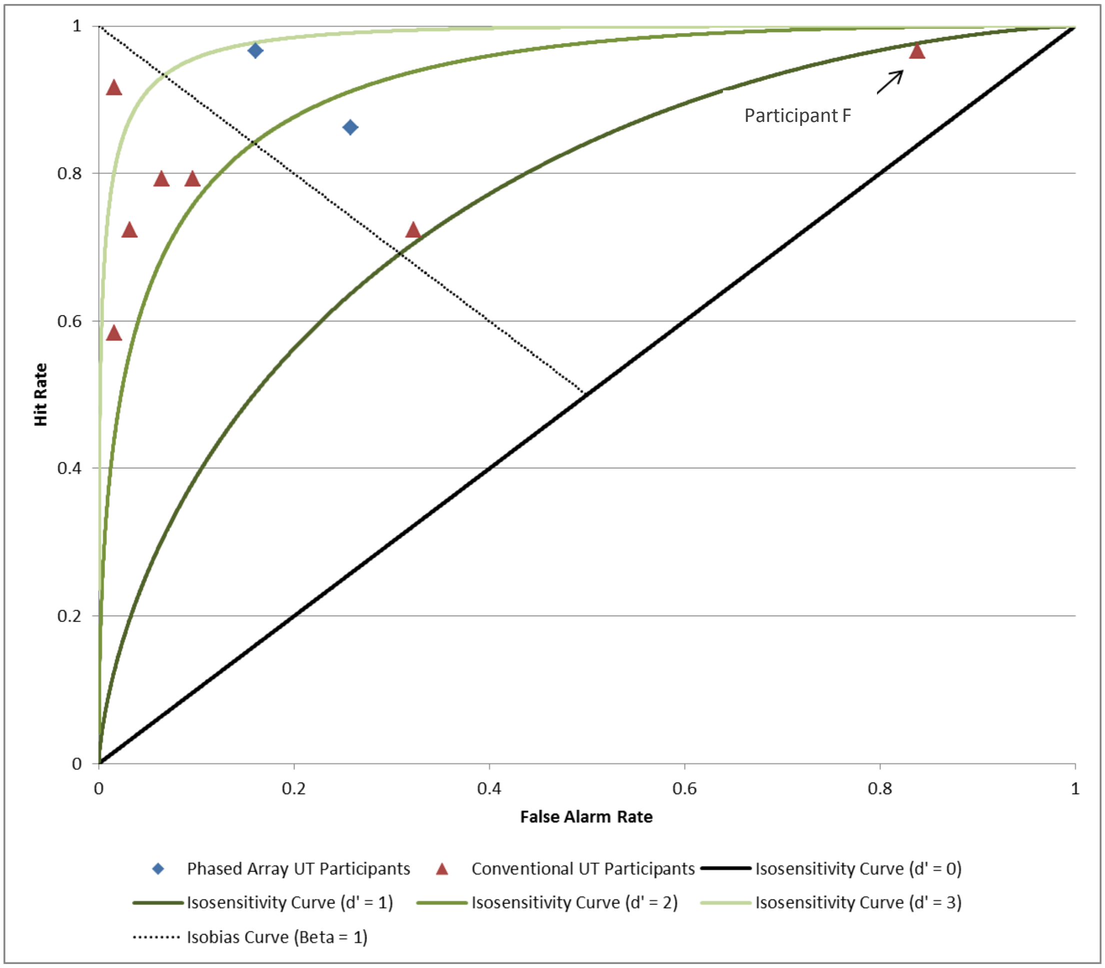


Figure 12. Receiver Operating Characteristic

Figure 13 shows the same data as Figure 12, but uses z coordinates rather than hit rate and false alarm rate for the y and x axis, respectively. As can be seen in Figure 13, isosensitivity curves are linear when plotted in z coordinates.

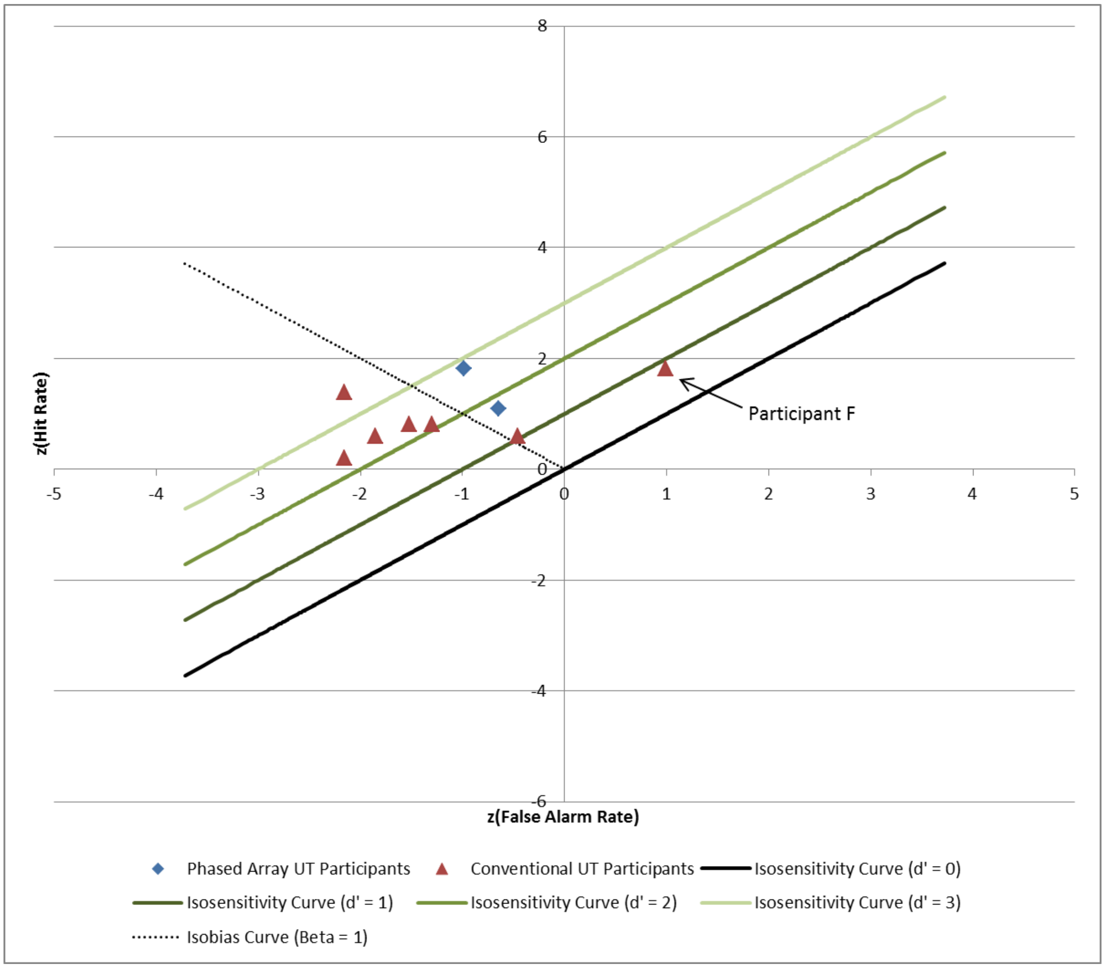


Figure 13. Receiver Operating Characteristic in Z-Coordinates

Also shown in the plots (Figure 12 and Figure 13) is a dotted line with negative slope. This line represents a line of constant bias, also known as an isobias curve. In signal detection theory, bias is commonly defined as:

$$\beta = e^{cd'}$$

where β is the metric for bias and is also referred to as the likelihood ratio (Macmillan 2005), *d'* is the sensitivity, and *c* is the criterion and is defined as follows:

$$c = -\frac{1}{2}\left[z(H) + z(F)\right]$$

The specific isobias line shown in the plot is for a criterion *c* value of zero and a resulting likelihood ratio β value of 1. When *c* equals zero, the false alarm (false positive) and miss (false negatives) rates are equal. It is clear that five out of seven of the conventional UT participants fall below the neutral isobias line while both of the phased array UT participants fall above that line. This indicates that the majority of the conventional UT participants had higher miss rates than false alarm rates, while the two phased array UT participants had higher false alarm rates than miss rates. In other words, the conventional UT participants, in general, were more inclined to say that there was not a defect present, while the phased array UT participants were more inclined to say that there was a defect present. However, as stated previously, there is not a statistically reliable difference in the *d'* sensitivity values of the two groups.

Table 5. Bias Values for Individual Participants

Participant	Bias (β)
A	0.68
B	0.31
C	9.95
D	3.91
E	1.67
F	0.31
G	2.27
H	0.93
I	4.63

A Student's t-test was performed to determine whether the biases of the participants using phased array UT equipment (Participants A and B) were different from the biases of the participants using conventional UT equipment (Participants C, D, E, G, H, and I). A p-value of 0.026 was obtained for a two-tailed test, which indicates that the two groups were in fact reliably different.

It is important to note that the participants using phased array UT technology did not have extensive experience performing UT scans on sections of rail; most of their experience in performing UT scans consisted of work in other areas, such as inspection of pipelines. The participants using conventional equipment, however, were generally well-versed in performing UT scans on sections of rail. It is not definitive, but it is possible that this experience gap may have contributed to the differences in bias between the two groups. However, the difference in bias could conceivably also be due to the differences in phased array technology and conventional technology. Phased array technology, due to its sweeping action, tends to saturate the rail with UT waves. Therefore, there is more potential for "noise" or false positives to appear due to the multiple looks (or sweeps through a range of angles) that the phased array technology is capable of performing.

3. Effect of Flaw Angle on Sensitivity

In Section 6, the potential benefits of phased array UT technology over conventional UT technology were briefly described. Phased array probes provide the ability to sweep through a range of angles, whereas conventional probes are limited to a single angle by the hardware being used, namely the angle of the probe wedge. The sweeping capabilities of phased array technology create the perception that the angle of a rail defect should not be a significant factor for detection of defects. However, with conventional techniques it is believed that there could be defects oriented at particular angles that may be harder to detect due to the single angle limitation of the conventional UT probes.

To that end, this section presents the same analysis that was presented in Section 7.2, but this time two sensitivities were calculated for each participant. The first sensitivity was calculated using the 31 samples with no defects, along with the 12 artificial defect samples oriented at a 20-degree angle with respect to the vertical; the second sensitivity was calculated using the same 31 samples with no defects, along with the 12 artificial defect samples oriented at a 0-degree angle with respect to the vertical. The five samples with field defects were not considered because their respective defect angles were not known exactly or were not constant.

What is expected or hypothesized is that the sensitivity for the participants using phased array UT technology would be approximately the same for both the group of 0-degree angle defects and the group of 20-degree angle defects; in other words, their sensitivity may be independent of the angle of the defect. For the participants using conventional equipment, it is possible that a greater variation in sensitivity between the two samples could be observed; in other words, their sensitivity may be dependent on the angle of the defect.

Table 6 shows the *d'* sensitivity parameter for each of the participants. Specifically, the second column shows sensitivities for the group with the defects oriented at an angle of 20 degrees with respect to the vertical, and the third column shows the sensitivities for the group with the defects oriented at an angle of 0 degrees with respect to the vertical.

Table 6. Sensitivity Values for Individual Participants (20° and 0° Angle Defects)

Participant	Sensitivity (*d'*) for Sample with Defects Oriented at 20° Angle	Sensitivity (*d'*) for Sample with Defects Oriented at 0° Angle
A	2.39	1.32
B	2.72	2.37
C	3.02	1.96
D	3.92	3.02
E	3.00	1.97
F	0.82	0.39
G	2.90	2.19
H	1.84	0.46
I	2.52	2.52

Two t-tests were performed. The first t-test compared the *d'* values for the sample with defects at an angle of 20 degrees with the *d'* values for the sample with defects at an angle of 0 degrees achieved by the participants using conventional UT equipment. In other words, the values in the six partially shaded blue cells in Table 6 were compared with the values in the six partially shaded orange cells. This one-tailed t-test resulted in a p-value of 0.05, which provides justification for accepting the original hypothesis that the sensitivity of the participants using conventional UT technology depends on the angle of the defect—a p-value of 0.05 indicates that the sensitivities were reliably higher for defects at an angle of 20 degrees compared with defects at an angle of 0 degrees with respect to the vertical rail cross section.

The participants using phased array UT equipment also had higher sensitivities for defects at an angle of 20 degrees than for defects at an angle of 0 degrees (Table 6). A second t-test was performed that compared the *d'* values for the sample with defects at an angle of 20 degrees with the *d'* values for the sample with defects at an angle of 0 degrees achieved by the participants using phased array UT equipment. In other words, the values in the two solid blue cells in Table 6 were compared with the values in the two solid orange cells. A p-value of 0.20 was calculated for this one-tailed t-test. Since the p-value is above 0.05, the two samples are not reliably different. However, it is important to note that there were only two numbers in each sample. This t-test result simply shows that the difference between the means for the two samples is not statistically significant, as it was for the participants using conventional UT equipment. Therefore, the possibility exists that the use of phased array technology could help improve detection of defects at certain angles, such as 0 degrees with respect to the vertical, where conventional UT equipment is less sensitive. However, with more phased array participants, the alternative possibility also exists that an increased number of phased array participants may lead to a t-test with a p-value of 0.05 or below, as was obtained for the participants using conventional equipment.

The plot in Figure 14 provides a nice visual; it shows the average d' values for the group using phased array UT equipment and the group using conventional UT equipment as a function of the defect angle. As can be seen, the lines are nearly parallel, indicating that the two group's respective sensitivities may be affected in the same way by changes in defect angle. However, as pointed out previously, it is important to note that there were only two phased array participants in the study.

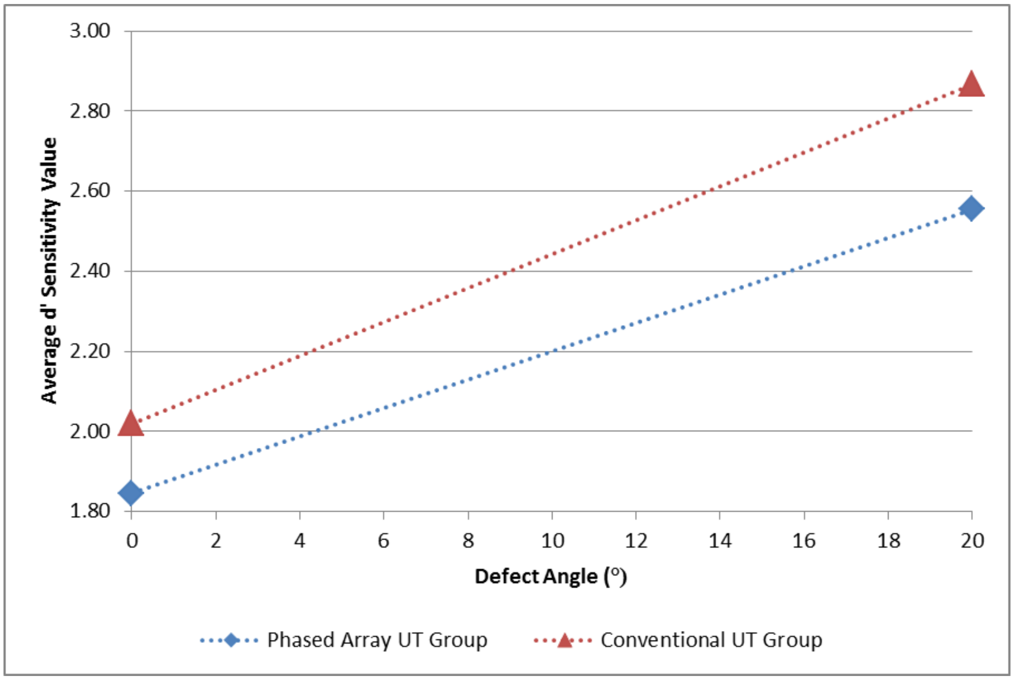


Figure 14. Average d' Sensitivity Value as Function of Defect Angle

4. Flaw Sizing Errors

Figure 15 and Figure 16 show histograms of the height and width errors, respectively, for the participants using phased array UT equipment versus the participants using conventional UT equipment. The data presented in these histograms includes only the width and height errors for hits on the master gauges with artificial defects; sizing error associated with misses and false alarms were not included. In addition, Participant F was not included in the data for the participants using conventional ultrasonics; this amounted to 44 data points for the participants using phased array UT equipment and 113 data points for the participants using conventional UT equipment. A positive error represents an incident where a participant sized a defect larger than its actual size, and a negative error represents an incident where a participant sized a defect smaller than its actual size.

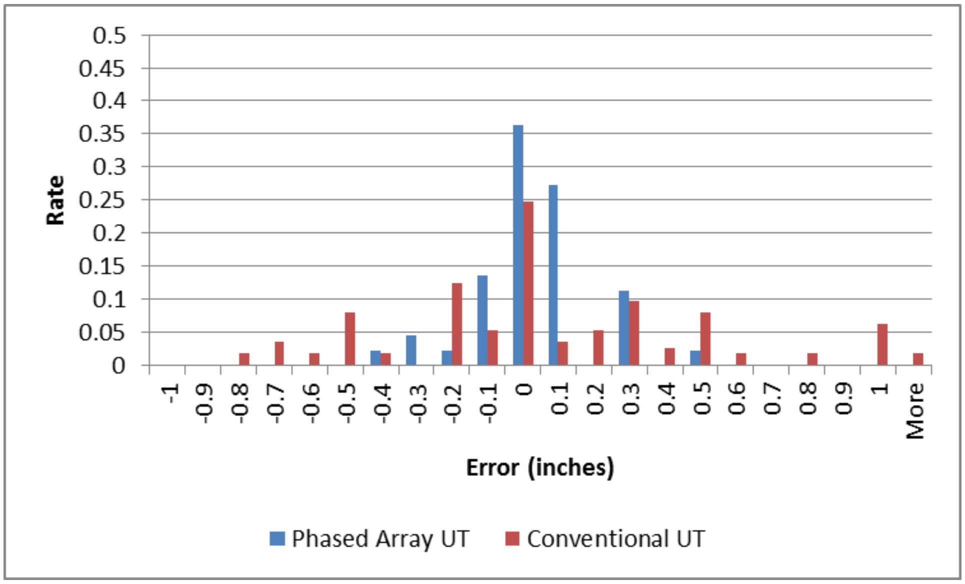


Figure 15. Histogram of Flaw Height Measurement Errors

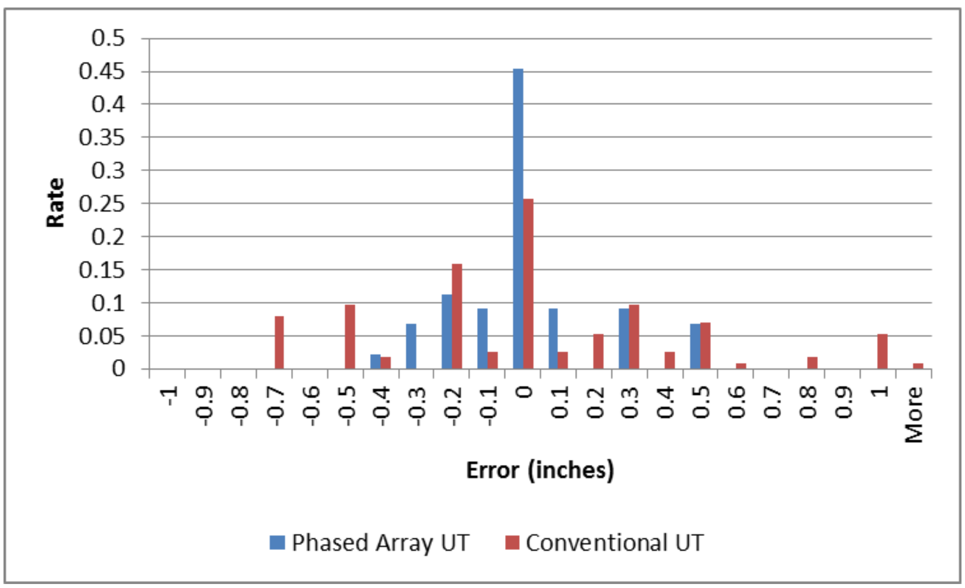


Figure 16. Histogram of Flaw Width Measurement Errors

Table 7 summarizes the means and standard deviations of the data presented in the histograms in Figure 15 and Figure 16. The participants using conventional UT equipment had larger errors. These large errors are made evident by the higher standard deviations for participants using conventional UT equipment versus participants using phased array UT equipment (Table 7). For both groups, the mean was close to zero, which indicates that there are no systematic errors in the equipment or process used by both groups. However, the higher variance for the group using conventional UT equipment is indicative of less precise hardware or a less precise data collection process.

Table 7. Means and Standard Deviations of Height Error and Width Error

	Phased Array UT Sample	Conventional UT Sample
Mean of Height Error (inches)	0.02	0.06
Standard Deviation of Height Error (inches)	0.16	0.49
Mean of Width Error (inches)	0.01	0.00
Standard Deviation of Width Error (inches)	0.21	0.44

In order to analyze the variance further, two F-tests were performed. The first F-test compared the variance of the height errors for the participants using phased array UT equipment to the height errors for the participants using conventional UT equipment, and a second F-test compared the variance of the width errors for the same two groups. Both F-tests had p-values well below 0.01. Therefore, there is a statistically reliable difference for both the height error variance and the width error variance between the participants using phased array UT equipment and the participants using conventional UT equipment.

5. Scan Times

The histogram in Figure 17 shows the distribution of scan times for participants using phased array ultrasonics and participants using conventional ultrasonics. As with the histograms in Figure 15 and Figure 16, only scan times from hits are included.

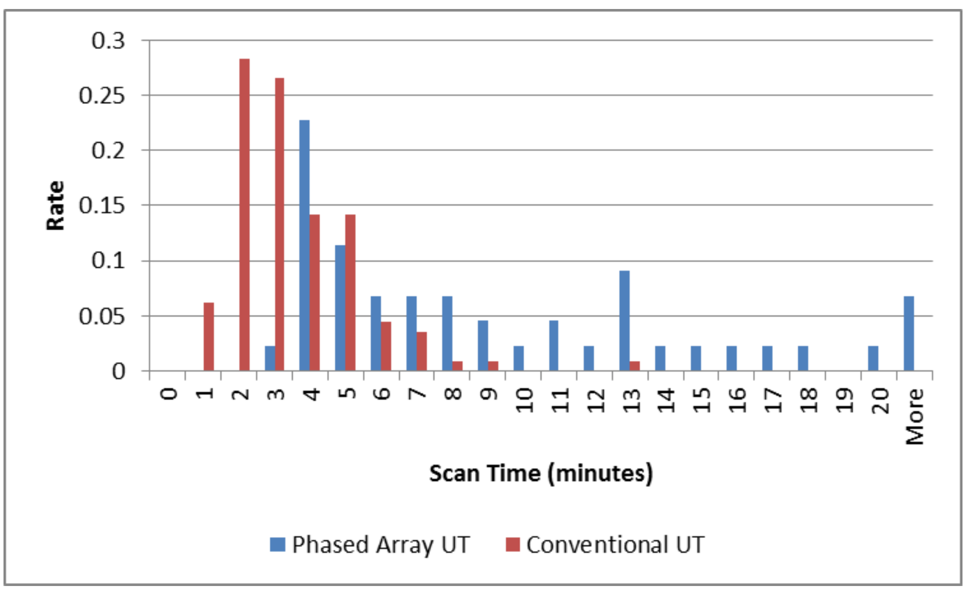


Figure 17. Histogram of Scan Times

As can be seen in Figure 17, the scan time distributions for both the phased array UT participants and the conventional UT participants are skewed. Therefore, the median or geometric mean, rather than the arithmetic mean, are more appropriate to use as measures of central tendency.[8] The median values and geometric mean values are presented in Table 8.

The average scan time of the participants using phased array technology was two to three times as long as the scan times of participants using conventional technology. This large difference in scan times might account for the smaller height and width sizing errors reported by the phased array participants put forth in the previous section. However, this conclusion is clearly not definitive, as the cause for the smaller sizing errors could also be due to the benefits of phased array technology (e.g., its ability to sweep through multiple angles) compared with conventional technology.

Table 8. Means and Medians of Scan Times

	Phased Array UT Sample	**Conventional UT Sample**
Geometric Mean of Scan Times (minutes)	7.3	2.6
Median of Scan Times (minutes)	7.5	3.0

[8] Measures of central tendency are discussed in introductory statistics books; for example, Bulmer (1979).

6. Table Summarizing Hits, Misses, False Alarms, and Correct Rejections

Table 9 summarizes the hits, misses, false alarms, and correct rejections for each participant.

Table 9. Hits, Misses, False Alarms, and Correct Rejections for Each Participant

Participant	Type of Ultrasonic Equipment	Number of Hits	Number of Misses	Number of False Alarms	Number of Correct Rejections	Hit Rate	False Alarm Rate
A	Phased Array	25	4	8	23	0.86	0.26
B	Phased Array	28	1	5	26	0.97	0.16
C[9]	Conventional	17	12	0	31	0.58	0.02
D[10]	Conventional	27	2	0	31	0.92	0.02
E	Conventional	23	6	3	28	0.79	0.10
F	Conventional	28	1	26	5	0.97	0.84
G	Conventional	23	6	2	29	0.79	0.06
H	Conventional	21	8	10	21	0.72	0.32
I	Conventional	21	8	1	30	0.72	0.03

[9] Participant C had no false alarms. Having a value of 0 or 1 for the hit rate or false alarm rate prevents the use of Equation 1. Therefore, a value of 0.5 was added to the number of hits, correct rejections, misses, and false alarms before calculating the hit rate presented in Table 9.

[10] Participant D had no false alarms. Having a value of 0 or 1 for the hit rate or false alarm rate prevents the use of Equation 1. Therefore, a value of 0.5 was added to the number of hits, correct rejections, misses, and false alarms before calculating the hit rate presented in Table 9.

6. Conclusions and Recommendations

This study focused on quantification of the effectiveness of conventional UT inspection equipment, as well as phased array UT inspection equipment. The information gathered in this study provides users with an estimate of the rates of hits (true positives), false alarms (false positives), misses (false negatives), and correct rejections (true negatives) for conventional and phased array equipment. This information can help to set inspection frequencies and optimize safety. By evaluating the methods and techniques used throughout the railroad industry, improved processes can be implemented to accurately detect, identify, and size flaws in rail.

Nine participants were included in this study. Seven of the participants used conventional UT equipment to inspect the rail specimens, and the remaining two participants used phased array UT equipment. Analysis of the data from the current study suggests that phased array technology is not more effective at finding rail internal defects than conventional technology, but a follow-up study with a revised methodology and additional attempts at controlling certain variables, such as operator experience and inspection times, would provide more conclusive results.

To that end, there are several recommendations for a follow-up study. First, the results from the study could be made more statistically significant if a larger sample size were obtained. This is particularly important for the participants using phased array technology, as there were only two such participants in the current study. Second, there should be an attempt to make the study more uniform and controlled. For example, the participants using conventional technology had significantly more experience with UT testing on sections of rail than the participants using phased array technology. It is hard to compare the effectiveness of phased array technology with the effectiveness of conventional technology when there is such a large experience gap in the operators of the equipment. In addition, inspection time should be limited to a set interval; by so doing, better performance by a set of individuals using one type of inspection equipment will be attributable to their inspection equipment, rather than the discrepancies in experience or inspection times.

It is also important to point out that the information collected in this report, namely the average sensitivity (or probability of detection) of the ground verification process, could be used as an input to a risk analysis model. Oftentimes, the ground verification process is known as “ground truth” because it is assumed that the absolute truth about the presence and size of a defect is being achieved. However, this study clearly demonstrates that, in the realm of rail inspection, there are errors in the ground verification process (for both conventional systems and phased array systems). If one could also learn the probability of detection of the current hi-rail detector car systems, then the two probabilities of detection (manual ground verification probability of detection put forth in this report as well as hi-rail detector car probability of detection) could be used in risk analysis models to optimize inspection frequencies, thereby minimizing the risk associated with an in-service rail failure and increasing the safety of the rail network.

References

Bulmer, M.G. *Principles of Statistics*. New York: Dover Publications Inc., 1979.

Garcia, G.A. *Application of Ultrasonic Phased Arrays for Rail Flaw Sizing* (DOT/FRA/ORD-06/17). Washington DC: July 2006.

Macmillan, Neal A. & Creelman, Douglas C. *Detection Theory: A User's Guide (Second Edition)*. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers, 2005.

Spiegel, Murray R. et al. *Schaum's Outline of Probability and Statistics (Fourth Edition)*. New York: McGraw-Hill Companies, 2013.

U.S. Department of Transportation, Federal Railroad Administration. *FRA Track Inspector Rail Defect Reference Manual*. Washington, DC: 2011.

Wickens, Thomas D. *Elementary Signal Detection Theory*. New York, New York: Oxford University Press, 2001.

Appendix A – NTSB Recommendations Related to Rail Maintenance and Inspection

In 1992, a derailment resulted in a toxic chemical spill into the water supply of Superior, WI. The National Transportation Safety Board (NTSB) formally investigated the accident. On the basis of its findings, the NTSB made the following safety recommendations to FRA and the Association of American Railroads:

- Research and develop inspection methods that will identify internal defects in rail that have significant shelling and other surface conditions. (Safety recommendation R-94-1 and R-94-4)
- Perform the necessary research and develop standards that (1) provide defined limits of allowable rail surface conditions (such as shelling) that can hinder the identification of internal defects, and (2) require remedial action for rail with surface conditions that exceed defined limits. (Safety recommendation R-94-2)
- Correlate rail surface conditions with detail fractures and with rail service failures, including the provisions of samples of rail with detail fractures. (Safety recommendation R-94-5)

In response to more recent derailments, the NTSB has also made the following safety recommendations to FRA:

- Review all railroads' internal rail defect detection procedures and require changes to those procedures as necessary to eliminate exceptions to the requirement for an uninterrupted, continuous search for rail defects. (Safety recommendation R-08-9)
- Require railroads to develop rail inspection and maintenance programs based on damage-tolerance principles, and approve those programs. Include in the requirement that railroads demonstrate how their programs will identify and remove internal defects before they reach critical size and result in catastrophic rail failures. Each program should take into account, at a minimum, accumulated tonnage, track geometry, rail surface conditions, railhead wear, rail steel specifications, track structure support, residual stresses in the rail, rail defect growth rates, and temperature differentials. (Safety recommendation R-08-10)
- Require that railroads use methods that accurately measure railhead wear to ensure that deformation of the head does not affect the accuracy of the measurements. (Safety recommendation R-08-11)

Appendix B – Operator Profile

Operator Profile

Name: ______________________________

Date: _____________________________

Company Name: ____________________________

Job Title/Position: __________________________

Training: ____
____
____

Certifications: ____

Number of Years with Company: ____

Number of Years in NDE: ____

Number of Years on Current Job: ____

Inspection Background: ____
____
____

Ultrasonic Background: ____
____
____
____

Comfort Level with Inspections: ____
____
____

Additional Information: ____
____
____
____

Abbreviations and Acronyms

CFR	Code of Federal Regulations
CSHA	Cross Sectional Head Area
FRA	Federal Railroad Administration
NDE	Non-Destructive Evaluation
NTSB	National Transportation Safety Board
ROC	Receiver Operating Characteristic
TSS	Track Safety Standards
UT	Ultrasonic Testing