PROBABILITY OF DETECTION OF AUTOMATED TAP TESTING FOR DISBOND DETECTION IN METALLIC HONEYCOMB STRUCTURES

Marc Genest, Shashank Pant, Muzibur Khan, Dmitrii Klishch, Catalin Mandache

[1] National Research Council Canada, Aerospace Research Centre,
marc.genest@nrc-cnrc.gc.ca

Abstract: The reliability of automated tap testing for honeycomb composite panels representative of those found in helicopter structures is investigated. A Probability of Detection (PoD) study was carried out for detecting disbond on honeycomb panels using manual and automated tap testing techniques. A process was developed to generate controlled disbond size in the honeycomb panels. The process was confirmed via ultrasonic testing and micro-computed tomography inspection. A reference and eight test panels containing a total of 70 different damage sites that included dents, disbonds, as well as combined dent and disbond were inspected by 11 inspectors using both manual and automated tap testing techniques. Overall, the $a_{90/95}$ value decreased by 15 mm (0.6 inch) equivalent diameter using the automated tap test as compared to manual tap testing. Furthermore, the automated tap testing led reduction in false calls, up to 3 times. It is important to note that inspections were carried out in ideal laboratory conditions, and, most likely the advantages of the automated tap tester over the manual tap test would be even more significant in noisy hangar environment.

Keywords: probability of detection, tap testing, metallic honeycomb, disbond detection

INTRODUCTION

The primary inspection method of honeycomb composite structures in the aircraft of interest in this work is by means of tap testing, which is performed by non-certified non-destructive testing (NDT) personnel. Current approved maintenance procedures to detect disbonds and crushed core in such structures involve having an operator to listen for a change in sound response after tapping with a hammer or other objects, and then infer changes in sound signature to the presence of disbonds or defects. It has been demonstrated that structural integrity of such panels depends largely on the exact type and amount of damage present, which makes differentiating the damage types important. It is suspected that current tap testing procedures result in an abundance of false calls and misidentification damage (disbond versus crushed core). Improving efficiency and reducing the false call rate of tap testing applied to honeycomb panels is highly desired.

In this work, a new process has been developed to generate controlled disbond without surface dent in honeycomb panels. Tap testing of these panels was carried out as per procedure developed. The aim of this work was to understand the differences in probability of detection (PoD) between manual tap testing and automated tap testing. The following sections describe the method developed to generate disbond in the test panels, and the tap test results and PoD curves obtained.

Tap testing

Tap testing is one the oldest methods for inspecting laminated, sandwiched, or bonded composite structures [1][2]. In this method the operator lightly taps the structure with a coin (approx. 25 mm in diameter) or a small tap hammer (max 2 ounces), and listens to the sound emitted by the structure. The characteristics of the impact due to tapping depend on the local stiffness of the structure in response to the coin or hammer being used, which can be altered due to the presence of damage. The acoustic response is compared to that of a healthy structure which produces ring as compared to the flat / dead / dull response by the presence of damage [1]. Tap testing methods are widely used as a quick evaluation procedure to detect delamination, disbond, and poor cure in composite structures.

Manual tap testing is highly dependent on the inspector's ability to hear and interpret the results. Therefore, any adverse working environment such as background noise can have a negative impact on the ability to detect damage. Automated tap tester allows reducing human errors by instrumenting the traditional tap hammer with force transducer to record the force-time history [3].

Probability of Detection

The reliability of the non-destructive inspection (NDI) techniques is quantified in terms of the probability of detection (PoD). The concept of PoD is used in various industries, including aerospace, to establish the capability and reliability of an NDI technique to detect flaws for a specific inspection application. PoD is generally expressed as a curve, which is a statistical measurement of the likelihood, with a specified confidence level, of detecting flaws against a characteristic parameter of the flaw (usually its size). Estimation of a PoD curve typically relies on the manufacture of large numbers of realistic defect specimens, followed by practical trials of the inspection procedure. The detectability of a defect is related to its size, human factors, defect geometrical characteristic, and many other operational conditions. PoD metrics have a variety of uses and provide a means to make improvements in inspection techniques, test procedures, or NDI systems, to establish the safe inspection interval and to perform risk assessments [4]. Availability of inspection PoD data is essential for the application of a suitable life cycle management strategy whether based on retirement-for-cause or damage tolerance approaches [5].

In quantifying NDI reliability using PoD curve, the $a_{90/95}$ value quantifies an inspection capability in terms of defect size. It purports to be the size of the target having at least 90% probability of detection in 95 of 100 experiments under nominally identical conditions [5]. Several factors influence the ultimate $a_{90/95}$ size including the statistical models used in estimation.

Various probability distribution models have been used in deriving equation for constructing the PoD curve as a function of discontinuity dimension, using curve-fitting methods of PoD analysis as demonstrated in [6] and also recommended in the United States Department of Defense MIL-HDBK-1823A [7]; however, log-logistic and log-normal are the most commonly used. Berens and Hovey investigated different methods of modeling NDI data to estimate PoD and recommended that the log-logistic distribution model was the most consistent distribution for modeling NDI data [8]. For this study, which generated "hit-miss" data, both log-normal and log-logistic probability distribution models were used. In log-logistic distribution model, which is also called log-odds or logit models, the mathematical relationship of probability of detecting a discontinuity of size, "a", can be expressed by the following equation:

$$POD(a) = \frac{e^{\{\beta_0+\beta_1 \, ln(a)\}}}{1+e^{\{\beta_0+\beta_1 \, ln(a)\}}} \tag{1}$$

Where, *PoD (a)* is the probability of detecting a discontinuity with size *a*, β_0 and β_1 are the constant parameters for curve fitting [4], where maximum likelihood method provides more reasonable estimation of these parameters.

Parameters β_0 and β_a in equation (1) do not have any physical interpretable term and another mathematical equivalent form of log-logistic model can define by following equations [9] which have more meaningful physical parameter (μ and σ).

$$POD(a) = \frac{1}{1 + e^{-\left[\frac{\pi}{\sqrt{3}}\left(\frac{\ln a - \mu}{\sigma}\right)\right]}} \quad (2)$$

In this form of log-logistic model μ is median detectable crack size (ln $a_{0.5}$) and σ is the standard deviation.

The MIL-HDBK-1823-1999 suggests using the log-normal model along with a curve fitting regression analysis for estimating PoD. This method does not include the false calls in the PoD calculations. Spencer extended the log-normal curve fitting model considering the negative effects of the false calls on PoD [10]. In log-normal distribution model, a continuous random variable, X, follows a log-normal distribution if its natural logarithm, ln(X), follows a normal distribution. Reference [8] suggested the cumulative log-normal distribution for modeling PoD data. Cumulative log-normal distribution for modeling PoD data is expressed by the following equation [11]:

$$POD(a) = \varphi\left(\frac{\ln a - \mu}{\sigma}\right) \quad (3)$$

Where, *a* is the discontinuity size (i.e. length or depth or area), $\Phi(z)$ is the cumulative log-normal distribution function, μ and σ are location and scale parameters of the PoD curve. The location and scale parameters were estimated using a maximum likelihood procedure to find the parameter values which maximized the observed data [9]. The variation in the confidence bound is mainly dependent on the data scatter and the specific procedure used for confidence interval calculation.

The tap testing PoD estimation from "hit-miss" data was carried out using log-logistic distribution model and statistical analysis software in R as detailed in the MIL-HDBK-1823 [7].

DAMAGE INTRODUCTION

Honeycomb sandwich panels with aluminum core and aluminum skin were used to introduce damage at various sites needed for the PoD experiment. The aluminum-aluminum (Al-Al) honeycomb sandwich panels were representative of some of the RCAF Griffon helicopters' panels. A Federal Aviation Administration (FAA) approved Al-Al honeycomb panel, 19 mm (0.75 inch) thick, was procured and cut into smaller 305 x 305 mm (12x12 inch). The panel has 0.5 mm (0.020 inch) thick 2024-T3 clad aluminum skins, and a 5052 aluminum core having 4.76 mm (3/16 inch) cell size and a 91 kg/m^3 (5.7 lbs/ft^3) density.

Damages at different sites were introduced in the test panels by using a combination of ultrasonic welder and impact drop tower. The ultrasonic welder approach (Figure 1) was employed at 1500 J and 138 kPa (20 PSI) pressure was applied to a piston to push the welder against the honeycomb panel surface as illustrated in Figure 1. It is known that impacts can result in dents and crushed cores, with or without any disbond. Therefore 3 types of damage features were introduced into the panels: 1) dent from impact damage (no-disbond); 2) disbond made with ultrasonic welder (disbond – no dent); and 3) disbond with dent, where the disbond were first generated using the ultrasonic welder approach and the disbond location was thereafter impacted at low energy level.

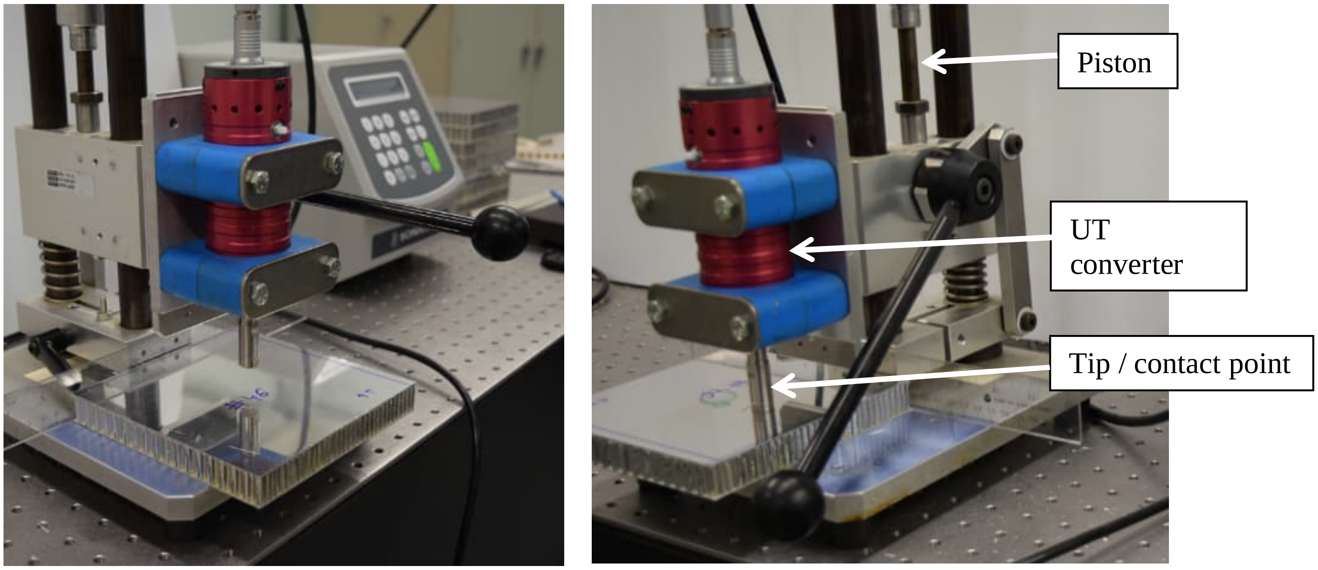

Figure 1: Ultrasonic welder setup to generate disbond.

The flaw distribution was set based around the 12.7 mm (0.5 inch) target flaw size to be detected, which field inspectors believe to detect using manual tap-testing. Thus, 6 to 25 mm (0.25–1.0 inch) diameter flaws were divided into five discrete sizes of 6, 10, 13, 19, and 25 mm (0.25, 0.375, 0.5, 0.75, and 1inches); and distributed throughout the specimen set. One large 76 mm (3 inch) diameter flaw was added to provide a flaw that should be found by all inspectors.

Based on MIL-HDBK-1823A [7] more precise estimates of a_{90} and narrower confidence bounds on the PoD curve results from target size that are uniformly spaced on a Cartesian scale and therefore this is the new recommended practice. Therefore, the flaw generation targeted the following damage sizes in the PoD specimen set.

- 12 x 6 mm (0.25 inch) diameter disbond
- 12 x 10 mm (0.375 inch) diameter disbond
- 12 x 13 mm (0.5 inch) diameter disbond
- 12 x 19 mm (0.75 inch) diameter disbond
- 12 x 25 mm (1.0 inch) diameter disbond

In addition, a minimum of 51 mm (2 inch) separation between flaws was maintained to eliminate signal cross-talk; expect for a few flaw pairs that are closely clustered to study ability to define flaw boundaries. A few dents with 1, 2 and 4 J were also included on the PoD panel set, to confuse the inspector, however those are not considered damage for this PoD study as we were only concerned about the disbonded areas. A minimum of 38 mm (1.5 inch) distance from flaws to edge of panels was maintained to avoid any edge effect. A reference panel containing a sound area, a 25 mm (1 inch) diameter dent, a 25 mm (1 inch) diameter dented disbond, and a 12 mm (0.5 inch) diameter disbond was provided to the inspector to be used as a calibration/reference sample.

Damage areas were sized by performing ultrasonic through-transmission measurements. Despite the targeted size above, the resulting damage size varies slightly from the targeted diameter, and are obviously not a perfect circle. Thus, the areas were converted into equivalent diameter to perform the PoD analysis. Ultrasonic inspections were also carried out after the completion of all the PoD tests, to confirm that no new damage was present and no damage grew during the test.

TAP TESTS INSPECTION RESULTS

Tap Test Summary Results

To ensure that all inspectors received the same information, an experimenter briefing and information package was developed and distributed to all 11 inspectors who performed the tap testing inspections. Photos of the inspection results were analysed and compared with the location of the flaws. Although the inspectors were to mark the outer diameter of the various damage sites, the sizing was not taken into consideration, as the ruler system and position of the panels was not consistent from inspector to inspector; despite the instruction provided in the information package.

From the collected photos, various disbond sites were identified and labelled as hit or miss; and false calls were also identified. Hit-Miss summary of the manual and automated inspections are provided in Tables 1 to 3. It was noticed that several of the false calls obtained with the automated tap test came from values close to the threshold of 0.32 provided in the information package. The observer also noticed that some of the inspectors kept using the automated tap tester despite the low battery LED being turn on. Despite being clearly informed to replace the batteries as soon as the LED battery turns on. For this reason, it was decided to investigate the effect of increasing the threshold to a higher value (0.34) as shown in Table 3, which substantially reduced the number of false calls while marginally decreasing the number of hits.

Table 1: Summary of Manual Tap Test Inspection

Inspector	*HIT*	*MISS*	*FALSE CALL*
Inspector 1	45	24	1
Inspector 2	58	11	8
Inspector 3	39	30	1
Inspector 4	32	37	2
Inspector 5	49	20	0
Inspector 6	40	29	7
Inspector 7*	62	7	7
Inspector 8	49	20	1
Inspector 9	36	33	5
Inspector 10	62	7	1
Inspector 11	53	16	10

Table 2: Summary of Automated Tap Test Inspection – normal threshold

Inspector	*HIT*	*MISS*	*FALSE CALL*
Inspector 1	41	28	1
Inspector 2	54	15	3
Inspector 3	59	10	1
Inspector 4	55	14	0
Inspector 5	60	9	0
Inspector 6	39	30	10
Inspector 7*	36	33	0
Inspector 8	50	19	2
Inspector 9	46	23	0
Inspector 10	66	3	11
Inspector 11	57	11	2

*inspector spent a significant more time on the manual tap test (>2hrs) than the automated tap test (<1hrs)

The manual inspection yielded an average of 48 hits and 21 misses per inspector, and a total of 43 false calls; while the automated inspection yielded an average of 51 hits, 18 misses, and a total of 30 false calls. The false calls that coincide with a dent only locations were 15 for the manual tap test, and 20 for the automated tap test.

By increasing the threshold to 0.34, as seen in Table 3, the number of false calls was reduced from 30 to 13, with an average of only 1.1 false call per inspector, and only 6 caused by dent damage, with the majority of inspectors having 0 false call. However, it also increased the total number of misses by 16, for an average of 50 hits and 19 misses. Most of the false calls were below 0.35, with only one inspector having a single false call above that value.

Table 3: Summary of Automated Tap Test Inspection – adjusted threshold

Inspector	*HIT*	*MISS*	*FALSE CALL*
Inspector 1(a)	41	28	0
Inspector 2(a)	52	17	2
Inspector 3(a)	59	10	1
Inspector 4(a)	55	14	0
Inspector 5(a)	60	9	0
Inspector 6(a)	29	40	6
Inspector 7(a)	35	34	0
Inspector 8(a)	49	20	2
Inspector 9(a)	45	24	0
Inspector 10(a)	66	3	0
Inspector 11(a)	56	12	2

Whenever feasible, the observer made notes of the inspection time and interview the inspector after they had carried out the inspection. Almost all inspectors spent as much time on the manual tap test then the automated tap test and most inspectors took between 1 and 2 hours to inspect the 8 panels. Except for inspector 7, who spent over 2.5 hrs for the manual tap test, and less than 1 hour for the automated tap test. This much shorter time while performing the automated tap test significantly affected the detection success (HIT) and therefore affects the $a_{90/95}$ PoD value obtained for the automated tap test. For this reason, another PoD curve was obtained by removing data from inspector 7, as this would be an issue of human factor (probability of inspection) rather than PoD. Additionally, PoD curves were generated by removing the data from both the best and the worst inspectors.

The damage location on the panel, and proximity between damage sites did not influence the inspection results, as no detection pattern was noticed among all inspectors.

Although the large 76 mm (3 inch) diameter damage was detected by all inspectors, some inspectors identified it as several small disbonds, while other grossly undersize the damaged area, with only a few marking the boundaries more closely aligned with the ultrasound indication.

Tap Test PoD Analysis

As expected, generally the smaller flaws tend to be missed more often than the large ones. However, some of the smallest flaws with less than 6 mm (0.25 inch) in diameter were detected by a few inspectors in the manual inspections, while they were all missed in the automated inspection. One possible reason for this, could be related to the tapping frequency and sweep speed used by the inspector. It is also possible that the inspector all missed those locations (did not inspect) or that the changes in material property is simply too small for the automated tap tester to detect the differences. It should also be noted that the tip diameter of the manual tap test was 3 mm (0.125 inch) while the tip diameter of the automated tap test was 6 mm (0.25 inch) which could also explain why the smaller damage were missed with the automated tap test.

The experiment was repeated after the PoD trial to confirm that changes can be detected with the automated tap test, if inspected properly. Values of 0.31 - 0.32 were obtained and confirmed that those create a change and lead to an amber light on the automated tap tester. However, minute changes in location quickly change the value to 0.30, which is below the threshold of detection.

Also troublesome was a 18 mm (0.75 inch) diameter flaw missed by all inspectors during manual tap test, while this site does sound slightly different, it was easily picked up by the automated tap tester with a value of over 0.34 in the post-test verification experiment.

A summary of the PoD values obtained is presented in Table 4. The logit link function was selected as it resulted in minimum variance in almost all cases. As can be seen in the Table 4, the $a_{90/95}$ value is consistently better for the automated tap test. For the inspector with the poorest detection, the manual tap test the PoD curve would not converge and therefore no $a_{90/95}$ was found. The improvement in PoD value for the best inspector was marginal; decreased by 3 mm (0.12 inch) from 25 to 22 mm (1.0 to 0.88 inches), while the overall $a_{90/95}$ value decreased by 15 mm (0.6 inch) equivalent diameter using the automated tap test compared to manual tap testing. Modifying the threshold from 0.32 to 0.34 did not significantly change the $a_{90/95}$ value. Example of logit PoD curve obtained are presented in Figure 2, while examples of log-normal approach are presented in Figures 3 to 4.

Table 4: a_{90} and $a_{90/95}$ equivalent diameter summary results.

Manual Tap Test					
	Logit			*Log-Normal*	
	a_{90} (mm)	$a_{90/95}$ (mm)		a_{90} (mm)	$a_{90/95}$ (mm)
Manual Cumulative	35	41.8		50.4	59.2
Manual Best Inspector	13.77	25.4		14.6	31.6
Manual Worst Inspector	43.3	N/A*			
Manual – minus worst and best inspectors	**35.3**	**43.3**			
Automated – Woodpecker					
	Logit			*Log-Normal*	
	a_{90} (mm)	$a_{90/95}$ (mm)		a_{90} (mm)	$a_{90/95}$ (mm)
Automated Cumulative	26.2	29		29.3	32.6
Automated Cumulative Modified Threshold	27.4	30			
Automated Best Inspector	6.2	22.5		7	14.3
Automated Worst Inspector	35.8	70			
Automated Cumulative – minus worst and best inspectors	**25.1**	**28**			

*No convergence, value not available

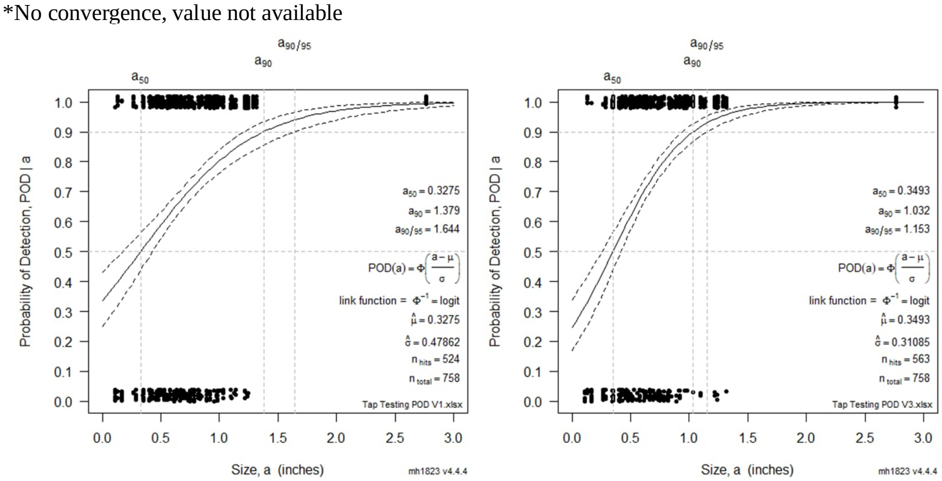

Figure 2: PoD curve – logit - cumulative for (left) manual tap test data, and (right) automated tap test data.

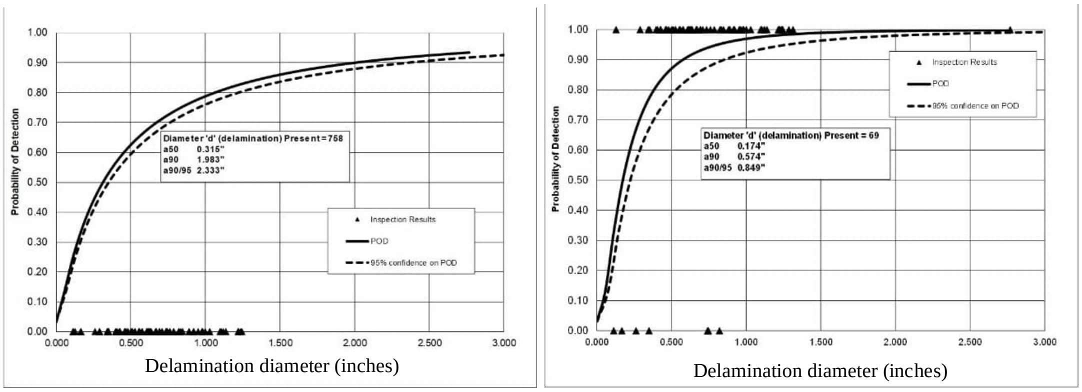

Figure 3: PoD curve manual tap test log-normal (left) cumulative, vs (right) best inspector.

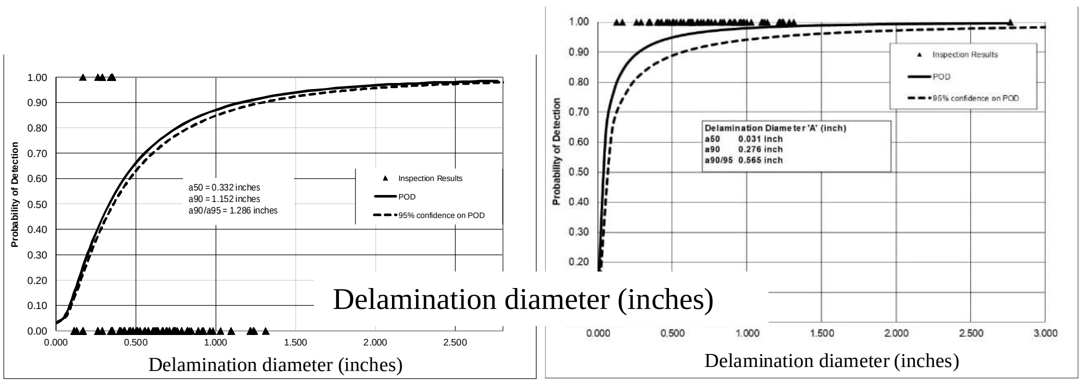

Figure 4: PoD curve automated tap test log-normal (left) cumulative vs (right) best inspection.

CONCLUSION

Manual tap testing performed by non-certified NDT personnel is used as a primary inspection method to inspect honeycomb composite structures of a helicopter fleet; and is highly dependent on the inspector's ability to hear and interpret the results. Therefore, any adverse working environment such as background noise can have a negative impact on the ability to detect damage. The goal of this project was to evaluate the performance of automated tap testing compared to manual tap testing, with the hope to improve inspection efficiency and reduce false call rate. Therefore, probability of detection (PoD) curves were generated for both manual tap and automated tap testing methods and corresponding $a_{90/95}$ values were obtained.

For inspecting the panels, and to generate the PoD curve, 11 inspectors were selected to carry out manual tap testing using a tap hammer and automated tap tester. All inspectors were provided with the same experimenter briefing package before carrying out the experiment. A "hit-miss" PoD analysis was performed using all the data acquired from the inspectors.

The PoD analysis results showed that the manual inspection yielded an average of 48 hits and 21 misses per inspector, and a total of 43 false calls; while the automated inspection yielded an average of 51 hits, 18 misses, and a total of 30 false calls. The false calls that coincide with a dent only locations were 15 for the manual tap test, and 20 for the automated tap test. It was noticed that several false calls obtained with the automated tap test came from values close to the threshold of 0.32, as provided in the information package. By increasing the threshold to 0.34, the number of false calls were reduced from 30 to 13, averaging 1.1 false call per inspector, and only 6 caused by dent only damage, with the majority of inspectors having no false calls. However, it also increased the number of misses by 16, for an average of 50 hits and 19 misses, which remains better than manual tap testing. Although the performance of automated tap testing overall is better, it did not result in reduction of inspection time; as both manual and automated tap testing took similar amount of time to carry out.

It was found that, in general, smaller flaws were missed more often than the large ones. However, some of the smallest flaws measuring 6 mm (0.25 inch) in diameter were detected by few inspectors using manual tapping, while they were all missed during automated tap testing by the same inspectors. This could be due to high tapping frequency and sweep speed, and / or because of the tip diameter of the automated tap tester being 6 mm (0.25 inches) as compared to the 3 mm (0.125 inches) for manual tapper. One area for potential improvement would be to machine the impact head used on the automated tap testing to make it smaller, and thereby, making it more sensitive to small damage. The current 6 mm (0.25) inch diameter hammer tip seems to be limiting the sensitivity.

Overall, the $a_{90/95}$ value decreased by 15 mm (0.6 inch) equivalent diameter using the automated tap test from 43 to 28 mm (1.70 to 1.10 inches); while this value is still above the targeted value the automated tap test performed better than manual. Moreover, it is important to note that these inspections were carried out in laboratory condition, and that the advantage of the automated tap tester over the manual tap test would likely be even more significant in noisy environment.

The significant reduction in false calls obtained using the automated tap test as compared to the manual tapping looks promising and a reasonable justification to employ automated tap testing. In the worst case, even if inspectors are reluctant to changes, it would be a good tool to confirm findings obtained by manual tap testing, which could significantly reduce the number of false calls.

REFERENCES

[1] Cawley, P., Adams, R. D. (1988), "The Mechanics of the Coin-Tap Method of Non-Destructive Testing", *Journal of Sound and Vibration*, vol 122, n. 2, pp. 299-316.

[2] Cawley, P., Adams, R. D. (1989), "Sensitivity of the Coin-Tap Method of Non-Destructive Testing", *Material Evaluation*, vol. 47, pp. 558-563.

[3] Mitushasho, K. et al., US Patent 5,048,320, method and apparatus for impact-type inspection of structures, September, 1991

[4] DOT/FAA/AR-01/96, A Methodology for the Assessment of the Capability of Inspection System for Detecting Subsurface Flaw in Aircraft, US Department of Transportation, 2002

[5] Khan, M., Yanishevsky, M., Fahr, A., et al. (2009) "Bolt Hole Eddy Current Testing Probability of Detection Part – I: Experimental Design and Data Analysis", *Proceedings of the 12th International Conference of Fracture.*

[6] Berens, A. and Hovey, P. W. (1981), Evaluation of NDE Reliability Characterization, USAF Report No. AFWAL-TR-81-4160.

[7] Department of Defense Handbook, MIL-HDBK-1823A (2009), Nondestructive Evaluation System Reliability Assessment.

[8] Berens, A., Hovey, A. (1982), "Characterization of NDE Reliability", *Review of Progress in Quantitative NDE*, Vol 1, Plenum Press, New York.

[9] Berens, A. (1989), "NDE Reliability Data Analysis", *Metals Handbook, Volume 17: Nondestructive Evaluation and Quality Control*, ASM International pp 689–701.

[10] Spencer, F. W. (1998), "Identifying sources of variation for reliability analysis of field inspections", in *RTO Meetings Proceedings 10: Airframe Inspection Reliability under Field/Depot Conditions*, RTO-MP-10, AC/323(AVT) TP/2, pp. 11-1 – 11-8.

[11] Petrin, C. Annis and S. Vukelich (1993), "A Recommended Methodology for Quantifying NDE-NDI based on Aircraft Engine Experience", AGARD-LS-190.