SKI Report 01:48

Research

Effects of Time Pressure and Noise on Non-Destructive Testing

Johan Enkvist
Anne Edland
Ola Svenson

December 2001

ISSN 1104-1374
ISRN SKI-R-01/48-SE

The SKI Perspective

Background

The SKI regulation of mechanical components demands qualification tests of the NDT systems to be used for inservice inspections. A significant part of the qualification concept is requirement on qualification of NDT personnel, as it is commonly understood that the human factors effect the total reliability of an NDT system. Development of valid and reliable tests for personnel qualifications takes commitment and effort and also connection to the human factor research.

SKI goal

The overall goal of the SKI research within the area of human factors is to obtain the fundamental understanding of factors, affecting human performance while NDT, and to further develop valid, reliable and economically effective personnel qualifications. The more detailed objective of that particular research was to study effects of stress conditions on manual ultrasonic testing.

Results

The results of the study suggest that a little stress pressure can increase performance in a familiar task. However, it is a delicate balance to avoid the negative consequences of stress. The study shows that the NDT operators that had started working under stress conditions established simplified decision patterns, which they continued to use under non-stress conditions. Thus, it is strongly recommended to avoid training in stress conditions - it may result in simplified decision patterns being used when they are not called for.

Further research

Further research is required to determine the long-term effect of operators simplifying the task.

Effect on SKI activity

The study has produced a new knowledge of NDT operator performance in stress conditions. That knowledge will be taken into consideration for further development of NDT personnel qualifications.

Project information

SKI project manager: Elena Österberg
Project number: 97179

SKI Report 01:48

Research

Effects of Time Pressure and Noise on Non-Destructive Testing

Johan Enkvist[1]
Anne Edland[2]
Ola Svenson[1]

[1]Department of Psychology
Stockholm University
SE-106 91 Stockholm
Sweden

[2]Swedish Nuclear Power Inspectorate
SE-106 58 Stockholm
Sweden

December 2001

SKI Project Number 97179

Table of contents

Summary

Manual ultrasonic testing (UT) is the most frequently used non-destructive testing (NDT) method for in-service inspection of components important to safety and/or plant availability. Earlier, great variations have been found in operator performance, often attributed to operator fatigue. However, no conclusive findings have been reported. According to the Yerkes-Dodson law there is an optimal arousal level where performance is highest, for simple tasks this optimum is higher than for more complex tasks. In the present study twenty operators performed manual ultrasonic inspections of six test pieces with manufactured flaws. The operators performed the inspections under stress (high arousal - time pressure and noise) and non-stress conditions; one condition the first day and the other the second and last day. It was hypothesised that the stress condition led to a level of arousal so high that it would affect the results negatively. The results confirmed that the operators were affected by the stress condition. However, contrary to the hypotheses it was found that the manipulation increased operator performance. Operators with the stress condition the first day performed better than the other operators did (under both the stress and the non-stress condition). This was interpreted as the "stress first" (group 1) operators had established efficient performance patterns the first day – affecting also the second day. Operators beginning with stress condition also tended to be more motivated. It was concluded that operator performance is affected by arousal. The operators with non-stress first (group 2) worked hard with the complex task but their arousal level was assumed to be above the optimal, resulting in a low hit rate.

Sammanfattning (summary in Swedish)

Manuell ultraljudsprovning (UT) är den mest använda metoden för oförstörande provning (OFP) av komponenter som är kritiska för säkerheten och/eller tillgängligheten för kärnkraftverken. Stora variationer i provares förmåga har visats i studier, ofta har trötthet nämnts som delorsak. Inga konkreta resultat har dock presenterats. Enligt Yerkes-Dodsons lag finns det en optimal vakenhetsgrad där prestationen är som högst. För enkla uppgifter ligger detta optimum högre än för mer komplexa uppgifter. I föreliggande studie genomförde tjugo provare manuell ultraljudsprovning av sex testblock med tillverkade fel. Provningarna utfördes under två arbetsdagar varav en där provarna utsattes för stress (hög vakenhetsgrad - tidspress och buller). Hypotesen var att provarnas prestation skulle vara lägre under stress än icke-stress. Resultaten visar att provarna påverkades av stressbetingelsen men tvärtemot hypotesen resulterade det i ökad prestation. Provare som hade stressbetingelse den första dagen presterade bättre än de andra provarna, både för provning under stress och provning under icke-stress. Detta tolkas som att stressbetingelsen ökade provarnas vakenhetsgrad så att de presterade närmare toppen av sin förmåga medan icke-stressbetingelsen inte lyckades med det (c.f., Yerkes & Dodsons lag). En potentiellt stressande situation kan hanteras antingen genom att 'bita ihop' eller genom att minska på kraven.Rapportens slutsats är att prestation påverkas av vakenhetsgraden Provarna som började med icke-stress (Grupp 2) arbetade hårt med provningen men vakenhetsgraden var över den optimala vilket resulterade i lägre resultat ('hit rate'). Grupp 1 däremot hanterade stressen genom att minska på kraven (förenkla uppgiften) vilket resulterade i att de kom närmare prestationsoptimum.

1 Introduction

1.1 General background

During an annual outage in a nuclear power plant (NPP) components are typically inspected with non-destructive testing (NDT) techniques (e.g., ultrasonics, eddy current and radiography). In order for an inspection to be reliable the system as well as its parts need to be reliable (i.e., equipment, procedure and personnel). The reliability of an NDT-system is typically demonstrated in qualification tests, where first the equipment, then the procedure, and last the operators are tested (using qualified equipment, following a qualified procedure).

1.1.1 The problem of varying performance

Studies of the reliability of NDT (in performance demonstration tests) have found great variance in operator performance (see below). For example, Spencer and Schurman (1986) found the variation across operators to be “as large as it could be” when examining a study on the inspections of fatigue cracking in aircraft parts. The variation in performance remains after controlling for equipment, procedure and formal NDT-qualification of the operator, suggesting, the quality of an inspection to a high degree depending on the individual operator (e.g., Murgatroyd, 1994; Wheeler, Rankin, Spanner, Badalamente, & Taylor, 1986; Taylor, Spanner, Heasler, Doctor, & Deffenbaugh (1989); Skånberg (1991). For instance, Taylor et al., 1989 found a wide variation of operators’ performance, ranging from one operator with acceptable performance to the majority of operators with unacceptable performance.

1.1.2 Why such great variation in performance?

Manual ultrasonic testing (UT) is an NDT-technique that is often employed and frequently used where other techniques are too cumbersome to apply. Manual UT is the focus of some studies of varying performance, for the reasons above, it is the focus of the present study. The work situation for operators performing manual UT in a NPP is to a large extent characterised as less than optimal. There is continuous noise from ventilation and there is intermittent noise from other work going on nearby (e.g., speech, welding, cutting, and things being dropped). Action 7 of the international PISC III programme made a thorough study of the working environments effects on operator performance (Murgatroyd, 1994). The operators in the PISC-study were tested under calm and simulated working conditions where a number of working environmental factors were imposed on the operators (e.g., long shift’s, high temperatures and humidity, and noise). In the PISC III study it was found that tiredness and motivation affected performance but no clear indisputable findings were presented concerning the effects of any of the working environment conditions used in the study. Recently Worrall reported an additional study with the same character as the PISC III study; all operators were experienced but still they varied in flaw detection capabilities from day to day (even from morning to evening) (Worrall, 2000). (This variation was found to be a result of differences in technical skill (scanning patterns and maintaining coupling) but foremost through loss of concentration and vigilance due to tiredness. Worrall explains the second day’s reduction in performance for one operator to be caused by either “taking extra care on the first day” and then “reverted back to normal”, or, behaving “as normal the first day” and then becoming “bored or de-motivated”. (The operator did not use enough coupling resulting in a poor connection between transducer and material.)

The varying NDT-performance level is, probably due to different operators using different criteria for assessments/decisions according to Skånberg (1991). That is, much of the problem lies in the assessment of the indications, not in their detection and perception of them (though it is important to find them first). Also, Skånberg (1994) states that the operators’ ability to perform, to a high degree is influenced by the physical and mental environment, and how the operator interprets them. Karimi (1988) explains that operator performance could best be understood as being a function of *contextual* and *motivational* factors. In the industry, the results of these variations are costly and time-consuming repairs. No matter if the varying performance results in misses or false alarms, rectifying a situation is always more difficult/expensive if made unplanned. Not least important, variation in performance (and unnecessary repairs) can have a negative effect on the credibility of NDT (Skånberg, 1994).

Manual ultrasonic testing (UT) is a very complex task, where the operators rely on their sensory, perceptual, cognitive, and motor skills. The operators interpret and assess the, sometimes ambiguous, information and make a decision concerning whether or not there is a crack in a material. The more complex the task, the more effort has to be invested into managing it. When a task is more complex, environmental and work situation factors become more important, and have a greater effect on performance.

1.2 Man in the working environment

Omissions and erroneous actions result from interactions between technical, organisational and human factors. From this perspective critical situations can be studied, in which the individual is exposed to risks and provoked to make erroneous actions and take unnecessary risks (Sundström-Frisk, 1990). Many human errors are the result of demands on the individual exceeding his or her temporary ability (although slips and lapses are more common). There are two types of erroneous actions, intended and non-intended. Violations, (i.e., purposely abandoning instructions (Reason (1990)) are the result of higher cognitive functions and will not be dealt with here. Non-intended erroneous actions may occur when consciousness and thought processes are not actively connected. Such actions can be the result of lack of concentration and/or miss-directed attention, in turn caused by internal or external disturbing factors.

Levi et al. have summed up the conditions/situations in the working environment that generally tend to promote stress reactions (Levi, Frankenhaeuser & Gardell, 1982). These conditions/situations are quantitative overload, qualitatively under-load, lack of influence and control over the work task, and lack of social support. Work by the assembly line is the archetype for a poor work task; a highly repetitious task, with qualitatively under-load, low influence over the work task, and low social support. However, there is work with other forms of monotony (e.g., surveillance work in the process industry) which may also lead to stress reactions. Demands on continuous attention, ability to react in a very short time, and the knowledge that one at any time can be required to act efficiently are examples of stress factors that are typical for that type of work (Sundström-Frisk, 1990). Rubenowitz (1990) presents five main factors for assessing the work situation. They are, interpreted control of ones own work, a positive management climate, stimulating work task, good relations with co-workers, and a reasonable work load.

The work situation for the NDT operators during NPP outages is not monotonous but there is much room for improvement. During outages at an NPP, much work is to be done in a short period of time (in order for the reactor to be up and running as soon as possible). Many NDT inspections also have to be performed in this time-window. That is, there is a real risk for a quantitative overload for the NDT operators (i.e., time pressure and high workload). Also, due to other personnel working close by, the operators do not have absolute control over their work place. This means that the operators can be restricted even further in time, they can be disturbed by the work of others, and the operators' access to the components can be limited. However, the local organisation of the NDT activities (e.g., responsibilities for planning and supervision) can affect the factors above. That is, the task of the NDT operators may or may not be stressful, depending on how well the task is organised.

1.2.1 Performance shaping factors

Performance shaping factors (PSF) are factors that in some way affect the performance of an individual. Some factors affect all individuals about equally. Fatigue, time pressure, and heat are examples of general factors. Other PSF's affect different individuals differently. They affect most people in most situations. How we perceive heights and smells are two examples of when we can come to different conclusions depending on who we are and the situation we are in. In a working environment situation PSF's appear in many forms, both in the physical working environment and in the organisation. PSF's can affect performance by reducing input (e.g., perception – noise), or by reducing the execution (e.g., vibration), in addition PSFs can annoy the individual, thus reducing the ability to process information. In NDT at a NPP there are typically many intermittent noises, though noise is unlikely to affect the perception in this case (auditory signals are seldom used) the noise might still pose an irritant for the operators. Also, complete access to the component is seldom achieved, thus execution is reduced by obstruction.

1.2.2 A model of performance

The attentional ability of a person is determined by his or her level of arousal. The arousal level also determines the amount of attentional resources that are available (Proctor & Zandt, 1994). This relation is known as the Yerkes-Dodson law. As early as in 1908, Yerkes and Dodson (1908) offered a model for how performance is affected by arousal (by e.g., time pressure, and noise). According to the Yerkes & Dodson model performance is at its peak when arousal is not too high, nor too low, see Figure 1.

The level of arousal depends on the physical activation at any given time, and varies from deep sleep (lowest) to high graded excitement (highest). In order for the human brain to function optimally it requires a steady flow of varied information through the sensory organs (Frankenhaeuser & Johansson, 1981). In the working environment the physical activation is affected by sensory stimuli (e.g., noise, light, and climate); social stimuli (e.g., co-operation, presence of others); and cognitive stimuli (e.g., from the task at hand). According to Zuckerman (1969), each individual has a level of arousal that is optimal for them. This means that in a situation where one individual is optimally aroused, another individual is under or over activated. Also, brief periods of time pressure are known to make people feel more energetic, while longer periods of time pressure leads to increased fatigue (Thayer, 1989). Maule, Hockey &

Bdzola (2000) found strong evidence for deadlines to produce changes in affective state, not just feelings of time pressure.

It makes sense that performance is poor at low arousal levels, it is less obvious that performance deteriorates at higher arousal levels. This effect is mainly caused by attention being focused, thereby restricting the information being attended to, and a decreased ability to discriminate between relevant and irrelevant cues (Easterbrook, 1959). More complex tasks typically have a greater number of features to be considered and co-ordinated, and as arousal increases attention is directed toward the task at hand (Proctor & Zandt, 1994). This explains why more complex tasks have lower performance than simpler tasks at low and high arousal levels. At low arousal levels the attention is enough for a reasonable performance of simpler tasks but not for more complex tasks. When arousal levels are high attention is too narrow for complex tasks but more appropriate for simpler tasks (Näätenen, 1973).

Arousal affects our cognitive behaviour, when arousal is low so is attention; when arousal increase so does attention and we start to filter out irrelevant information. As arousal gets even higher a focus is put on the most important information; finally this narrowing of attention results in lowered performance. For more complex tasks this occurs sooner, and for simpler tasks later. At first the increased arousal is beneficial – the individual gets more focused on the task. When arousal gets too high the individual can experience stress. Both psycho-physiological and cognitive functions (e.g., heart rate and decision making) are affected by stress (i.e., heightened arousal) (Bergström, 1990). There is an optimal level of activation (arousal) where both performance and health is at their highest. Stress is generally considered to be a condition of high activation that is bad for the health of the individual. The combination of high arousal and low control in the long run promotes ill-health (e.g., cardiovascular diseases). However, a situation with low arousal and low control can also be bad in the long run. With high arousal and a feeling of control the individual can experience the positive stress – eustress (Frankenhaeuser, 1986).

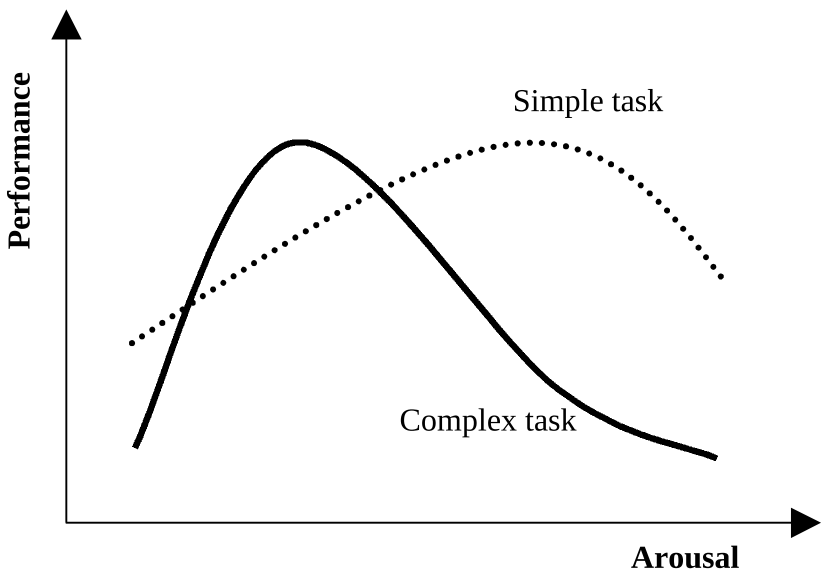

Figure 1: Graph over the relationship between arousal and performance according to Yerkes & Dodson (1908). For a complex task the optimal level of arousal is lower than for a simple task.

1.2.3 Time pressure and non-destructive testing

Since psychological processes take time to execute, time is evidently an important factor in decision making and judgement (Maule & Svenson, 1993); time pressure has been shown to reduce the quality of decision making (Payne, Bettman & Johnson, 1993). For a fuller description of the effects of time pressure on human judgment see Maule & Edland (1997). Time pressure occurs when a task is given a time limit to complete a task with less time than is usually required for that task. The feeling of being time pressured can occur when a person believes that there is not enough time to complete a task satisfactory. A shortage of time is assumed to induce affective changes through a generally increasing arousal level (Maule & Svenson, 1993). As time pressure increases a stress reaction can occur. Stress is generally defined as occurring when there is an imbalance between perceived demands and perceived resources (e.g., Chalmers, 1981; Cox & Mackay, 1981). Stress can be avoided by a more efficient use of available resources, for example by using a simpler decision procedure or focusing more resources to one single task. In 1960, Miller suggested seven strategies for how the human mind deals with information overload (i.e., too much information to process)(Miller, 1960). Three of these are valid in the NDT context, namely: acceleration, filtering, and omission. By accelerating the thought processes more information can be processed during the same time. By filtering the least important information is discarded. Omission is less selective than filtering and whole areas of information can be ignored. Whereas filtering and omission can go on forever, acceleration can only take place during shorter periods of time. In NDT the operators can deal with time pressure either by increasing the available resources (acceleration), or by reducing the demands on performance (filtering/omission). Increasing the available resources can work for shorter periods of time but will wear out the operators in the long run. Decreasing the demands on performance is a more likely strategy when working for prolonged periods of time. In either way, Kruglanski & Freund (1983) found that the effects of anchoring and primacy were larger under time pressure. That is, when under time pressure the decision is more than otherwise affected by information closer in time - primacy, and the final decision is more likely to be close to the initial (preliminary) decision – anchoring (Svenson, 1996).

In research on decision making, strong time pressure often leads to qualitatively poor decisions. One reason for this is the decision-makers basing their decision on only part of the information available. Another reason is decision-makers using different procedures when deciding under time pressure and when not under time pressure. It is reasonable to assume that these findings also hold true for NDT operators and their work situation. Some findings suggest positive effects from time pressure (e.g., Svenson & Benson, 1993a). Also, Freedman & Edwards (1988) reports increased satisfaction for tasks performed under time pressure. However, the task of manual UT is so complex that the effect of time pressure is believed to be negative.

1.2.4 Noise and NDT

The most often used definition of noise is unwanted sound. This means that noise is subjectively defined and can include even hardly noticeable and meaningful sounds such as speech and music (Kjellberg, 1997). Noise, at reasonable levels, up to about 70 dBA, has no direct influence on performance of non-auditory tasks. The indirect effect via disturbed concentration and increased arousal level can, however, have an effect, especially on such complex task as manual ultrasonic testing with high demands on concentration. There is no simple and non-ambiguous theory of how performance is affected by noise. Some findings point to no effects for noise below 95 dBA (e.g., Broadbent, 1971). These findings, however, are based on subjects performing simplistic laboratory tasks. Some researchers hold that the effects of noise on performance reasonably are larger than what currently have been shown. Individuals who notice a sudden noise display an ‘orientation reflex’, that is, non-wilful focusing on the noise, which could result in disturbed momentary attention. There is no generally acceptable noise level; under some conditions barely audible noise is enough to be disturbing. The effects of this disturbance might not be directly measurable in the individual’s performance but if the disturbance is prolonged findings points to negative effects on performance (Lundberg & Frankenhaeuser, 1978). By redirecting its resources the human mind can keep performance unaffected despite poor preconditions, for a limited time. Lundberg (1982) found two ways in which the subjects reacted to noise. Some subjects decreased their performance while keeping their adrenaline levels stable (non-stress). Other operators kept their performance stable while increasing their levels of adrenaline (stress). Findings point to stress factors that are hard to predict and difficult to control result in a greater stress reaction than when the individual experiences control over the situation (e.g., Kjellberg & Sköldström, 1996).

Noise can mask signals, that is, make critical audio signals undetectable (e.g., alarms). Masking can also reduce auditory feedback, a feedback that can be important for the performance of a task (e.g., checking the status of a component). Sudden noise can also make an individual focus attention on something else than the primary task for a very short time. In many tasks such short distractions in attention do not lead to negative consequences. In other, more complex, tasks any distraction may result in the process having to be aborted and restarted. Traditionally, noise has been dealt with within arousal theory. Noise has been considered to increase the level of arousal and thereby stop the drop in arousal that usually occurs after about 30 minutes. It should be noted, however, that there are reasons to believe that the effect may have an opposite direction, that is, decreasing the level of arousal if the noise is repetitive or continuous. Continuous noise, especially low frequency (e.g., ventilation fan) make people tired (Bohlin, 1971; Landström, Lundström & Byström, 1983).

To conclude, noise is an additional cognitive load in the performance of many tasks. Irrelevant speech or other sudden noises can disturb concentration. Continuous noise can cause fatigue and reduced attention. It is therefore reasonable to assume that noise can have a negative effect on manual ultrasonic inspection.

1.2.5 Non-destructive testing in nuclear power plants

In the present study, it is assumed that performance is affected by two factors, the NDT-system’s ability to perform, and the opportunity to perform. The first (ability, is typically controlled in qualifying the NDT-system. The second is mainly made up of the working conditions and working environmental factors (e.g., noise, time pressure, and accessibility). It is expected that by manipulating the working environment of the operators (imposing time pressure and noise) the operators’ performance would be affected negatively. As could be seen in the model of performance (figure 1), when at the optimal level, additional increase in arousal will result in decreased performance. For complex tasks the optimal level is lower than for less complex tasks. Manual UT is a highly complex task and the operators being aware of their performance being tested, the task itself would generate heightened arousal. The introduction of time pressure and noise in the present study will be referred to as the "stress" condition. Consequently the other condition (lack of time pressure and noise) will be referred to as "non-stress".

The nuclear industry (in Sweden) typically set higher demands on NDT than other industries, with the effect that operators work with that particular procedure only part of the time. Also, test blocks with implanted flaws has yet to produce exactly the same signals as "real" flaws do. Taking both these facts into consideration, it is reasonable to assume that operators feel a bit "rusty" the first day and more at ease the second. Thus, we have two factors that might affect performance and/or the subjective ratings. The present authors assume the first day effect to be negative (despite the findings in PISC III and Worrall studies above).

As mentioned above, the reliability of an NDT-system is controlled by qualification of equipment procedure and personnel. However, there are a number of working environment factors present when working in a NPP that are not present when performing a qualification. Examples of such factors are noise, heat, time pressure, poor preparations, bureaucracy, and not least radiation.

A high degree of vigilance is assumed in the task of manual NDT. The operator must be observant of indications of a possible crack. When operators work for full days (e.g., during outages) the demand of prolonged attention must be considered an additional load/constraint. There are findings showing that the longer a task has been performed (without rest) the more signals (per time unit) are missed (e.g., Davies & Tune, 1970). Grandjean (1988) sums up the most important findings on vigilance. In the field of NDT three of these findings are considered more important. First, performance under prolonged attention generally decreases after 30 minutes. Second, performance generally increases if the individual receives feedback on the performance. Third, performance decreases when the individual is exposed to adverse conditions (e.g., noise, heat, humidity etc.). Operators performing NDT under outages at NPP typically work long periods, without feedback, and under adverse conditions such as radiation, time pressure, heat and noise. That is, according to the above findings the risk is high for the operators committing erroneous actions (errors/misses).

For manual ultrasonic inspection the above concerns would be more important during the detection phase. During the characterising phase keeping vigilance should pose no problem. First, since the operator has found something and tries to characterise an indication, the operator's interest is awake. There are certain tasks to perform and all sufficient information is later assessed holistically, thereby the operator has better opportunity to notice any lack of information. Second, the operator changes between different probes, and looks for specific information. Vigilance poses a bigger problem in the first phase, detection. Detection is a supervision task, typically there are no indications. Since the operators do not become aware of any information missing, they will not go back and correct any mistakes.

As mentioned earlier, one way of coping with decision making under time pressure, is to choose a simpler, less time consuming process. In NDT, however, the operators have to follow a written procedure. That is, the procedure pattern for how to determine the integrity of a material is regulated. In the steps of the procedure there might be room for sub-procedures that the operators create themselves. Findings show (e.g., Wheeler et al., 1989; Behvaresh et al., 1989) that operators behave differently in different situations, even though they follow the same procedure. That is, operators might inspect in one way during qualification and in another way during testing in the field. Under qualification the operators know that their ability to detect and characterise cracks is tested, therefore the operators expect a large number of cracks present in the materials. When inspecting in real environments (e.g., NPP) cracks are rare. Operators being convinced that under qualification indications are more common and indications are more likely to be a crack than when testing in the field can explain the findings above. If an operator at the first glance believe an indication to be a crack, it is more likely that he will find the evidence to support his conception. The opposite also holds true; it is hard to find things you are convinced do not exist. Harris & McCloskey (1990) point out avoiding premature conclusions as one part of a successful inspection. Time pressure is definitely a working environment factor that the operators are exposed/subjected to, with a potential of negatively affecting the performance results.

1.3 The present study

In a critical evaluation of research on human factors in NDT, Pond, Donohoo & Harris (1998) recommend future research on NDT to be conducted in a controlled experimental environment. Pond et al. also recommend future research to be more concerned with performance, managerial policies and job design than with physiological responses and the effect of noise and heat. The present authors firmly believe that further knowledge can be reached by using methods and results from research in working environment and applying them to the particular field of NDT. The complex task of NDT operators requires proper experimental studies in order to reach results that can be generalised to operator performance in their proper working environment factors.

It was hypothesised that the stress condition affects performance negatively as an effect of sufficiently increased arousal (c.f., Figure 1). It was also hypothesised that there might be an effect of time. That is the operator being "rusty" the first day, and slightly improved their performance the second day. Furthermore, it was hypothesised that the above effects would be smaller for more experienced operators as a result of them being less affected by stress and being less affected by feeling "rusty" (well established tasks). Hence, the model for hypotheses is presented in Figure 2.

Hypotheses:

1. Stress condition results in higher stress/lower performance
2. Performance is increased the second day
3. Interaction effect, stress and second day – twice negative

	Stress	Non-stress
Day 1	— —	— +
Day 2	+ —	+ +

Figure 2: Matrix for the prediction of results according to theory. Performance is affected negatively by "stress" and by "Day 1" separately, making the upper left corner a double negative.

2 Method

The main goal of the present study is to look at the effects that working environment factors can have on the performance in manual ultrasonic inspection. This is done by exposing the participants (hereafter referred to as the operators) to a combination of stress factors associated with bad working conditions and stressful environments; namely noise and time pressure.

In the present study each of the operators performed manual ultrasonic inspection of three test pieces during each of two days. One day an operator worked under non-stress conditions, the other day he worked under time pressure and noise. The order in which the conditions appeared for the operators was balanced. In addition to the performance demonstration test (PDT) above, the operators also performed two ability tests (Aros and DIAT), and filled out a background questionnaire on the first day. After the PDT on each day the operators filled out a questionnaire concerning that particular day. The particular questionnaire concerned assessments of working conditions (e.g., noise and time pressure) and assessments of how important different signal characteristics were for detection and characterisation respectively.

2.1 Participants

Twenty-one professional operators, certified at Nordtest level 2 or 3, were hired to participate in the present study. (Eleven of them participated in an earlier study, Enkvist, Edland & Svenson, 2000.) The mean age of the operators in the present study was 40 years, the youngest 26 and the oldest 55 years. The operators also differed in years of NDT experience, from 5 to 37 years with an average of 19 years.

2.2 The task

The operators performed manual ultrasonic inspection according to the Swedish procedure (UT01) developed for manual ultrasonic inspection of stainless austenitic material (no welds). The operators inspected the heat-affected zone (HAZ) on both sides of the weld. The task was to detect circumferentially oriented indications (i.e., geometries and manufactured flaws), and at a later stage describe them according to the procedure (characterisation). During characterisation the operators mainly follow a flowchart in order to describe the indications as embedded or surface breaking (inner surface), volumetric or planar. The transducers used were 45^{o} probe for detecting, and WSY-70, WSY-60 and a normal (0^{o}) probe for characterising. No through wall sizing was required in the present study. The operators were told to work as they would have done, if they had worked in a nuclear power plant.

2.3 The conditions

In the present study the operators performed inspections in two conditions, non-stress and stress. In the non-stress conditions the operators had ample time to perform the inspections, they could take brakes at will and there were no loud noise. In the stress condition, however, the operators were subjected to both time pressure and noise (80-83 dBA).

2.3.1 Time pressure

Time pressure has been shown to be a powerful stress factor (Svenson & Maule, 1993). In order to make subjects time pressured one common method is to first measure the time a task normally takes and then decrease that mean with one standard deviation. Hence, using the times measured in a previous study without time pressure (Enkvist, Edland & Svenson, 2000), average times for the operators to complete the task were calculated for the test pieces. Time for detection was between 24 and 26 minutes, and the time for characterising was between six and eight minutes per indication, all depending on which block was inspected.

2.3.2 Noise

In the stress condition the operators were also subjected to loud noise. The noise consisted of a recording of ambient noise in a reactor containment of a Boiling Water Reactor during an outage. At the time of recording the sound level was between 80 and 83 dBA. In the present study the recording was presented via a tape recorder, measurements of the noise level near the head of the operators gave a sound level between 80 and 83 dBA.

2.4 Design

In the present study, both NDT related and psychology related material were used. First, the material necessary for performing the inspections will be presented.

2.4.1 Test pieces

The six test pieces consisted of a pipe (168 mm diameter by 7 mm), butt-welded to a bend (168 mm diameter by 11 mm), both in austenitic stainless steel. The bend bevelled down to 7 mm on the inside, closest to the weld. The test pieces contained both manufactured cracks (simulating IGSCC) and geometric indications. The flaws were implanted on the inner diameter of the HAZ, circumferentially orientated (skew +/- 20^{o}). The cracks had a through wall size larger than 2 mm, and were between 10 and 70 mm long.

The six test pieces used for the performance demonstration test (PDT) were chosen on the basis of the accumulated performance results generated under qualification tests. The test pieces were assessed to yield some variation but still with all operators being able to complete the inspections in the time allowed, without feeling too much time pressure (in the non-stress condition).

2.4.2 NDT equipment

The NDT equipment used in this study (instrument, transducers etc.) was the equipment that the operators ordinarily use. The equipment had to meet the requirements according to the procedure (UT-01).

2.4.3 Questionnaires

In the present study each operator completed two questionnaires, one concerning background information and one concerning the inspections made the same day. The latter questionnaire was completed at the end of each day. Also, immediately before inspecting the first test piece and between the second and third test piece each day, the operators made subjective ratings of their mood and working conditions at that moment.

The reason for putting the second self-ratings between the second and third test piece instead of after the third test piece was that this was considered rendering a better picture of the situation. Had the second ratings been given after the third test piece the ratings might have been affected by the operators feeling that the day's work was over. After the second set ratings the operators knew that they still had more work to do before leaving. Subtracting the former from the latter ratings results in the difference between the two ratings (c.f., reactivity). Reactivity is the difference between a base level (baseline) and a "condition" level. Ideally, the baseline is recorded during rest, and represents the individual's profile when not affected by the environment. In the present study, the first measure is taken just before starting the inspection. That is, when the operators knew and had mentally prepared for what was to come (i.e., time pressure and noise).

The background questionnaire contained questions regarding: age; number of years of NDT experience; training; certification etc. The questionnaires concerning the inspections were the same as in an earlier study by the present authors (Enkvist, Edland & Svenson, 2000). In this questionnaire the operators, among other things, rated the importance of certain signal characteristics for their inspections for that particular day. The eight signal characteristics were taken from the Harris & McCloskey (1990) study. The operators also rated their trust in the written procedure, their trust in their own gut feeling, and how well the inspections had gone on that particular day.

2.4.3.1 Subjective ratings

The subjective ratings questionnaires consisted of statements were the operators rated how much they agreed to each of a set of statements. Examples of statements used are -"I feel… *irritated*", the ratings were made using the Borg CR10 scale (Borg, 1982 & 1998) presented in Figure 4. The CR10 scale combines category and ratio scales, and was developed to compare ratings between individuals and over time. It is a general scale for measuring intensity of experience, but has mainly been used for ratings of pain and effort. The scale goes from 0 "Nothing at all", to 10 "Extremely strong", with the possibility for higher ratings in order to avoid truncation effects. In between, there are other numbers, and decimals, some of which are verbally anchored in verbs (e.g., "Very weak" and "Moderate"). The mood statements in the subjective ratings questionnaire were taken from Edland's (1994) stress index questionnaire. The items are: worried, tensed, discontent, irritated, time pressured, stressed, and insecure, which ratings are added, and: calm, relaxed, at peace, and content, which ratings are subtracted, resulting in the stress index. (See tables 6,7 & 8, statements with "+" are added and statements with"-" are subtracted). The stress index is not developed to be an absolute measure but to be used to compare individuals over time (repeated measurements). However, by incorporating the CR10 scale in the stress index comparisons between groups are believed to be justifiable. The rating behaviour of the operators was controlled by ratings of blackness (Borg & Borg, 1991).

2.4.4 Ability tests

The operators in this study performed two ability tests before starting the inspections. The ability tests were Aros (a paper and pencil test) and DIAT (a computer administered test). Aros measures spatial ability and DIAT measures a dynamic combination of aptitudes relevant for manual UT. However, all operators in the present study are seen as, at least, fairly good at manual ultrasonic inspection.

2.4.4.1 Aros

Aros is a test of spatial ability, designed to predict the ability to perform technical-mechanical work, among other things. The task is to compare two geometries (”pieces of cardboard”) presented on paper, and to match digits on sides of the two-dimensionally drawn (spread) with letters on the three-dimensionally drawn (folded). Aros was administered according to provided guidelines (See Glossary for reference).

2.4.4.1 DIAT

DIAT (Dynamic Inspection Aptitude Test) is designed to measure a dynamic combination of aptitudes, the principal aptitudes being general cognitive ability, abstract reasoning and spatial visualisation. DIAT was designed as a tool for early screening of personnel applying to be manual ultrasonic operators (Harris & Spanner, 1998). In the present study, however, it was administered to operators already active in manual UT. In this way a selection process had already taken place. Thus, the tested group is not the one intended for the DIAT. Apart from this, the DIAT was installed and administered according to provided guidelines.

0	**Nothing at all**	**”No I”**	
0,3			
0,5	**Extremely weak**	**Just noticeable**	
0,7			
1	**Very weak**		
1,5			
2	**Weak**	**Light**	**Easy**
2,5			
3	**Moderate**		
4			
5	**Strong**	**Heavy**	**Difficult**
6			
7	**Very strong**		
8			
9			
10	**Extremely strong**	**”Strongest I”**	
11			
≁			
●	**Absolute maximum**	**Highest possible**	

Figure 3: The Borg CR10 scale used for subjective ratings of feelings. Copied with kind permission. © Gunnar Borg, 1982, 1998.

2.5 Procedure

All inspections where performed in the same general environment, with the test pieces mounted on two identical rigs. In the present study the test pieces were mounted knee-high (0.4 m above ground). The two test-rigs were placed next to each other; half of the operators performed their inspections on one rig, and the other half on the other. The operators brought their equipment (transducers, gel etc) into their respective room. After detecting and characterising each test piece respectively, the operators wrote their reports sitting by a desk before starting on the next test piece.

The first day started at 10 o'clock a.m., this was done in order for the operators to travel to the test site the same day (the operators came from different parts of Sweden). After a short oral presentation of the study, the operators performed the DIAT ability test. After a break for lunch the operators where administered the AROS test of spatial ability. After the ability test were completed the operators performed a rating control test (Borg & Borg, 1991). At one o'clock the operators started the inspection of the three test pieces for that day. The operators were randomly assigned to one of two groups. Ten operators had the non-stress condition the first day and the stress condition the second day. For the other 11 operators the order was the reversed. The second day started at 8 o'clock with inspection straight away; there was a break for lunch when both operators had characterised all three test blocks.

2.5.1 Non-stress condition

During the non-stress condition day the operators were told that they had 45 minutes to complete the detection phase and 20 minutes per indication for the characterising phase for each of the three test pieces. The operators were also told that this was ample time, and that they would get more time if they should need it.

2.5.2 Stress condition

During the stress condition the operators were told that they had 24 to 26 minutes (depending on the test piece) to complete the detection phase and six or seven minutes per indication to complete the characterising phase for each of the three test pieces. They were also told that this time would be short and that they would have to work fast in order to complete the task on time. Similar instructions have been found to be more important for the adoption to time pressure, than the actual time for the task (Svenson & Benson, 1993b) (Maule & Maillet-Hausswirth, 1995). The operators were asked to please accept the challenge and do their best to complete the task in the time given. There was no time pressure on the operators when they wrote the inspection reports after each test block.

In addition to time pressure, the operators were also subjected to loud noise in the stress condition. The noise started right after the operators had completed the mood ratings and continued throughout the day until the last test piece was reported. The noise consisted of a tape recording of ambient noise in a reactor containment during an outage of a Swedish BWR (Forsmark III). The noise varied between 80 and 83 dBA, which corresponds to the level measured at the plant at the time of recording (c.f., Berglund & Lindberg, 1990). The operators were, due to ethical and practical reasons, not subjected to radiation even though operators themselves believe this to be one of the most significant performance shaping factors.

In summary, a total of 21 operators participated in the study, all male. One of the 21 operators was excluded due to missing data. All operators inspected six test pieces with manual ultrasonic inspections (UT), three test pieces on each of two days. The inspections were performed under two conditions, “stress” and “non-stress”, with one condition one day and the other condition the other day. Half of the operators started with the stress condition (Group 1). All inspections were performed individually, one test piece at a time.

3 Results

The outlay of the results will be as follows. First, some background variables and the results of the ability tests will be presented. Second, the operators’ subjective ratings of disturbances and feelings will be presented. Finally, the operators’ results on the performance demonstration test will be presented.

3.1 Background variables

Before inspection the first day, the operators performed two ability tests and completed a questionnaire of background variables (age and years of experience etc.). In Table 1 some background data on the operators are presented.

As can be seen in Table 1, there were no significant differences between the two groups in the above background variables. In the background questionnaire the operators were also asked questions concerning the testing procedure used; how they perceived the procedure; and how difficult they considered different phases in the inspection process to be. There were no differences between the two groups on any of these questions either. In Table 2 the results of the two ability tests, Aros and DIAT, are presented.

As can be seen in Table 2 the results on DIAT suggest the operators in Group 1 to be better suited for manual ultrasonic inspection. However, this difference is not significant. The correlation between the ability tests and performance (see below) were not statistically significant (Aros r_{xy}= 0,27 and DIAT r_{xy}=0,32). There were, however, significant correlation between age of operators and results on ability tests (Aros r_{xy} –0,74 p< .01; and DIAT r_{xy} –0,56 p< .05). This means that the older the operator in this study - the lower the score on the ability test. For further results on performance see below (section 3.4 Performance). In sum, the background data and ability tests suggest that the two groups did not differ significantly from each other prior to the study.

*Table 1: Means (**bold**) and standard deviations (in parenthesis) for age and experience of the operators. Presented for all operators and each of the two groups.*

	All	Group 1	Group 2	t-value	p
Age	**39.6** (8.9)	**38.8** (7.7)	**40.3** (10.2)	-0.37	n.s.
NDT-experience (years)	**17.5** (9.2)	**16.2** (7.6)	**18.7** (10.9)	-0.60	n.s.
NDT-experience in NPP (years)	**8.7** (6.7)	**7.4** (6.0)	**9.9** (7.4)	-0.83	n.s.
Weeks/year at NPP (average)	**5.2** (2.9)	**5.3** (2.7)	**5.0** (3.3)	0.22	n.s.

*Table 2: Means (**bold**) and standard deviations (in parenthesis) for results of ability tests AROS and DIAT. Presented for all operators and each of the two groups.*

Test	All	Group 1	Group 2	t-value	p
Aros	**32.7** (12.4)	**33.7** (11.1)	**31.6** (14.1)	0.37	n.s.
DIAT	**20.7** (9.0)	**22.9** (8.4)	**18.4** (9.4)	1.13	n.s.

In addition to measuring performance (see below), the present study also collected subjective ratings from the operators concerning their situation during the study. Below follows the results of the performance demonstration test and the subjective ratings on disturbances and stress related feelings.

3.2 Subjective ratings of disturbances

The operators performed subjective ratings in a questionnaire concerning their emotional state in the situation before beginning the inspection each of the two days. The operators also rated their emotional state after completing the second (of three) test block each day. This makes it possible to look for possible differences before each day as well as differences during each day. It is also possible to look at the operators' reactivity (difference between before and during) which can add to the understanding of the situation. Before the inspections started there were no significant differences within or between the groups, and there were no differences between the two groups during inspection. However, as expected, both groups rated noise more disturbing for the stress condition than for the non-stress condition. Group 1 had a mean of 3.05 (sd. 1.12) for the stress, and 0.83 (sd. 0.64) for the non-stress condition: t= 6.81, p< .01. Group 2 had a mean of 3.87 (sd. 1.96) for the stress, and 0.45 (0.38) for the non-stress condition: t= 5.31, p< .01. Also, with both groups collapsed noise was rated more disturbing for the stress condition (t= 7.47, p< .01). In Table 3 the operators' reactivity in ratings of disturbances are presented.

*Table 3: Means (**bold**) and standard deviations (sd) for the reactivity (difference between before and during) for operators subjective ratings concerning disturbances. Presented for both groups in both conditions.*

	Stress				Non-stress			
	Group 1		Group 2		Group 1		Group 2	
Disturbance	Mean	sd	Mean	sd	Mean	sd	Mean	sd
Noise	**1.98**	(1.45)	**3.39**	(2.07)	**0.20**	(0.35)	**0.02**	(0.40)
Heat	**-0.20**	(1.21)	**1.41**	(0.99)	**-0.05**	(0.37)	**1.11**	(1.59)
Lighting	**-0.13**	(0.74)	**0.35**	(0.88)	**-0.20**	(0.67)	**0.35**	(1.68)
Posture	**-0.10**	(0.84)	**0.70**	(0.98)	**-0.05**	(0.60)	**0.83**	(1.05)

As can be seen in Table 3, the operators in Group 1 had lower reactivity for noise during the stress condition than Group 2 had. This is in part due to Group 1's higher before-value. There was a significant increase in rated disturbance of noise under the stress condition, for Group 1 (t= 3.72, p< .01), Group 2 (t= 5.03, p< .01), and both groups collapsed (t= 5.85, p< .01). There were also differences in disturbance of heat between the two groups; Group 2 increased their ratings of heat more than Group 1, in the stress condition (F= 10.64, p< .01), and in the non-stress condition (F= 5.08, p< .05). Finally, the operators in Group 2 became overall more disturbed than the operators in Group 1 by their posture in the non-stress condition (F= 5.27, p< .05), also with both conditions merged (t= -2.30, p< .05).

In sum, the subjective ratings of disturbances suggest the manipulation was, on the whole, at least in part effective. The operators were more disturbed by noise when this was presented. Group 2 reported being more disturbed by heat and their posture even though this was not manipulated.

3.3 Subjective ratings of feelings

In an attempt to capture each operator's subjective feelings (e.g., of stress), the operators completed questionnaires concerning their feelings before and during each day. Each item in the questionnaire describes either a generally positive or generally negative feeling in the form of a statement. The operators rated how much they agreed with each statement. The ratings were used to calculate a stress index (Edland, 1994), the items in Table 4 marked with a "+" were added and the items marked "–" were subtracted making up the stress index score (marked *). In the tables below each item is presented as well as the stress index and operators' ratings of motivation. (Ratings were made in the Borg CR10-scale, Borg 1982 & 1998).

Before starting the inspections under the stress condition Group 1 had a higher stress index score than Group 2 had (-2.60 vs. –14.45; F= 6.98, p< .05). Before starting inspections under the non-stress condition, there was no significant difference between the groups (-7.57 vs. –12.46; F= 1.95 n.s.). Group 1 also tended to have significantly higher stress index before the stress than before the non-stress condition (–2.60 vs. -7.57 t= 2.19, p= .056); this was not true for group 2. The reactivity scores (difference between first and second measurement) are presented in Table 4.

*Table 4: Means (**bold**) and standard deviations (sd) for the reactivity (difference between before and during) for operators subjective ratings on situation. (Items marked with a +/- are added/subtracted, making up the stress index marked *)*

	Stress				Non-stress			
	Group 1		Group 2		Group 1		Group 2	
Feeling	Mean	sd	Mean	sd	Mean	sd	Mean	Sd
Worried+	**-0.47**	(2.18)	**0.36**	(0.99)	**0.30**	(0.63)	**0.12**	(1.88)
Calm-	**-1.00**	(1.87)	**-1.25**	(2.90)	**0.20**	(0.92)	**-1.35**	(1.83)
Tensed+	**-0.15**	(1.68)	**0.69**	(1.03)	**0.37**	(1.29)	**0.05**	(1.50)
Discontent+	**0.53**	(1.28)	**0.79**	(0.77)	**0.55**	(1.26)	**1.09**	(1.06)
Irritated+	**0.43**	(1.39)	**1.03**	(1.42)	**1.20**	(2.21)	**1.18**	(1.50)
Time pressured+	**2.03**	(1.86)	**1.94**	(2.89)	**-0.65**	(1.29)	**0.27**	(2.04)
Stressed+	**1.10**	(1.37)	**1.35**	(1.84)	**0.67**	(1.97)	**0.33**	(1.41)
Insecure+	**0.33**	(0.88)	**1.10**	(1.19)	**1.30**	(1.89)	**0.62**	(0.97)
Relaxed-	**0.35**	(1.06)	**-1.00**	(2.21)	**0.10**	(1.29)	**-0.65**	(1.29)
Peaceful-	**0.00**	(1.63)	**-1.00**	(2.40)	**-0.20**	(1.23)	**-1.00**	(0.78)
Content-	**0.70**	(2.00)	**0.40**	(0.84)	**0.00**	(0.82)	**-0.20**	(2.16)
Motivated	**-0.90**	(1.29)	**-0.60**	(0.97)	**0.50**	(0.85)	**-0.50**	(1.35)
Stress index*	**3.75**	(11.48)	**10.11**	(11.24)	**3.64**	(9.69)	**6.86**	(9.77)

As can be seen in Table 4, the operators increased their stress index score during both conditions, Group 2 more than Group 1. However, there were no significant differences in stress index reactivity between the two groups (F= 1.57 n.s. for the stress condition; and F= 0.547 n.s. for the no-stress condition). All operators were less calm during the stress condition then before the condition. However, the operators in Group 1 remained calm during the non-stress condition. The operators in Group 1 were also more calm than the operators in Group 2 during the non-stress condition (F= 5.75, p< .05). Also, Group 1 tended to be more motivated than Group 2 under the non-stress condition (F= 3.91, p= .063).

In sum, the subjective ratings of the situation under inspection suggest all operators being stressed by the task (manual UT) and more so for the stress condition.

3.4 Performance

The operators inspected three test blocks during each of two days with different conditions (stress and non-stress). There were 13 cracks in the test blocks, Table 5 presents the hit rates in percent for each cell in the matrix (c.f., Figure 2).

*Table 5: Hit rates (%) for each of the cells in the matrix, and for day 1 and day 2 as well as for the stress and non-stress conditions respectively. Means in **bold** and standard deviation in parenthesis.*

	Stress	Non-stress	
Group 1	**92.4** (8.1)	**85.0** (12.4)	**88.5** (8.2)
Group 2	**81.4** (9.5)	**77.6** (14.0)	**79.4** (10.3)
	86.9 (10.3)	**81.3** (13.4)	

As can be seen in Table 5, in the stress condition Group 1 performed better than Group 2 (F= 7.70, p< .05). Also, Group 1 had a higher overall hit rate than Group 2 (88.5 versus 79.4%) (F= 4.79, p< .05). However, there were no significant differences in performance between the two groups for the non-stress condition (F= 1.56, n.s.); nor for either of the groups between the stress and the non-stress conditions (Group 1: t= 1.87, n.s.) (Group 2: t= 0.99, n.s.), or. In order to find an effect of the manipulation the results of both groups for each condition was merged. The combined average hit rate for operators in the stress condition was 86.9% (sd 10.3) and for the non-stress condition it was 81.3% (sd 13.4). The difference is marginally significant (F= 4.21, p= .054). That is the stress condition itself tended to increase performance. In Table 6 the numbers of false calls in each cell are presented.

As can be seen in Table 6, most false calls were made in the non-stress condition by Group 2 (the first day). Table 6 is congruent with Table 5. Thus, looking at Tables 5 and 6, the results of the performance demonstration test are to some extent the opposite of what was hypothesised (c.f., Figure 2). The best performance (highest hit rate, fewest false calls) is found in the upper left corner; the lowest performance and most false calls are found in the lower right corner. In short, Tables 5 and 6 show the operators tended to perform better under the stress condition, and for both conditions the operators starting with stress (Group 1) performed better than the others.

Table 6: Number of false calls for each of the cells in the matrix.

	Stress	Non-stress
Group 1	0	1
Group 2	1	3

Summing up, there were no significant differences between the two groups in the above background variables. The results on the ability tests were correlated to age and not to subsequent results on performance demonstration test. Group 2 was more disturbed by heat and posture even though this was not manipulated. Operators tended to perform better under stress condition, and the operators with the stress condition the first day (Group 1) performed better than the others in both conditions.

4 Concluding remarks

In the present study operators performed manual UT inspections on three test blocks under each of two days with different conditions. Half the operators inspected under "stress" condition (time pressure and noise) the first day and "non-stress" condition the second day; for the other half of the operators the order was reversed. There were no 'a priori' differences between the two groups. Contrary to the hypotheses performance was higher under the stress condition. Also, the first day's condition had a transfer effect on the performance the second day.

Before starting inspection, the only difference between the two days was that the operators knew what condition they were going to work in that day. The operators who were going to start with the stress condition the first day were stressed already before inspection. This indicates these operators being more prepared than the other operators in the same condition (i.e., mentally prepared for a stressful event). The operators who were going to start with the stress condition primed them selves for a stressful event, in part by increased arousal but also by simplifying the task. The operators who had the stress condition the second day were less stressed due to their 'warm-up' the first day. They became more irritable due to their more complex (non-simplified) task depleting their resources, and they were more irritable (i.e., more disturbed by heat and posture) which results from them exerting more effort in solving their (non-simplified) task.

The hit rates in the performance demonstration tests suggest there be a main effect of the stress condition (high arousal) resulting in higher performance for all operators. Most false calls were made the first day by the operators who began with the non-stress condition. Thus, the results are to some extent the opposite of what was hypothesised. It was originally hypothesised that the stress condition would affect the performance (hit rate) negatively, possibly with a smaller effect when occurring the second day. There was no direct (order) effect of day. However, the condition (stress/non-stress) the first day seems to have a transfer effect on day two, resulting in the operators beginning with the stress condition having higher performance each of the two days. In a study of framing, Svenson & Benson (1993a) also found participants to perform better under time pressure.

The stress manipulation in the present study is believed to have caused the operators to reduce their demands (i.e., simplifying the task) in order to cope with the stressful situation. The operators that started with time pressure (Group 1) established patterns for making quick decisions, patterns that they used the second day as well. The operators that started with non-stress conditions established patterns that were more elaborate and therefore not as effective when there was little time. The second day, these latter operators had to change their patterns (way to work) due to time pressure. This resulted in them being less effective than the operators who used a simpler pattern from the beginning. Thereby, the operators starting with the stress condition performed higher than the other operators as a result of not having to deal with as much conflicting information the first day. The second day, the operators in Group 1 (stress condition the first day) used the same decision procedure as they had the first day, but with more time on their hands they, in some cases, sought out additional (conflicting) information, resulting in slightly lower performance. This lower performance was, however, not as low as for the operators who began with the non-stress condition. Studies of decision making under time pressure have shown this transfer of decision pattern (e.g., Keinan & Friedland, 1984 and Edland, 2001 manuscript). But in these studies the performance decreases under time pressure, and the ones who started under time pressure are worse off in both conditions.

The operators who began with the non- stress condition tried to make use of all information the first day (conflicting information as well) resulting in lower performance. The second day (stress condition), these operators started out using the same decision procedure as day one but were forced to adapt to the restricted time. The adapted decision procedure was more efficient and their performance subsequently rose. It should be pointed out that the test pieces used were chosen because they in the past have yielded varying results. That is, they are not easy to assess correctly. The difference in performance between the two groups for the stress condition was significant. The operators starting with the stress condition performed better, and were less disturbed by heat and posture. Also, the operators starting with the stress condition were more calm and motivated during the non-stress condition.

Although all operators were aroused by the stress condition, the operators who began with the non-stress condition were more irritable (disturbed by heat and posture). It is concluded that the reason for this difference is the unsuccessful coping of these operators. That is the operators who began with the non-stress condition became more frustrated as a result of them having a more complex (non-simplified) task. A tentative model of these relationships is presented in Figure 4 (c.f., Figure 1).

As mentioned, the findings that the performance was higher under the stress condition go against the hypotheses of the present study. However, they go well with the information volunteered by most operators, they reported to feel more challenged under the stress condition. That is, they felt the positive effects of heightened arousal (i.e., eustress (c.f., Frankenhaeuser, 1986)). A similar explanation is offered by Worrall (2000), where he suggests an operator’s decrease in performance resulting from “taking extra care” the first day and then “reverted back to normal”, or, behaving “as normal the first day” and then becoming bored or de-motivated. Within the PISC III action 7 programme (see Murgatroyd, 1994) human errors were attributed to the operators becoming tired and their technical ability thereby decreasing and that the operators becoming less observant of this occurring. With the operators feeling challenged (eustress) tiredness is unlikely to occur. However, it is plausible that the feeling of challenge is diluted over time, in the PISC III study the operators worked for eleven days compared to two days in the present study. Again, this points to the difficulty of this area. How performance is affected by a stress factor depends on the situation and the individual exposed. In some situations some individuals will perform poorly due to

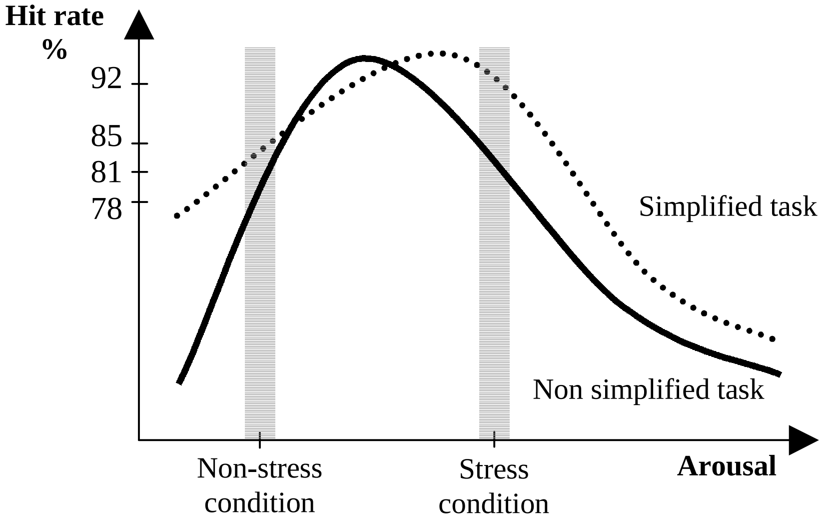

Figure 4: A tentative explanation for the performance results in the present study using the Yerkes & Dodson model. The shaded areas being the approximate mean arousal for the operators in the Non-stress and Stress condition respectively. The non-simplified task is the inspection task unaffected by coping (operators beginning with non-stress condition). The simplified task is the inspection

stress (distress) where as other individuals will perform well because they feel challenged (eustress).

The consequences of a decision also tend to affect the decision-making. With the present work description, over-reporting (false calls) is typically accompanied by negative consequences for the decision-maker (operator). It seems like small differences can make the operator assess an indication as a crack in one situation and not a crack in another situation (c.f., Wheeler et al. 1989). Thereby it is reasonable that working environment factors can have an effect on NDT performance, at least when the information is ambiguous. The amount of ambiguous indications thereby affects the reliability of NDT.

One problem with NDT in the nuclear industry is lack of training. Even experienced operators perform inspections according to this particular procedure (UT01) only a few weeks a year, resulting in them operators feeling “rusty”. Therefore the first day of the study in one sense served as training on the procedure for the operators. In accordance with the Keinan & Friedland (1984) and the Edland (manuscript) findings, the patterns used under training affect the patterns used later on (i.e., a transfer effect). However, whereas individuals who train under time pressure usually perform worse during both time pressure and non-time pressure conditions; the operators who “trained” under time pressure (stress) in the present study performed better under both conditions. The explanation for this contrast is believed to lie in the complexity of the task. In the Keinan & Friedland and Edland studies, the task benefited from considering all information before making the decision. In the present study, the results are interpreted to show that when the task is too complex the operator performs worse when trying to incorporate too much information. If the operator instead use fewer cues a decision can be made. An operator with enough resources to assess the most important cues comes to a better decision than an operator using more cues but who does not have the resources to make a proper assessment.

It is concluded that the operators under stress reduce the demands of the task by simplifying it. In the present study working under the stress condition increased performance contrary to other research findings. However, the effect that this simplified task has on the quality of NDT performance in the long run is not known. Also, the point where an operator shifts from trying to keep up to reducing the demands is individual and depends much on the work situation in general, the organisation, and probably on the personality of the individual operator.

The results presented in this study suggest that a little pressure can increase the performance in a familiar task. However, it is a delicate balance to avoid the negative consequences of stress. Since an area reported to be flawless is rarely double-checked it is important that the qualified procedures are followed during inspection. It is strongly recommended to avoid training in pressing conditions, it may result in simplified patterns being used when they are not called for. Further studies are required to determine the long-term effect of operators using a simplified pattern (simplifying the task).

The results presented here may also have implications for the qualification of personnel. The purpose of the qualification test is for the operator to show that he/she is qualified to perform NDT at a predetermined level of reliability. If operators display different decisional patterns for calm and pressing situations to a degree that affect the reliability, the design of the personnel qualification test might have to be reconsidered.

The findings suggest the effects of stress on performance to be modulated by motivation; that is, performance is less affected by stress when the individual is motivated. If this is true, it is of utmost importance that the organisational climate is good in order to keep the NDT operators motivated. It can be argued that the operators in the present study were fairly motivated since their ability was tested. On the other hand, the operators typically have a social network that can keep motivation high when working during outages, this network was not present during the present study.

Acknowledgements

This study was funded by grants from the Swedish Nuclear Power Inspectorate (SKI) to Anne Edland through Stockholm University. Many thanks to Arne Eriksson, Hans Lundberg, Lars Skånberg, Tommy Zetterwall, Elena Österberg, and of course to all the operators who participated in the study.

References

Behvaresh, M.M., Karimi, S.S., & Ford, M.E. Human Factors Affecting the Performance of Inspection Personnel in Nuclear Power Plants. Nuclear Plant Journal. January – February, 1989. p 16, 19, 23, 52, 54 & 63. 1989.

Berglund, U. & Lindberg, S. Human reliability of Non Destructive Testing. Unpublished SKI-Report. 1990.

Bergström, B. Theories of stress and performance. A short overview. FOA rapport A 50007-5.2. 1990.

Bohlin, G. Monotonous stimulation, sleep onset and habituation of the orienting reaction. Electro-encephalography and Clinical Neuropsychology. 31, 593-601. 1971.

Borg, G. A Category Scale with Ratio Properties for Intermodal and Interindividual Comparisons. In H.-G. Geissler and P. Petzold (Eds.), Psychophysical Judgment and the Process of Perception. Berlin: VEB Deutcher Verlag der Wissenshaften. 1982.

Borg, G. Borg's Perceived Exertion and Pain Scales. Human Kinetics:Chamaign, IL. USA. 1998.

Borg, G., & Borg, E. A general Psychophysical Scale of Blackness and its Possibilities as a test of Rating Behaviour. Reports from the Department of Psychology. Stockholm university, Sweden. 1991.

Broadbent, D.E. Decision and Stress. London Academic Press. 1971.

Chalmers, B.E. A selective review of stress: Some cognitive approaches taken a step further. Current Psychological Reviews, 1, 325-344. 1981.

Cox, T. & Mackay, C. A transactional approach to occupational stress. In E.N. Corlett & J. Richardson (Ed.s) Stress, Work Design & and Productivity. New York:Wiley. 1981.

Davies, D.R. & Tune, G.S. Human Vigilance Performance. Staples Press, London. 1970.

Easterbrook, E.A. The effect of emotion on cue utilization and the organization of behavior. Psychological Review *66*, 183-210. 1959.

Edland, A. Time Pressure and the Application of Decision Rules. Scandinavian Journal of Psychology, 35, pp 281-291. Scandinavian University Press. 1994.

Edland, A. Decision making at a catastrophe site: Effects of time pressure in training. (Unpublished manuscript). 2001

Enkvist, J., Edland, A., & Svenson, O. Operator Performance in Non-Destructive Testing: A Study of Operator Performance in a Performance Test. SKI-Report 00:26. 2000.

Frankenhaeuser, M. A psychobiological framework for research on human stress and coping. In M.H. Appley & R. Trumbull (Eds.) Dynamics of Stress. pp. 101-116.New York: Plenum Press, 1986.

Frankenhaeuser, M. & Johansson, G. On the psychophysiological consequences of understimulation and overstimulation. In L. Levi (Ed.) Society, stress and disease. Vol 4 Working life. London: Oxford University Press. 1981

Freedman, J.L. & Edwards, D.R. Time pressure, task performance and enjoyment. In J.E. McGrath (Ed.) The social Psychology of Time. London: Sage. 1988

Grandjean, E Fitting the task to the man. A textbook of occupational ergonomics. Taylor & Francis, London. 1988.

Harris, D.H & McCloskey, B.P. Cognitive correlates of ultrasonic inspection performance. EPRI Report NP-6675. 1990.

Harris, D. & Spanner, J. Jr. Dynamic Inspection Aptitude Test (DIAT) User's Guide. EPRI TR-106304. 1998.

Karimi, S.S., *Human Factors Affecting the Performance of Inspection Personnel in Nuclear Power Plants*. Electric Power Research Institute: Palo Alto, California. EPRI Report NP-6000. 1988.

Keinan, G. & Friedland, N. Dilemmas concerning the training of individuals for task performance under stress. Journal of Human Stress. 10, 185-190. 1984.

Kjellberg, A. Noise. In H.A. Waldron & C. Edling (Eds.) Occupational Health Practice. Butterworth Heinemann. 1997.

Kjellberg, A., Sköldström, B., Andersson, B., & Lindberg, L. Fatigue effects of noise among airplane mechanics. Work and Stress, 16, 62-71. 1996.

Kruglanski, A.W. & Freund, T. The freezing and unfreezing of lay inferences: Effects on impression primacy, ethnic stereotyping and numerical anchoring. Journal of Experimental Social Psychology, 19, 448-468. 1983.

Landström, U., Lundström, R., & Byström, M. Exposure to infrasound – perception and changes in wakefulness. Journal of Low Frequency Noise and Vibration. 2, 1-11. 1983

Levi, L., Frankenhaeuser, M., & Gardell, B. Work stress related to social structures and processes. In G. R. Elliot, & C. Eisdorfer (Eds.) Stress and Human Health. New York, Springer Publ. Co. Pp 130-147. 1982.

Lundberg, U. Psychophysiological aspects of performance and adjustment to stress. In H.W. Krohne &L. Laux. (Eds.) Achievment, stress, and anxiety. Hemisphere, New York. 1982.

Lundberg, U. & Frankenheauser, M. Psychophysiological ractions to noise as modified by personal control over noise intensity. Biological Psychology 6 51-59. 1978.

Maule, A.J. & Edland, A. The effects of time pressure on human judgement and decision making. In R. Ranyard, W.R. Crozier & O. Svenson (Eds.) Decision making: Cognitive models and explanations. 1996

Maule, A.J., Hockey, G.R. & Bdzola, L. Effects of time pressure on decision-making under uncertainty: changes in affective state and information processing strategy. *Acta Psychologica* **104** pp 283-301. 2000

Maule, A.J. & Hockey, R.J. State, Stress, and Time Pressure In O. Svenson & A. J. Maule (Eds.). Time Pressure and Stress in Human Judgment and Decision Making. pp. 83-101. Plenum Press: New York. 1993

Maule, A.J. & Maillet-Hausswirth, P. The mediating effects of subjective appraisal, cognitive control and changes in affect in determining the effects of time pressure on risk taking. Paper presented at the 15th Research Conference on Subjective Probability, Utility and Decision Making (SPUDM15), Jerusalem, August. 1995

Maule, A.J. & Svenson, O. State, Stress, and Time Pressure In O. Svenson & A. J. Maule (Eds.). Time Pressure and Stress in Human Judgment and Decision Making. pp. 323-329. Plenum Press: New York. 1993

Miller, J.G. Information input overload and psychopathology. American Journal of Psychiatry, 116, 695-704. 1960.

Murgatroyd, R. A. Human Reliability in Inspection, Final Report on Action 7 in the PISC Programme, PISC II Report 31. Nuclear Energy Agency. 1994.

Payne, J.W., Bettman, J.R., & Johnson, E.J. The adaptive decision maker. Cambridge: Cambridge University Press. 1993.

Pond, D.J., Donohoo, D.T., & Harris, Jr., R.V. An Evaluation of Human Factors Research for Ultrasonic Inservice Inspection. NUREG/CR-6605, PNNL-11797. 1998.

Proctor, R.W. & Zandt, T.van. Human Factors in Simple and Complex Systems. Boston: Allyn and Bacon. 1994.

Reason, J. *Human Error*. Cambridge University Press. 1990.

Rubenowitz, S. Arbetsorganisation och ledarskap. In L. Lennerlöf (Ed.) Människan i arbetslivet, beteendevetenskaplig arbetsmiljöforskning. Allmänna Förlaget, Stockholm Pp.140-157. 1990.

Skånberg, L., Bedömning av tillförlitlighet vid återkommande provning av huvudcirkulationskretsar. Report (Swedish Nuclear Inspectorate) SKI TR 91:28 October. 1991.

Skånberg, L., *Kvalificering av OFP-sytem.* SKI Rapport 94:25 [*The qualification of NDT systems*. Swedish Nuclear Inspectorate Report 94:25(In Swedish)] 1994.

Spencer, F. & Schurman, D., Reliability assessment at airplane inspection facilities, Vol III: results of an eddy current inspection reliability. Report No. DOT/FAA/CT-92/12, III. 1986.

Sundström-Frisk, C. Arbetsolycksfallen och den mänskliga faktorn. In L. Lennerlöf (Ed.) Människan i arbetslivet, beteendevetenskaplig arbetsmiljöforskning. Allmänna Förlaget, Stockholm Pp. 98-121. 1990.

Svenson, O. Decision making and the search for fundamental psychological regularities: What can be learned from a process perspective? Organizational Behavior and Human Decision Processes, 65, 252-267. 1996

Svenson, O., & Benson, L. Framing and time pressure in decision making. In O. Svenson & A.J. Maule (Eds.) Time Pressure and Stress in Human Judgment and Decision Making. pp 133-144. Plenum Press: New York. 1993a

Svenson, O., & Benson, L. On experimental instructions and the inducement of time pressure behavior.. In O. Svenson & A.J. Maule (Eds.) Time Pressure and Stress in Human Judgment and Decision Making. pp 133-144. Plenum Press: New York. 1993b

Svenson, O. & Maule, A.J. (Eds.) Time Pressure and Stress in Human Judgment and Decision Making. Plenum Press: New York. 1993

Taylor, T.T., Spanner, J.C., Heasler, P.G., Doctor, S.R. & Deffenbaugh, J.D., An Evaluation of Human Reliability in Ultrasonic In-Service Inspection for Intergranular Stress-Corrosion Cracks through Round-Robin Testing. Materials Evaluation, 47, 338 - 344. 1989.

Thayer, R.E. The biopsychology of mood and arousal. New York: Oxford University Press. 1989.

Wheeler, W.A., Rankin, W.L., Spanner, J.C., Badalamente, R.V., & Taylor, T.T. Human Factors Study Conducted in Conjunction with a Mini-Round Robin Assessment of Ultrasonic Technician Performance. Batelle Human Affairs Research Centers. 1986.

Worrall, G.M. A Study of the Influence of Human Factors on Manual Inspection Reliability and a Comparison of the Results with those from the PISC 3 Programme. In Proceedings of the 2nd International Conference on NDE in Relation to Structural Intergrity for Nuclear and Pressurized Components, May 24-26, 2000, New Orleans, Louisiana USA. 2000.

Yerkes, R.M. & Dodson, J.D. The relation of strength of stimulus to rapidity of habit formation. Journal *of Comparative Neurological Psychology,* **18**, 459-482. 1908.

Zuckerman, M. Theoretical Formulations. In J.P. Zubek (Ed.) Sensory deprivation. Fifteen years of research. New York: Appleton Century Crofts. Pp 407-432. 1969.

Appendix

Material used in the present study

Noise
A tape recording of the noise in the reactor containment (dry well) during an outage (Forsmark III, a Boiling Water Reactor (BWR) manufactured by ASEA Atom). The noise-level measured at the time of recording was between 80 and 83 dBA. During playback the noise-level was measured to be between 80 and 83 dBA also.

Tape recorder
used for playback was JVC RV-B55. The tape was played back at volume 49 (of 50) with highest loudness (BASS 2) without sound modification/equaliser (FLAT).

Tape
TDK Endless EC-6M

Rigging
The test blocks were mounted in a test rig with the possibility to impose different constraints. In the present study all test blocks were mounted knee-high with the weld approximately 500 mm from the floor.

Testing equipment
The equipment used was according to the procedure for manual ultrasonic inspection of stainless austenitic steel, UT01.

Sound-level meter
YFE YF-20. The meter was calibrated (against a Brüel & Kjaer, Type 2231) with the same equipment and noise source that was later used in the study.

Test blocks
168 by 7 mm pipe but-welded to a 168 by 11 mm bend, both in stainless austenitic steel and manufactured by Sonaspection UK

Acronyms and abbreviations

ABB – ASEA Brown Boweri (currently Westinghouse Atom AB)

BWR – Boiling Water Reactor, a nuclear reactor where the nuclear process boils water and the subsequent steam propels the generators.

dbA – Decibels Attenuation, one way of measuring noise where the frequencies that the human ear is adapted to are compensated for.

DIAT – Dynamic Inspection Aptitude Test, a test developed by ANACAPA Science and EPRI for the early screening of possible manual ultrasonic operators.

EPRI - Electric Power Research Institute

F-value – The result of a statistical test (ANOVA)

FAA - Federal Aviation Administration

HAZ – Heat Affected Zone, the part of the material closest to a weld that is affected by the heat from the welding process.

IGSCC – Inter Granular Stress Corrosion Cracking, a form of crack typically found in austenitic materials.

MTO – Man Technology Organisation

NDT – Non-destructive Testing, a way of examining materials' integrity without affecting it. Also called, Non-destructive Inspection (NDI) Non-destructive Examination (NDE).

n.s. – Not Significant, the result of a statistical test when the influence of other factors are too large.

PISC – Programme for Inspection of Steel Components, An international research programme aimed at exploring the capability and reliability of NDT

PSF – Performance Shaping Factors

Glossary

Aros – A test of spatial ability, measures the subject's ability to perform technical mechanical work [Utvikning: Art. nr.634-002, Psykologförlaget AB Box 47054, SE-100 74 Stockholm, Sweden]

Hit rate – The percentage of correctly reported cracks

Spatial ability – The ability to twist and turn objects mentally

Stress index – A subjective measure of stress, based on seven negative and four positive feelings (Edland, 1994)