Optimisation of risk/cost ratio for the maintenance of in service components through the assessment of NDT reliability

C. Cappabianca, (ENEA Centro Ricerche Casaccia, S. Maria di Galeria, Roma, Italy, E-mail: cappabianca@casaccia.enea.it);
C. Ozzano, (Ansaldo Ricerche srl, Genova,Italy, E-mail: ozzano@ari.ansaldo.it)

Abstract

The topic of NDT reliability was urged to the authors during a preliminary study aimed at verifying the feasibility of traditional ultrasonic inspection of train wheels in service.

At first, the defects to be detected where defined, then test apparatus performance was quantified and finally ultrasonic probes were chosen. Some test blocks with calibrated artificial defects were designed. Keeping in mind that wheel condition should have a considerable effect on test results, two different test blocks were machined, one out of a new wheel and one out of an aged wheel.

As defects characterisation was going on, unexpected factors decreasing test reliability emerged. It was found out that the reproducibility of the results was also depending on some probe details that are usually not mentioned by the probe manufacturer, such as shaping of the probes due to service.

An important role was also played, as expected, by the wheel conditions. This is probably due both to rolling surface shaping and to metallurgic modifications induced by the severe thermal cycles that wheels undergo in service.

In the light of these preliminary evaluations, an informal survey was done on test procedures proposed by partner companies. It was found out that only a few of the named factors were duly considered; for example, acceptance criteria are not dependent upon component wear (system calibration being only performed on reference blocks machined out of new wheels). Moreover, when multiple channel system are employed, probe sensitivity is assumed to be the same.

Some key questions arose:

- Is it possible that, in some applications, acceptance criteria were set to conservative levels because NDT reproducibility seemed poor, due to lack of information on relevant parameters?
- If this holds true, how a better confidence in NDT reliability could affect the acceptance criteria and inspection scheduling?
- Could this help in the risk/cost ratio optimisation?

In this paper the authors are proposing tentative answers.

1. Introduction

In the framework of a project aiming at developing a monitoring system for in-service train wheels by the means of ultrasonic inspection, an impressive data dispersion was observed on the investigation of Flat Bottom Holes (FBHs) with identical rated characteristic. As the noticed phenomenon was far from being negligible, it was decided to carry out a new experimental campaign, explicitly pointing to aspects concerning reliability, in order to get a clearer picture of the emerging problem.

A specific issue was defined to avoid variables proliferation: the detection of FBHs ϕ3mm parallel to the rolling surface, simulating circunferential defects at an early stage.

Test blocks were machined out of both new and aged wheels (Fig. 1): FBHs ϕ3 were drilled having different depth and axial position. Tests were then carried out with a set of four probes with identical rated characteristics. Four different operators performed the tests.

Leaving now aside discussions on how much FBHs are representative of real defects (orientation, shape, roughness,…), it is necessary to clearly identify all the factors which can interfere with reliable detection of defects in the wheel.

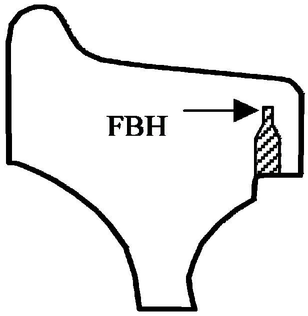

Fig. 1: sketch of FBH representing a delamination parallel to the rolling surface

This problem has a special meaning in those cases where acceptance criteria for indications coming from ultrasonic inspections are based on defects size; here, size estimation is based on the amplitude of the reflected echo, increasing with the echo's amplitude itself. As a consequence, any factor bringing an echo loss is causing a dangerous underestimation of the defect size. For this reason, a deeper understanding of the problem is necessary.
During the preliminary phase, many factors having an influence on the tests were singled out: some were investigated in the activity here described, while others will be subject for subsequent work.
The authors would like to emphasise that no exhaustive considerations are coming yet from the present study, whose main aim was to underline a phenomenon that, if neglected, could bring severe consequences, especially in the light of the growing importance being ascribed to the in-service inspection of rolling stocks.

2. Defects detection and sizing

Sizing of defects much larger than the beam section comes out more from the number of subsequent positive calls than from the evaluation of reflected echoes amplitude (investigation is supposed to be done in pulse-echo mode). Nevertheless, as a more desirable aim would be the detection of defects as small as possible, sizing must be based on the reflected echo amplitude. For this reason, leaving aside the many possible approaches to define acceptance criteria, sizing should be carried out with the highest reachable reliability.
Standard current procedures prescribe to estimate the defects size by comparing the reflected echo with the echoes coming from reference FBHs machined on calibration blocks; as obvious, a safety factor is taken into account when performing the tests on real wheels, but:

- Is this factor really safe?
- If it is, isn't it too much safe?

No answer was up to now given to these questions, mainly because acceptance criteria were based on experience and no severe problems occurred, so that a deeper understanding of the problem was never urged.
As already said, defect sizing is carried out comparing the reflected echo amplitude to the echo coming from FBHs of a calibration block. But, it must be considered that there are other differences in the measuring apparatus that lead to uncertain conclusions.
Table 1 summarises the identified types of errors.

Table 1: type and category of error sources

Error category	Error type		
Test apparatus Intrinsic Capabilities (IC)[1]	Probes: statistic dispersion of sound beams	Probes: wearing	Probes: overlapping of sound beams for multiple probes set
Test piece condition (AP)	Aged material	Worn-out surface	Type of material
Human factor (HF)	Operator		

3. Error quantitative estimation

3.1 Ultrasonic probes: ageing

Double crystal twin focussed probes, 4 MH, were employed, which were previously slightly worn out through an automatical device in order to simulate natural wearing. Such probes allow the rim investigation for depth in the range 2÷50 mm.
As tests were going on, it was noticed that reflected echoes amplitude was different if the probe was rotated by 180°. Fig. 2 is an example of comparison of probes response in case of orientation 0° (identification label towards the operator) and 180°. The phenomenon was noticed on all the tested probes, but the dispersion was quite high.

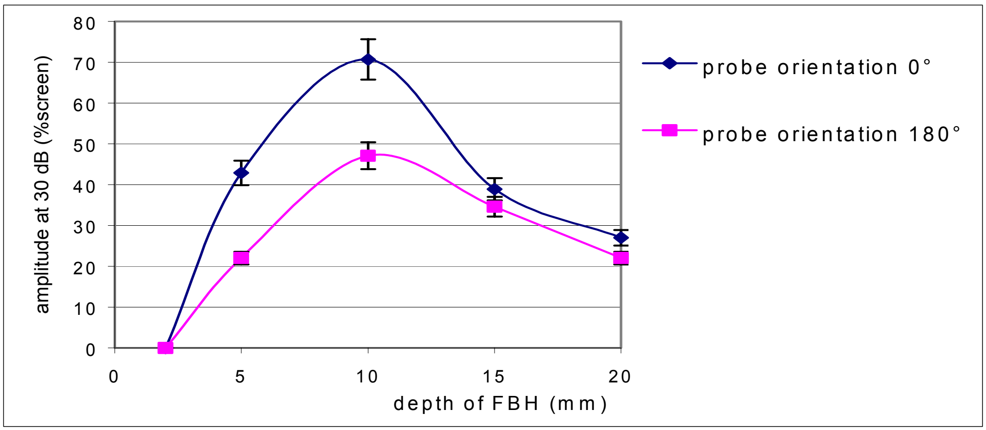

Fig. 2: a 4 dB dispersion is noticed depending on orientation

As this behaviour was discovered during an advanced stage, it was decided to plan a new test specifically aimed at studying the wearing effect; so, a new probe was acquired and tested both as new and after wearing (Fig. 3).
Two aspects could be noticed: an improving of sensitivity for the used probe, probably caused by a better coupling due to shaping, and the arising of a response dependence on orientation due to usage, probably produced by the fact that, as the wheel surface is slight inclined, the shaped surface is not symmetrical any more.

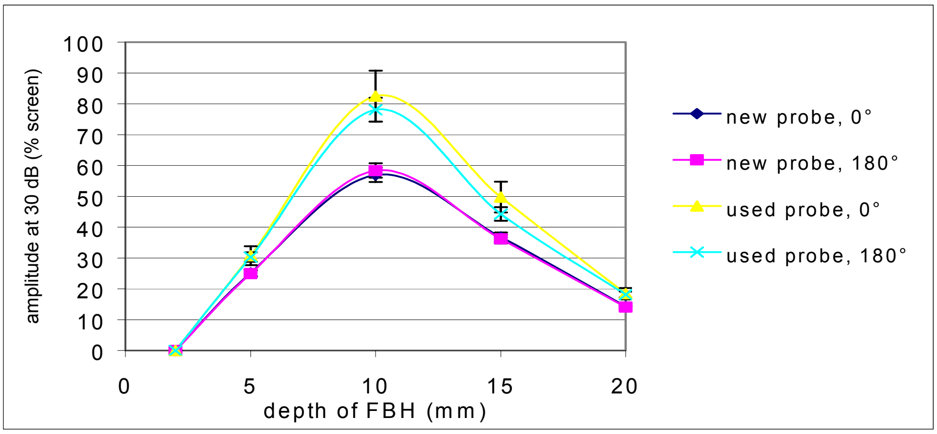

Fig. 3: used probe shows higher response, but a dependence on orientation (though slight in this case) is arising.

3.2 Overlapping of probes array sound beam

Reflected echoes amplitude is as higher as smaller is the interaxis probe-FBH. It follows that, if the rolling surface investigation is performed with a set of probes held by a holding device, the probes interaxis must be properly calculated to guarantee a satisfactory scanning.

While evaluating the sound beams overlapping, it must also be taken into account that it depends on depth, especially for focussed probes; tests were carried out on ϕ3 FBHs at 5-10-15 mm depth. The comparison of the energy decay curves versus probe-FBHs interaxis is given in Fig. 4.

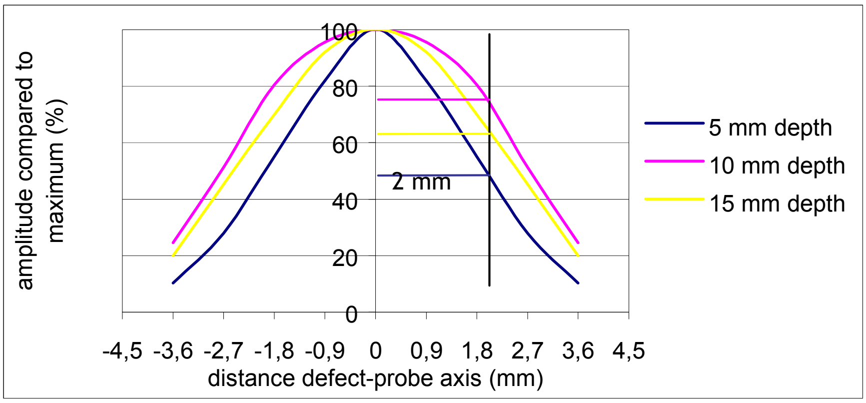

Fig. 4: energy decay curves versus probe-FBHs interaxis.

As probes have their own size, their interaxis cannot be reduced as desired when using a single row. It is anyway possible to reduce theoretical interaxis by using multiple parallel arrays.

3.3 Probes: differences among units with identical rated characteristics

Verifications performed on four probes of a set showed that the existing difference in behaviour wasn't negligible (Fig. 5), either as new or after wearing.

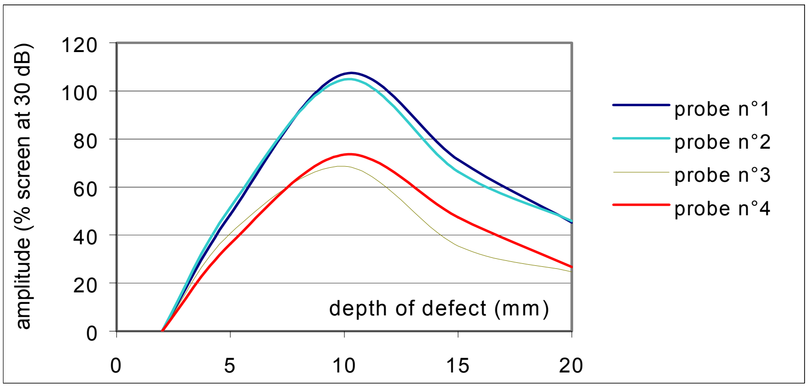

Fig. 5: a 4-dB dispersion was noticed for the reflected echo amplitude of the units of the tested lot (curves obtained from mean values are here used).

3.4 Material (type and condition)

Another important test variable is the wheel material, as some types of steel are more 'friendly' to ultrasonic inspection than others. Moreover, thermal fatigue due to service induces structural changing in the wheels, which must be considered as calibration blocks are always drawn from new wheels. This aspect was not specifically tested in this work.

3.5 Wheel's surface wearing

Comparing the reflected echoes amplitude of FBHs machined on new and used wheels, a 10 dB dispersion was observed, as showed in Fig. 6 where experimental data from a new wheel and a used one are compared. Together with the curves (mean values) obtained form FBHs on new and aged wheels, indications from two critical points are visible, located at the boundaries of the portion which can be tested by means of ultrasonic probes (worn out wheel).

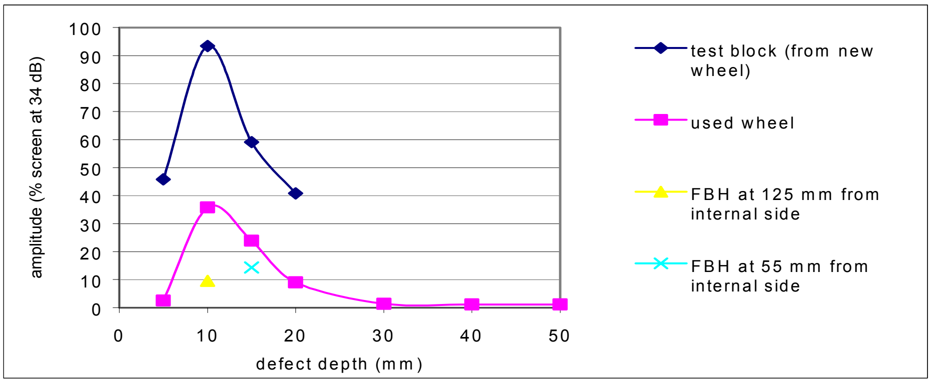

Fig. 6: Data obtained from an aged wheel are compared to those coming from a new wheel.

3.6 Operator

Among the considered factors, the human factor showed a minor influence, even though operators with different kind of experience were selected. This is not surprising, as the defects' size and location were well known.

4. Factors composition

For each factor, starting from the statistical distribution of an adequately rich data collection, it is possible to define the energy loss that has to be considered in order to guarantee the desired POD.

Let's consider, as an example, a complex application for which a suitable POD has been chosen.
For a reliable ND evaluation, each influencing factor should have been previously characterised by dedicated tests: for each, the relation POD-'energy loss to be taken into account' must be known.
In case many factors are present, the composed total POD, let's call it POD_t, comes out multiplying the many specific POD, let's call them POD_i: as a consequence, when a proper POD_t has been defined, a POD_i array has to be selected in order to generate an adequate acceptance criteria.

An example:
Let's assume that POD_t, the minimum POD which can been accepted for the application, is 90%.
'a', 'b', 'c', 'd' are the influencing factors: then (0.97, 0.97, 0.97, 0.97) and (0.96, 0.96, 0.98, 0.98) are two possible arrays (POD_a , POD_b , POD_c , POD_d) satisfying the problem.
It is clear that the probability to find an acceptable solution decreases as far as the number of influencing factors increases.
Once found out the best POD_i array, for each factor the energy loss to be considered is calculated on the basis of relevant data collections. It is now possible to define the acceptance criteria of the ND investigation on the basis of the total loss, obtaining summing up the calculated values.
Coming back to the ultrasonic evaluation of train wheels, supposing that the results described in the previous sections are not coming from preliminary tests but from exhaustive investigations, supposing that POD_t is defined and proper POD_i have been chosen, it would be possible to calculate the total loss as follows:

Wheel wearing:	-10, +0	dB,
Differences among units:	-4, +0	dB
Beams overlapping:	-4, +0	dB (interaxis 5mm, neglecting depth effect)
Probe wearing:	-0, +3	dB
Total loss:	*-18, + 3*	*dB.*

That is: to guarantee the defined POD_t, the level to be set during tests must be at least 18 dB lower than the value coming from tests on the calibration block. This means that, *supposing that a wheel must be removed from service when a Φ 3mm defect is detected, the risk to reject a wheel because of a Φ 1mm defect must be accepted.*

5. How can the solution be improved?

Generally POD_i cannot be changed, even though quite 'unconfortable', but sometimes the situation can be improved reducing the uncertainty related to each factor, that is, increasing knowledge.
In order to propose an example about possible methods to reduce the test severity without decreasing reliability, let's come again to the proposed example, that is a wheel being inspected with a multiple probes set.

5.1 Probes: wearing

As the geometry of in-service wheels is critical, it is advisable never to change probes position once that the run-in of the set has been performed: in fact, as it was found out, probes shaping depends on their position on the wheel, and it has a great influence on coupling quality. Moreover, it is also advisable to record the DAC of each probe and to use a single channel for each one: if multiplexing is necessary, proper selection of probes to be managed with the same channel must be carried out.
A slight preliminary wearing of new probes should be also done, to avoid in-service changing of performance, even though an improvement can occur.

5.2 Beams overlapping

Beams overlapping can be optimised reducing the probes distance and placing them on multiple rows: unfortunately, the probes number is limited by the system capability. Once the optimum probes set is defined, the minimum energy loss related to overlapping is found.

5.3 Probes: difference among units with identical rated characteristics

Once again, characterisation and separate managing of each probe of the set is necessary. As a more economical solution, a properly calculated gain can be set for each probe when using multiplexed solutions.

5.4 Material

Material must be taken into account: information on energy loss due to ageing and material composition must be evaluated by dedicated tests. Once such information is available, acceptance criteria can take benefit of it. Only in case information on the wheel being tested is not available, the safest energy loss level must been considered.

5.5 Used wheel worn-out surface

Two are the possible solutions, depending on the degree of data processing which is allowed by the available system. The most profitable method is to acquire the quantitative relation 'axial position versus energy loss' by means of properly designed test blocks. This function could be employed as signal modulating curve, that is, for each probe an energy loss value could be applied on the basis of its axial position. The simplest solution is to adopt the most conservative loss value.

5.6 Operator

As mentioned, the human factor was negligible in the activity here reported.

6. Revised example

Coming back to the considered case, the following improvements could be reached:

Wheel wearing:	-5, +0	dB, (supposed corrected value)
Differences among units:	-0, +0	dB (probes are singularly managed)
Beams overlapping:	-4, +0	dB (unvaried)
Probe wearing:	-0, +3	dB
Total loss:	*-9, + 3*	*dB.*

That is: the minimum defect which can lead to wheel rejection is Φ 1,8 mm instead of Φ 1 mm as in the previous situation.

7. Future work

The tests described in this paper and the associated considerations were not aimed at proposing final solutions, but, on the contrary, at trying a definition of the problem of in-service wheels inspection and proposing provisional solutions taking advantages from discussions on reliability.
A more detailed analysis would need a huge experimental campaign, to be performed with a proper set of specifically designed test blocks.
A round-robin test involving different test labs would surely help to perfect the approach and collect relevant information.

8. Conclusions

To conclude the discussion, tentative answers to the arisen questions might be proposed.

Is it possible that, in some applications, acceptance criteria were set to conservative levels because NDT reproducibility seemed poor, due to lack of information on relevant parameters?
The final goal of the activity, whose preliminary phase has been here described, is to set up an acceptance criteria for ultrasonic inspection of in-service train wheels which could be based on real knowledge of the application.
This is necessary as the fact that current criteria always proved to be safe and severe accidents due to wheels break-down never happened (speaking of Italian reality) doesn't mean that their they are the safest and the most convenient from the economical point of view; moreover, it could happen that, due to the evolution of railways transportation and the introduction of new materials, a revision of current criteria is required.
Only a more detailed knowledge of the factors affecting the ND evaluation can allow a better awareness in the definition of inspection procedures.

If this holds true, how a better confidence in NDT reliability could affect the acceptance criteria and inspection scheduling?
A more accurate POD isn't by itself enough to carry on a really profitable review of maintenance strategies, but it can be regarded as a fundamental tool in a more sophisticated and innovative method that the authors are planning to propose. The new approach that is being faced is aimed at supplying, starting from the defect ND characterisation and from information on service load conditions, an estimation of the residual life, calculated on the basis of fracture mechanics. To reach this goal, it is therefore necessary to have a good estimation of the reliability related to defects characterisation.

Could this help in the risk/cost ratio optimisation?
It comes out that a higher confidence in the reliability of periodical inspections will surely have an impressive effect on maintenance strategies, especially in case a residual life estimation would be possible thanks to reliable ND evaluations.

8. References

[1] C. Müller, M. Elaguine, M. Scharmach, B. Redmer, U. Ewert, L. Schaefer, P. Th. Wilrich", Reliability investigation of NDT systems by modular analysis of recorded data", 8th ECNDT Barcelona 2002