INTEGRATED **HUMAN** FACTORS

BASELINE NC™

Advanced Fatigue Monitoring System

IHF Digital

www.ihf.co.uk

+44 131 201 2086

Contents

1.What is BaselineNC™?

“BaselineNC™ is a stakeholder project, which instigated the development of an Integrated Human Factors proprietary wearable and real-time fatigue monitoring device, to aid in the prevention of ensuing incidents or accidents within the workplace.”

The Integrated Human Factors (IHF) Baseline project was born out of a need to research, create, develop, and manufacture a viable unobtrusive product that could house a system whose functionality could accurately monitor, and alert increases in cognitive fatigue levels.

As a workplace Human Factor Consultancy group, we recognised the need to ensure individual fatigue thresholds are never exceeded and suitable safety margins and processes are put in place to prevent or minimise the introduction of human error, which is often attributed to impeded arousal levels.

The Baseline programme followed a system Verification and Validation (VnV) development and evaluation process. As a lead into that activity, a requirement capture procedure was conducted, which highlighted the need for the following key operational features:

1. The system's algorithm must accurately follow essential applicable parameters that correctly reflect an individual's fatigue bio-rhythm.

2. The product must be non-invasive - The device should collect data without irritating or breaking the subject's skin or invading the body.

3. The product can be wearable for ease of use and compact to be discreet - The wearable should be lightweight and small so that it can be used in the workplace without impacting the user's activities and movements.

4. The product must be affordable - The price of the device affects its adaptation at the workplace.

5. The device must have ease of use - The hardware should provide an easy-to- use interface with minimal user intervention.

6. The information presented to the user must be rapidly displayed and feedback must be unambiguous and quickly understood, to ensure good situational awareness.

7. The device must be durable and robust - The wearable should have a durable power source to ensure usability for at least one complete work shift to collect significant data. Its structure must be capable of withstanding the operational rigours brought about within the environment it is used.

From these key requirements, the IHF Baseline System was produced. The two tier system offers real-time monitoring, housed in an unobtrusive wearable fatigue device (WFD) worn around the wrist, with a small hub which can be attached to a belt clip or placed in the cab or workplace area. It was designed to be worn as a wristwatch, which is lightweight and comfortable and not aesthetically different to normal attire people would wear. This compatibility allows for extended wear without causing discomfort and can continuously monitor fatigue levels throughout the day, without interfering with day-to-day activities.

The system uses built-in sensors to monitor signals related to fatigue and employs a predefined algorithm to map bio-rhythms and identify fatigue threshold levels. These signals are wirelessly transmitted to a remote data logger, where they are processed into actionable information. Supervisors receive alerts in a traffic light system—Red (fatigued), Amber (approaching fatigue threshold), and Green (safe operation)—indicating when an individual needs to rest or be substituted from their role.

2.Fatigue is a Workplace Hazard

Fatigue is often defined as a decline in mental and/or physical performance caused by cognitive overload (prolonged physical and/or mental workload) or underload (boredom), physical exertion, sleep deprivation, circadian phase/circadian rhythm disruption, or illness [1]. Various definitions of fatigue have been proposed however the multidimensionality, interaction of different variables and in many cases, the subjective nature of perception of fatigue have not allowed for a unified definition..

Fatigue quantification is of crucial relevance in areas such as occupational health and transport safety. In part it is a physiological response of the human body preventing its overload. Fatigue impairs cognitive and/or motor performance, reducing situational awareness, work efficiency, productivity, and the introduction of human error, as well as increasing risks for injury and fatality [3]. Therefore, this physical and mental impeding attribute has the utmost importance for work safety in occupational settings such as, for instance, transportation, mining, aviation, or construction. As reported by police records, globally, 1–4% of the registered road crashes occur due to sleepiness and fatigue [4]. However, these values underrate the impact of fatigue on road safety, partially due to the inability to assess drivers' fatigue at the crash scene. Questionnaires, observational studies and in-depth investigations indicate that the actual value is around 10–20% [5]. That share rises to 20–50%, when considering only commercial vehicle accidents [6]. Long working hours, and consequent fatigue and stress, were found to increase the hazard rate among workers [7]. Furthermore, fatigue is involved in 4–8% of aviation mishaps [8].

3.The BaselineNC™ Inititial Operating Environment

Transport for London (TfL) officials were told of concerns about driver fatigue on the Croydon tram network two years before a fatal crash which killed seven people – believed to have been caused by a driver falling asleep. An independent audit of the Croydon Tramlink carried out in March 2014 identified “fatigue management” as one of the “seven weaknesses” of the transport system's safety management system. The report said that fatigue management measures in place “do not match current Office of Rail and Road (ORR) expectations in respect to the management of fatigue”. The independent auditors also recommended the development of a fatigue management system.

The Croydon Tramlink is operated by Tram Operations Ltd (TOL), a subsidiary of FTSE-250-listed First Group and provides drivers and managers for the network. In March 2014 , CIRAS – the body for confidentially reporting safety issues on the railways – received a complaint detailing concerns about fatigue arising from TOL's driver roster management. In 2014 an Abs Tracked Solutions audit was conducted, and several concerns were highlighted. In Abs Tracked Solutions' report, it stated that with regards to fatigue management, TOL “focuses on the management of hours worked rather than the management of fatigue”. The report called for the “development of a fatigue risk management system”.

On 9 November 2016, a tram derailed and overturned at a junction near the Sandilands stop in the worst rail accident in the UK in a decade. As a consequence, 62 people were injured, 19 seriously. An investigation later found that the driver had most likely had a “ microsleep” episode when approaching a bend.

Based on these outcomes, IHF were selected by UK tram to partner, design and develop a wearable device system that would pre-empt and flag fatigue related issues. This was conducted through a fatigue monitoring Driver Innovation Safety Challenge (DISC) project (see Figure 1), which involved City of Edinburgh Council, Transport for Edinburgh, Edinburgh Trams, and the Scotland Can Do Fund, working in partnership with UK Tram.

Objective Safety Scenario

The aim of integrating a BaselineNC™ device into a working environment is to highlight to both the operator and monitoring supervisor of any potential ensuing cognitive fatigue that could lead to human error and a catastrophic event, similar to that seen at Sandilands in 2016. Within Figure 1, a concept scenario is played out where a tram driver picks up their BaselineNC™ device at the depot, straps it on and engages it functionality. Both the driver and the monitoring supervisor (located in a central control room) will be informed of the system going live, as associated signals will be sent via a cellular or wi-fi connection from the BaselineNC™ device to the control room.

The supervisor will have a set display with all operational drivers wearing such a device – against each name there will be a set of three colours, Red, Amber, and Green, each highlighting the fatigue status of each driver. During each shift the supervisor will continually monitor the status display, where a signal passes from green to amber the supervisor will pay more attention to the fatigue condition of the associated driver. When the signal moves from amber to red, the supervisor will inform that driver to pull up to a stop at the next station and disembark to allow another driver to take over. Thus, managing fatigue levels, reducing human error, conceivable ill health of a driver and the potential of an ensuing catastrophic event.

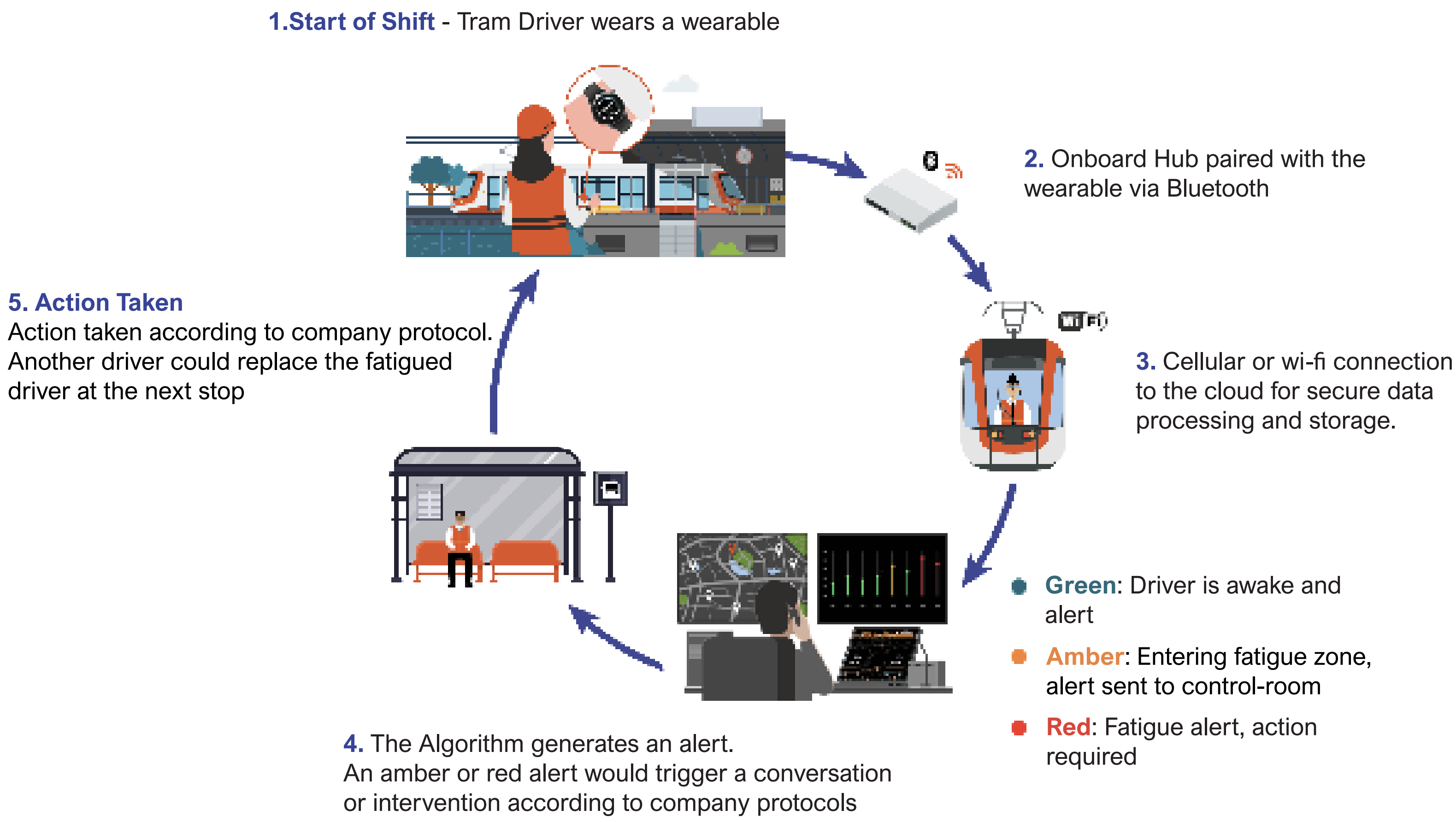

Figure 1. *Driver Innovation Safety Challenge (DISC) project assessment.*

Figure 2. *BaselineNC™ Functionality*

BaselineNC™ was designed and built based on research by:

A. Chowdhury, R. Shankaran, M. Kavakli and M. M. Haque - Sensor Applications and Physiological Features in Drivers' Drowsiness.

Malathi, D., Jayaseeli, J.D., Madhuri, S. and Senthilkumar, K., April, 2018. - Electrodermal activity based wearable device for drowsy drivers. In Journal of Physis: Conference Series.

Lu, K., Dahlman, A.S., Karlsson, J. and Candefjord, S., 2022. - Detection: A Review. Detecting driver fatigue using heart rate variability: A systematic review. Accident Analysis and Prevention.

4.BaselineNC™ as a wearable fatigue device

The previously mentioned facts show the important role of continuous monitoring of fatigue levels in an accurate and unobtrusive manner, through the detection and management associated with the onset of fatigue. Past methods of detection have been cumbersome and only useable within lab conditions. However, among recent and developing technologies, wearable sensors are starting to meet realistic portable requirements. Wearables enable continuous, long-term monitoring, paving the way for the development of accurate models for fatigue monitoring in real time. Therefore, their application for fatigue monitoring is seen as viable safety products and mitigation devices.

A wearable device is defined as a **small electronic mechanism** consisting of one or more sensors intended to be worn on or attached to a single body location where there can be a connection between sensors and a wireless remote data logger. As previously indicated the BaselineNC™ is a wrist- worn device, with resembling watch features, which has inbuilt sensors to monitor signals linked to fatigue factors. These wireless signals can be transmitted to a remote data logger and re- formatted into key information.

The **Bluetooth**, **Wi-Fi** enabled BaselineNC™ WFD, forms part of a personal health monitoringsystem, a concept introduced in the late 1990s, with the purpose of placing an individual in the centre of a censor delivery process, using remote monitoring by professionals as ameans of scrutinising and managing key biosignals that would indicate fatigue thresholds - allowing them to intervene and prevent human error- based events, induced by degraded cognitive and/or physical capabilities. The use of WFDs allows the portable acquisition of fatigue status monitoring over extended periods (days/weeks) within normal workplace environments.

Regarding healthcare and medical environments, in which fatigue monitoring falls under, it is expected to trend,worldwide, above

$15 Billion

According to Statista, in 2019, the wearable device market had a worldwide revenue of almost

$34 Billion

"Baseline was three years in the making, having evolved using a combination of data analytics, biometrics, and Human Factors - best practice approach to safety. We were proud to have been selected to develop this cutting- edge wearable technology as part of the Driver Innovation Safety Challenge project. The project involved City of Edinburgh Council, Transport for Edinburgh, Edinburgh Trams, and the Scotland Can Do Fund, working in partnership with UK Tram and IHF.

It represents a significant leap forward in proactive fatigue monitoring and workplace safety. I am excited that we have had the system endorsed by the Light Rail Safety & Standards Board and can now go on and fully commercialise the system. The ground-breaking solution has the potential to revolutionise fatigue detection in safety-critical roles across a wide variety of hazardous industries.

While initially tested by light-rail drivers, Baseline has cross- industry applicability and is suitable for anyone in safety-critical roles across various organisations and industrial sectors." Baseline TM is set to reshape the landscape of safety-critical roles, offering a proactive approach to fatigue prevention and management." - Neil Clark, IHF CEO.

5.Test & Evaluation

For the purposes of this outline document and the methodology of the Test and Evaluation (T&E) process that refers to, the issue of fatigue has been classed according to the type of load, which is considered as both physical (i.e., of physical causation, resulting from physical effort, leading to a decrease in physical performance – in this case, prolonged seated positions using a set of driving controls up to 8 hours per shift) and mental (i.e., of psychological attribute, resulting from sustained cognitive activity and leading to a reduction in cognitive and behavioural performance – in the case of this outline paper, long shift patterns, requiring intensive concentration while driving a tram) [2]

Assessment Fatigue Measuring Parameters

The BaselineNC™ WFD system was assessed using existing non-invasive methods, which are mainly based on five measuring principles: subjective measures, performance- related methods, bio-algorithm models, behavioural-based methods, and physiological signal-based methods of which not all are useful for online monitoring during work. However, for the BaselineNC™ WFD project, its VnV assessment process uses a cross-reference combination of Psychomotor Vigilance Testing (PVT), The Karolinska Sleepiness Scale (KSS), Behavioural-based observations and Situational Awareness as a human performance attribute indicator.

PVT

Where PVT is considered a performance-related method relying on the fact that participants' cognitive and consequently motor performance on specific tasks reflects their level of fatigue. Objectively assesses fatigue-related changes in alertness associated with sleep loss, extended wakefulness, circadian misalignment, and time on task. The difference between vigilance and attention is that attention is a fundamental cognitive process that is important to higher-level cognitive processes. Vigilance (the ability to sustain attention over prolonged time periods) requires attention and describes an individual's state of alertness, watchfulness, and preparedness to attend to critical information that is not yet present.

KSS

The KSS is a subjective measure and consists of assessing self-reported fatigue through questionnaires and scale measurements. It helps provide insights into mental and emotional processes underlying performance in a task. It accounts for subjective levels of sleepiness at a particular time during the day. Using a predetermined scale, subjects indicate which level best reflects the psycho- physical state experienced within a sequence of 10 - 15 min time frame.

An example of the scale can consist of the following:

KSS Description	Scale
Extremely Alert	1
Very Alert	2
Alert	3
Rather Alert	4
Neither alert nor sleepy	5
Some signs of sleepiness	6
Sleepy, but no effort to keep awake	7
Sleepy, but some effort to keep awake	8
Very sleepy, great effort to keep awake, fighting sleep	9
Extremely sleepy, can't keep awake	10

Table 1. *Karolinska Sleepiness with algorithm fatigue RAG status superimposed.*

The KSS is a measure of situational sleepiness. These subjective measures are therefore helpful as gold standards to compare with the results of fatigue models. As colour-coded in Table 1 and outlined in Figure 1, alerting BaselineNC™ algorithm Red/ Amber/ Green (RAG) fatigue status was aligned with KSS scores presented below -

- **Green** — **Alert** (A KSS score below 4.5 is considered as true alertness).
- **Amber** — **Transitional** (A KSS score of between 4.5 and 7.5 is considered as transitional (not truly alert and not truly fatigued).
- **Red** — **Truly fatigued** (A KSS score of greater than 7.5 is considered as truly fatigued).

This representative form of colour coding provides quick and accurate situational awareness for monitoring supervisors, of how fatigued a driver may be. The output from this correlation is discussed in the results section of this document.

Behavioural-based methods follow an observational approach to detect fatigue and include external signs, such as yawning, sighing, head nodding and eye closure – the latter in particular relating to microsleep. The term microsleep refers to very short periods of sleep that can be measured in seconds, rather than minutes or hours.

Even if you are not familiar with the words microsleep, you have likely experienced this phenomenon or witnessed someone else experiencing it. Classic recorded examples have been observed during simulated car driving assessments. A person might nod off during microsleep or keep their eyes open and continue to look awake i.e., starring into the distance without taking in their surroundings, hence impeded situational awareness (Situation awareness (SA) refers to a person's perception and understanding of their dynamic environment, and it is critical in making correct decisions and achieving performance levels. This awareness and comprehension are critical in making correct decisions that ultimately lead to correct actions) [11].

Regardless of how someone appears during a microsleep episode, their brain is not processing external information like usual. Microsleeps are most likely to occur after sleep deprivation. Because of this, many people with sleep disorders, such as shift work disorder or obstructive sleep apnoea, experience microsleeps. People who experience microsleep do not always recognise that they briefly fell asleep. Instead of recognising that you fell asleep, you might think you briefly stopped paying attention to whatever was around you. During a microsleep episode, people show a reduced response to external stimuli, such as sound or visual cues, they also display less accurate physical interactions. The primary risks of microsleep are accidents that could occur while driving or operating heavy machinery. These episodes are associated with a decrease in human performance. Since microsleep makes people briefly less responsive or unresponsive to stimuli, any high-stakes situation that requires quick reaction times becomes risky when a person microsleeps.

Mental fatigue has a detrimental effect on a wide range of prefrontal-cortex loaded cognitive functions, reduces the willingness to exert further effort, and is frequently accompanied with reduced performance efficiency and an increased number of errors [12]. Fatigue and its reduced task-specific motivation are often suggested to be accompanied by a decreased physiological arousal level [13].

A person's Level of Arousal can be described as a function of alertness, situational awareness, vigilance, level of distraction, stress and direction of attention. The function of a WFD can in effect, monitor how ready a person is to perform appropriate tasks in a timely and effective manner, including highlighting extreme under- arousal when it is manifested by unconsciousness, possibly caused by tiredness, associated with fatigue.

In the same respect, a WFD can monitor extreme over-arousal which can be manifest via a range of symptoms that will be peculiar to the individual, the environment, the task, and other factors. Such symptoms may include panic, aggression, submission, resignation, withdrawal, irrational behaviour mood swings, as well as unconsciousness.

Somewhere in between the two extremes, a point of optimum arousal, which is appropriate and effective for the situation, will exist allowing for optimum performance. At this point it seems that our mental capacity, situational awareness, alertness, attention, vigilance, decision-making, and actions are all heightened in sensitivity and execution – we are in the zone.

Our performance at work is also affected directly by the number, complexity and intensity of stressors present and the amount of subjective stress that we are experiencing. There is also a direct relationship between stress and arousal. It is therefore useful to connect levels of arousal with levels of stress. The Yerkes-Dodson Stress Curve is a simplification that can help to explain the relationship between arousal and performance.

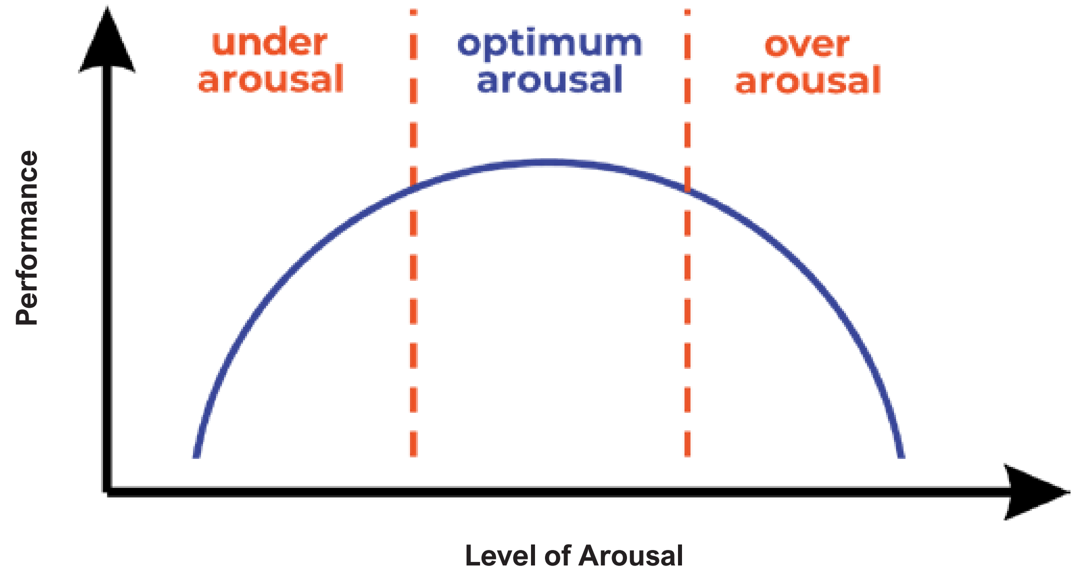

Figure 3. Yerkes-Dodson Stress Curve

At low levels of arousal, our performance is also low, which is fine when we're relaxing and not needing to perform tasks requiring skill or energy. However, if fatigue or lack of motivation causes our performance to fall below the required level, we're more likely to miss information, skip actions, be less vigilant, scan insufficiently, make mistakes, and react more slowly.

On the other hand, when arousal is too high, perhaps due to overload, we also make more errors, struggle with decision-making, and react more slowly. Over-arousal often leads to a narrowed focus, causing us to ignore or miss other important information. We may also fail to assess risk properly and prioritise tasks effectively, often focusing only on what we can easily achieve. In extreme cases, this can lead to a complete breakdown in functioning.

Factors Affecting Levels of Arousal

The factors that facilitate a person to perform at an optimal level are variable and peculiar to the individual, however, certain elements are often common, such as:

Workload
high (over stressed) or low
(under stressed – compliancy)

Task Complexity & Difficulty

Task Familiarity & Past Experience

Circadian rhythm
i.e previous rest/duty cycles
and time of day or night

Degree of Associated risk if failure occurs

Personal Motivation

The Working Environment

Levels of personal stress

Health

Interpersonal relationships at work
cooperation and coordination

Factors Affecting Levels of Arousal

- Decreased or limited cognitive situational awareness - inability to absorb attributes in the moment of project the potential introduction of future events.
- Increase in human error – through inaccurate or inadvertent activities

Human Performance

As part of the array of assessment parameters a Human Performance Situation Awareness Rating Technique (SART) was introduced as part of the observational behavioural-based approach.

The SART is a simplistic human performance non-evasive subjective rating technique that was developed for the assessment of Situational Awareness (SA). It uses the following **ten dimensions** to measure operator SA:

- familiarity of the situation
- focussing of attention
- information quantity
- information quality
- instability of the situation
- concentration of attention
- complexity of the situation
- variability of the situation
- arousal
- spare mental capacity

SART is administered post-trial (or non-intrusive stops e.g., tram stops) and involves participants subjectively rating each dimension on a seven-point rating scale (1 = Low, 7= High) based on their performance of the task under analysis. The ratings are then combined in order to calculate a measure of participant SA.

SART is a self-rating technique which elicits the subjective opinion on how aware a person was during task performance.

The reason that SART measures different components is that the SART developers feel that, similar to workload, SA is a complex construct; therefore, to measure SA in all its aspects, separate measurement dimensions are required. Because information processing and decision-making are inextricably bound with SA (because SA involves primarily cognitive rather than physical workload), SART has been tested in the context of skill, rule, and knowledge-based behaviour.

SART is a multidimensional scaling technique that consists of a series of questions that have bipolar responses. The number of dimensions varies between different SART forms. In its original form there are ten-dimension types (e.g., observation, questionnaire, interview, checklist, measurement instrument, etc.):

Domains	Contruct	Definitions
Attentional demand	Instability of situation	Likeliness of situation to change suddenly
	Variability of situation	Number of variables requiring attention
	Complexity of situation	Degree of complication of situation
Attentional supply	Arousal	Degree that one is ready for activity
	Spare Mental Capacity	Amount of mental available for variables
	Concentration	Degree of focus for each activity
	Division of attention	Amount of division of attention between tasks
Understanding	Infromation quantity	Amount of understood information presented
	Infromation quality	Degree of information accuracy
	Familiarity	Degree of aquaintance and experience

Table 2. *The SART ten dimensions (construct) to measure operator SA*

1. **Demands on attentional resources** (complexity, variability, and instability of the situation).

2. **Supply of attentional resources** (division of attention, arousal, concentration, and spare mental capacity).

3. **Understanding of the situation** (familiarity, information quantity, and information quality).

This was used in conjunction with the PVT, KSS and Behavioural-based observations to provide an overall objective and subjective set of results to be analysed.

Test and Evaluation Equipment and Drivers

To ensure the outcomes from the BaselineNC™ project met the expectations of both the internalIHF assessors and those from an independent resource (Ian Rowe & Associates Limited (IRAL)), it was essential that the operational environment and the associated tram drivers were fully representative in terms of equipment and training.

A number of stakeholder meetings were conducted, and it was agreed that the full-size replica simulator cab unit located at Edinburgh Trams' "Gogar Tram Depot" would be used as a staging ground for representative T&E. The simulator allows both trained and prospective drivers to experience almost five kilometres of new track and eight stops connecting Leith and Newhaven to the city centre, as well as the existing route from Edinburgh Airport to St Andrew Square, complete with lifelike imagery and landmarks along the route. The simulator helps drivers increase hazard awareness through inbuilt scenarios involving trouble spots and complex interactions with signals, traffic, pedestrians and other trams. The simulator is anthropometrically adjustable, with control layout accurately mirroring those within a real tram cab. All functionality and control inputs are representative, including warnings, cautions and advisories. The 180- degree visual field of view is as expected and allows the driver to look upon a high- fidelity external view, where real life scenarios can be played out and interacted with.

Figure 4. *Edinburgh Trams' Simulator at "Gogar Tram Depot"*

A small sample set of Suitably Qualified Equipped Personnel (SQEP) drivers was used during the initial IHF T&E phase. All were experienced drivers and simulator time slots were woven into their regular operational day, to ensure normal fatigue levels were present during all assessments and to ensure results were not biased or skewed.

IHF Test and Evaluation Procedure

The evaluation of the BaselineNC™ wearable device was non-gender specific and was conducted via a combination of verified and validated subjective SART and KSS feedback, plus objective PVT assessments, in conjunction with observational data gathering from a non-interacting Human Factor Subject Matter Expert observer during simulated scenarios. The outcomes were compared against BaselineNC™ recorded data and associated RAG status to see if indicated fatigue levels correlated.

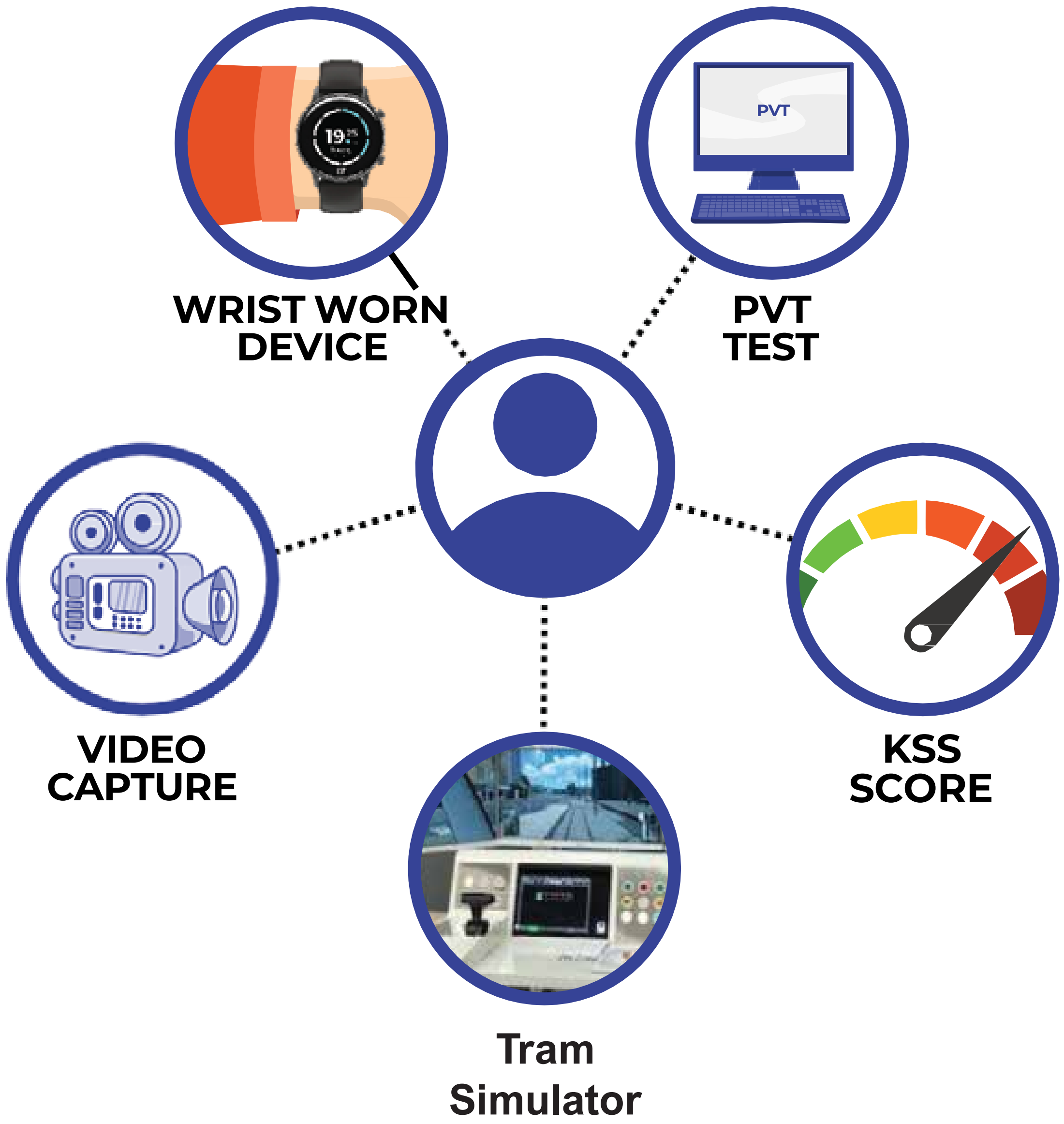

Figure 5. *Elements of the Test and Evaluation*

Each participating driver who took part in the simulator Baseline project initially went through a baseline evaluation. This provided a personalised calibrated model using self-assessed fatigue scores and informed Human Factors knowledge of shift patterns and fatigue criteria. The initial steps involved a training period (During an initial algorithm refining period, sensor data and self-assessed fatigue and well-being scores were collected), followed by baseline creation of physiological parameters (From this data, everyone's statistically usual range of physiological parameters was estimated) including physiological monitoring (Live sensor data from the BaselineNC™ wearable was compared to their baseline, allowing deviations* to be flagged in real-time). The combination of these factors forms Baseline Monitoring (Incoming data is monitored for long-term drifts away from the original baseline)

Note* - Deviations (statistically unusual states) can reflect a lack of well-being. Live flagging allowed (self-) assessment and intervention before it might otherwise be triggered.

Test Procedure Steps

Algorithm Calibration Phase:

1 - Using an IHF-developed algorithm, a personalised model was created and deployed for each driver.

2 - Objective data was gathered in parallel via a 10min PVT - to measure behavioural alertness via reaction speed to visual stimuli using a computer-based reaction assessment.

3 - The personalised model was refined during the calibration phase, using initial representative tram SIM scenarios - A standard route was used, with no additional operational scenarios, to provide a task with driver stimuli typical of a 'normal' eventless drive. Analysis of driver performance on the simulator was available but this was only used as an additional indicator if required.

4 - Self-assessment data was gathered in parallel via a KSS 1-10 scale at the end of each SIM scenario. For the purposes of improving self-assessment correlation, independent test work was requested by each participant. This was collected by asking each participant to wear the Baseline device for five evenings and provide KSS scores every hour for those evenings.

5 - During a normal 20min scenario, observational data was recorded, looking for elements of human performance degradation e.g.; decreased situational awareness, visual indications of micro sleeps and inept control interactions.

6 - Data was cross-referenced and compared, assessing correlation points, to ascertain algorithm accuracy to a goal level of 98%. The BaselineNC™ provides a Red/ Amber/ Green (RAG) output every one minute (i.e. continuously) whereas the KSS provides a subjective result (self-reported) at the sample time and PVT was used to provide an average 10-minute test.

7 - Once a suitable calibration level was achieved the alerting RAG fatigue status was aligned with the PVT data and KSS scores presented below.

- **Red** – Truly fatigued (A KSS score of greater than 7.5 is considered as truly fatigued).
- **Amber** – Transitional (A KSS score of between 4.5 and 7.5 is considered as transitional (not truly alert and not truly fatigued).
- **Green** – Alert (A KSS score below 4.5 is considered as true alertness).

8 - An independent verification was conducted using a similar calibration method with different participants and a tram simulator.

Full Assessment Phase:

1 - Participants were selected to conduct a set of SIM tram shifts as per their normal working schedule, which covered day, late and night periods.

2 - At the beginning of each scenario the participant was asked to complete a PVT assessment – those results were logged and compared against the KSS scores.

3 - Each SIM approx. 15 to 20min scenario was representative of their normal routes – these were repeated for a full shift pattern.

4 - During each run the participant was asked a set of SART questions at non-intrusive points of the journey. These were recorded by an observational Human Factors subject matter expert. Observations associated with microsleeps, human error and inept control interactions were also recorded and time stamped for correlation with other data.

5 - At the end of each run the participant was asked to complete a KSS assessment.

6.Summary of Results

Using a representative scenario and simulated operational environment (See Figure 1 and 4), the objective of the VnV process was to ascertain if the Baseline algorithm provided adequate fatigue level monitoring accuracy of individual participants. This was done by correlating accumulated measurements of subjective KSS and objective PVT recorded values - and matching them against the assigned WFD RAG status outcomes shown in Table 1. These were also correlated against subjective SART scores and observer findings, to match real-life issues associated with unwanted micro sleeps, low workload levels, low arousal levels, poor situational awareness and human error.

A succession of assessments to verify the accuracy of the KSS as a means of subjective measurement for this evaluation were conducted, by correlating KSS scores against PVT Reaction Times (RT). Where RT is defined as the time from stimulus onset to initiating adequate reaction. In particular, PVT lapse times were considered applicable (where a lapse during the psychomotor vigilance task, is usually defined as a response longer than 500ms) It is well recognised that increasing fatigue levels correspond to increases in RT responses. The PVT RT was assessed using a standard automated computer RT test, with randomly timed and sequenced stimuli, requiring the participant to engage a button when recognised.

The outcomes presented in Figure 6 indicate that PVT lapsed times or response times greater than 500ms does increase with higher KSS scoring and associated increases in fatigue levels. This agrees with the literature evidence found in references 14 and 15.

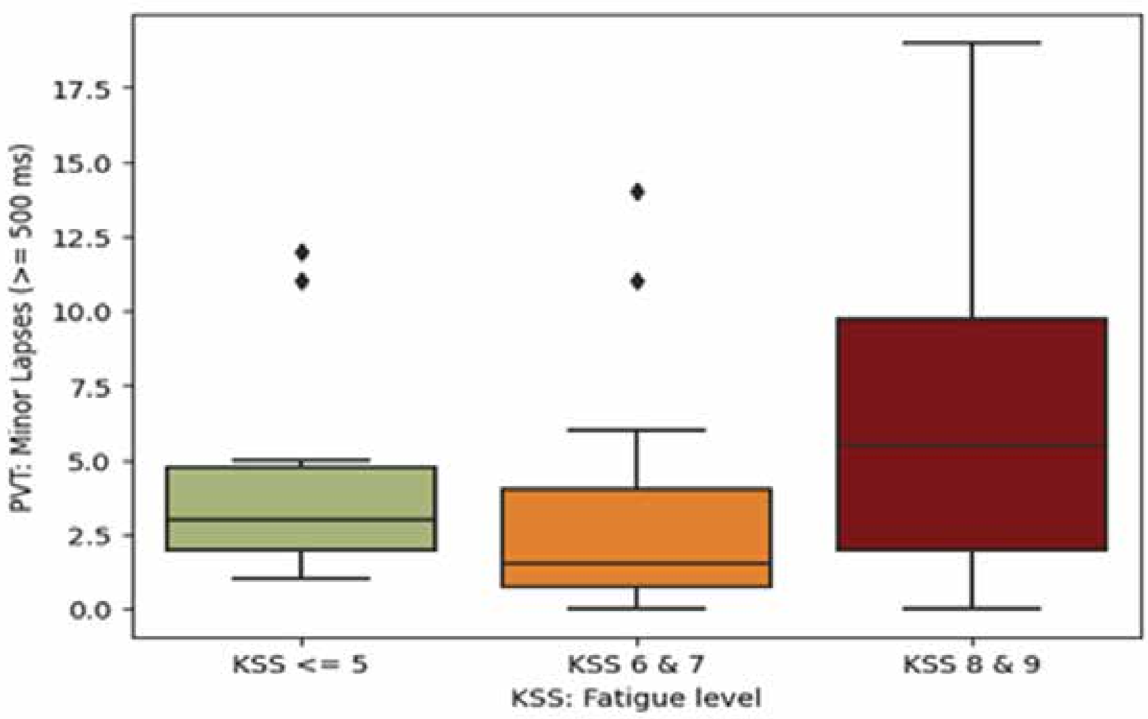

Figure 6: *Three key grouped KSS scores measured against PVT RT scores.*

Based on the evidence of this correlation, subjective KSS scores were used as a baseline marker for fatigue levels. These markers were then matched against the RAG status produced by the BaselineNC™ algorithm to assess the system's accuracy.

Karolinska Sleepiness Scale vs IHF BaselineNC™ output

The aim of the VnV process was to establish how accurate the WFD algorithm was in accounting for key fatigue parameters and presenting them in a RAG format that suitably alerted an onlooking supervisor as to a team member's fatigue level.

There are no standardise fatigue levels for humans; lethargy for individuals have their own discrete criteria and settings. These can vary from age, fitness, home life, shift work etc. However, what is common is a measure of the "accumulation of fatigue", especially in the workplace - and what we see in various cases, is that people will start a shift with more energy and better situational awareness. This can be, in part attributed to, people having had some type of rest period before a shift - also the environment and tasks are different to the previous period, which induces a form of additional arousal, even if those persons have done similar actions before. However, over time this arousal diminishes and is replaced by increased fatigue levels, both from a physical and cognitive perspective.

A key factor of the evaluation was to ascertain if the BaselineNC™ could follow an accurate accumulation of fatigue, tracing a lethargy timeline for each participant wearing the WFD. The expectation was to see a succession of green RAG indications over the first phases of a simulated shift, with the introductions of ambers as a participant's fatigue level increased to a threshold period, with the onset of red indications becoming more prominent at the latter phases of a shift.

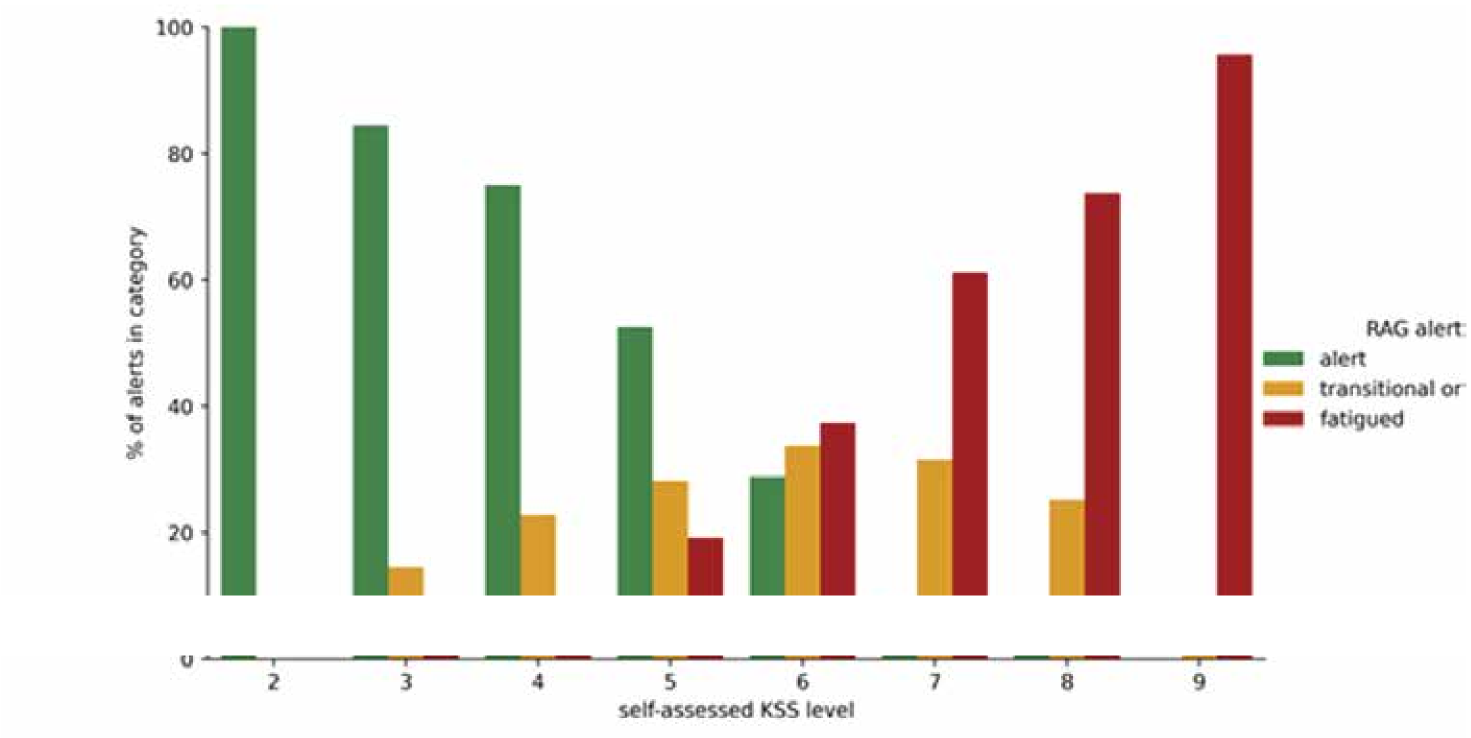

Figure 7: *KSS scores vs RAG status percentage for the Baseline.*

As can be seen from Figure 7, the results show there is a clear transition from dominantly green alerts associated with the KSS "subjective alert" category to dominantly red alerts in the KSS "subjective fatigued" category. These results were captured over a number of shift periods. The outcomes meet both the initial operational premise and the literature link between fatigue over time. Some overlap between is observed among fatigue states, particularly with amber alerts – this is expected due to the very individual nature of fatigue and because the KSS subjective fatigue evaluation captures only snapshots in time compared to the more continuous alerts of the RAG status.

Independent Assessment Results VS IHF Assessment Results

To ensure there was no bias in the results of the IHF concept testing of BaselineNC™; the company requested the services of a qualified independent assessor, which was also in line with the wishes of associated stakeholders.

A sequence of 58 independent evaluations, using high-fidelity tram simulators, at a separate site and participants who required training to ensure an acceptable skill set; was compared to a sequence of 21 evaluations conducted by IHF using similar simulators but with qualified tram drivers. Both sets of evaluations looked for misclassification of fatigue RAG status

From a IHF assessment perspective, the statistical results indicated that the self-identified KSS 8 and 9 scores (what can be considered trending true fatigue) only 0.4% of alerts were green, with 12 % of alerts were amber and 87.5% were red. While at the self-identified KSS 1 – 4 scores (trending true alertness): only 2.8% of alerts were red, 51% of alerts were amber and 47% were green. This demonstrated the BaselineNC™ had a very low rate of false negatives and misclassification.

The correlation of the independent results equated to a 98% positive outcome and appears to be relatively consistent with IHF assessment results which also had a positive outcome.

A combination of both IHF and Independent test results indicated the following for misclassification rates, where predicted RAG status for particular key KSS scoring was measured against actual outcomes and assigned a percentage of accuracy:

For KSS scores greater than 7 (True Fatigue Stages):

- The predicted RAG status was Green, but the actual outcome was Red = 0.86%
- The predicted RAG status was Amber, but the actual outcome was Red = 21.7%
- The predicted RAG status was Red, and the actual outcome was Red = 77.4%

For KSS scores less than 4 (True Alertness stages):

- The predicted RAG status was Red, but the actual outcome was Green = 1.38%
- The predicted RAG status was Amber, but the actual outcome was Green = 15.4%
- The predicted RAG status was Green, and the actual outcome was Green = 83.2%

Note 1: Ambers have been considered “uncertain or transitional” due to the variations of individuals, whose tolerance to off-set fatigue is different, based on stress levels and prelude events prior to the assessment.

Note 2: The key markers are when predicted RAG status equals actual status for True Fatigue and True Alertness stages.

From the combined results it can be clearly noticed that there is a significant success rate for fatigue indications when KSS scores are within the True Fatigue zone (outlined in Table 1). In contrast there also a significant success rate for non-fatigue indications when KSS scores are within the True Alertness zone (Outlined in Table 1). Both these results provide evidence of the devices high accuracy as a fatigue monitoring tool.

Overall statistical analysis indicates that the standard deviation between KSS scores reported at different alert levels and those relating to recorded RAG alerts, function as expected, with a clear separation between the Red and Green RAG states when required - see Table 3.

Rag Status	Mean KSS	Standard Deviation
Green	4.4	1.3
Amber	6.3	1.4
Red	7.3	1.1

Table 3. *Statistical Standard Deviation between RAG status and Mean KSS.*

Observational and Situation Awareness Rating Technique Outcomes

As an additional evaluating dimension, impartial observational and SART techniques were introduced during operational simulated test scenarios. These were considered unobstructive and non-biased in their application.

The aim of these techniques was to assess the human performance capability and to ascertain if it diminished over time – if so, did its decline correlate with the WFD algorithm RAG outputs. From an observational perspective, the observer recorded any physical or cognitive degeneration in terms of sloppy control inputs, micro sleeps or human error e.g. slowing down incorrectly while approaching a tram stop or engaging a control inadvertently at the wrong time. In particular micro sleeps were a key attribute to fatigue threshold and in the real world a prelude to a potential accident. All observations were recorded and matched against a combination of KSS scores and WFD RAG status, both over time.

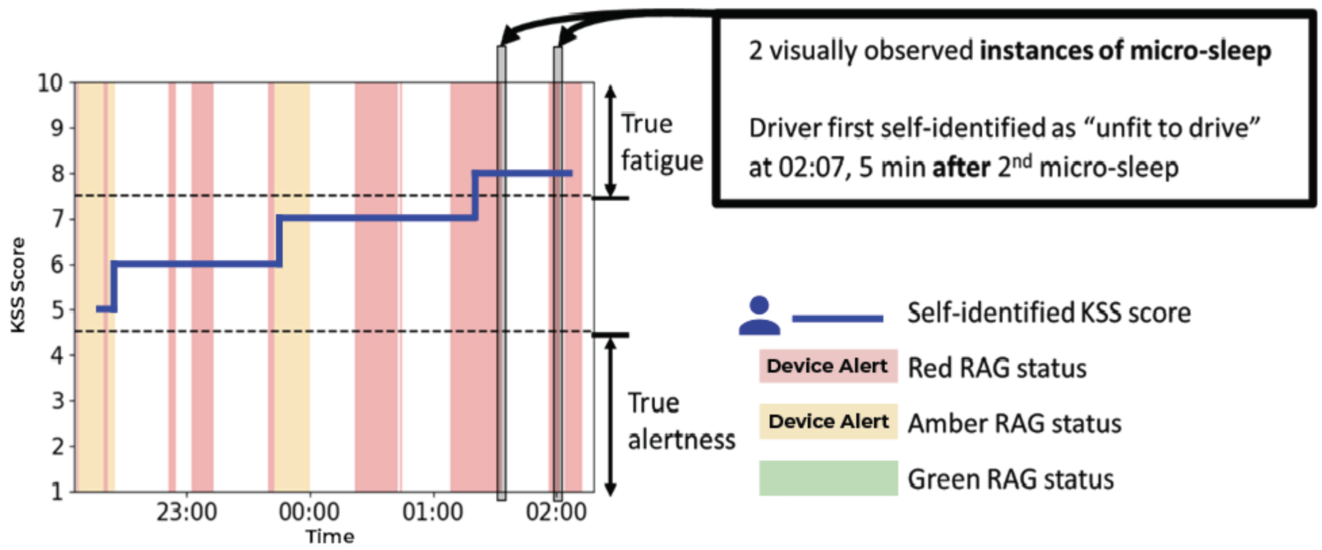

Figure 8: *A sporadic snap shoot of KSS scores, WFD RAG status over time.*

The example in Figure 8, recorded on a night shift scenario, provides a prime example of the lead-up to fatigue and how micro sleep can creep in. Recorded observational data indicates there were a number of physical manifestations prior to two microsleep episodes. These exhibitions ranged from incorrectly engaging the simulator ignition, a sequence of yawning, increasing eye blinking, a tram station overshoot and a number of over speeds, in comparison to previous runs. Recorded SART scores indicated a steady reduction in situational awareness from the start of the shift to the end, when the participant declared they were not fit do drive due to fatigue. Arousal levels were deemed low due to a lack of physical and cognitive demand, and repetitive actions became more wearisome, especially when approaching the early hours of the morning.

This observational evidence was matched against associated KSS scoring and WFD RAG indications over the period of that shift, where Figure 8 clearly identifies, in particular, two instances of microsleep corresponding with a WFD red status and a high KSS scoring of 8, which links to a true fatigue value. There were also a number of other microsleep episodes within this time frame – 01.47/ 02.04/ 02.05/ 02.06, all of which additionally correspond to a set of red RAG status and high KSS scores. This indicates a consistency in the algorithm's capability and accuracy to monitor fatigue levels.

Overview of Findings

The combination of the comprehensive IHF and independent assessment results show a significant correlation and accuracy (87.5% IHF testing, 98% independent testing) between real-life individual fatigue levels, including associated cognitive and physical characteristics, and those indicated by the Baseline algorithm RAG status.

These assessment connections provide operational evidence that the system delivers effective situational awareness monitoring of fatigue onset. It shows the device can be used as an effective tool to minimise potential catastrophic events associated with human error, linked to fatigue. Its applications are numerous within the workplace environment, with functionality having the capability to be adapted to meet the operational domains of various industries.

The BaselineNC™ system not only aligns with other commercially viable products, but rather surpasses them in terms of specificity, unobtrusiveness, and performance.

7.References

- Mohanavelu K., Lamshe R., Poonguzhali S., Adalarasu K., Jagannath M. (2017). Assessment of human fatigue during physical performance using physiological signals: a review. Biomed. Pharmacol. J. 10, 1887–1896. 10.13005/bpj/1308.

- Aryal A., Ghahramani A., Becerik-Gerber B. (2017). Monitoring fatigue in construction workers using physiological measurements. Autom. Constr. 82, 154–165. 10.1016/j.autcon.2017.03.003.

- Yung M. (2016). Fatigue at the Workplace: Measurement and Temporal Development. Thesis, UWSpace.

- Li Y., Yamamoto T., Zhang G. (2018). Understanding factors associated with misclassification of fatigue-related accidents in police record. J. Safety Res. 64, 155–162. 10.1016/j.jsr.2017.12.002.

- European Commission (2018). Fatigue. Available online at: https://ec.europa.eu/transport/road_safety/sites/default/files/pdf/ersosynthesis2018-fatigue.pdf (accessed March 30, 2021).

- Davidović J., Pešić D., Antić B. (2018). Professional drivers' fatigue as a problem of the modern era. Transp. Res. F Traffic Psychol. Behav. 55, 199–209. 10.1016/j.trf.2018.03.010.

- Dembe A. E. (2005). The impact of overtime and long work hours on occupational injuries and illnesses: new evidence from the United States. Occup. Environ. Med. 62, 588–597. 10.1136/oem.2004.016667.

- Caldwell J. A. (2005). Fatigue in aviation. Travel Med. Infect. Dis. 3, 85–96. 10.1016/j.tmaid.2004.07.008.

- Statista, B.I. Wearable Device Sales Revenue Worldwide from 2016 to 2022 (in Billion U.S. Dollars); Statista Inc.: New York, NY, USA, 2017.

- Yussuff, V.; Sanderson, R. The World Market for Wireless Charging in Wearable Technology; IHS: Englewood, CO, USA, 2014.

- Endsley MR. Situation awareness global assessment technique (SAGAT). Proceedings of the National Aerospace and Electronics Conference (NAECON). New York: IEEE, 1988: 789–95.

- Ackerman, P. L. (Ed.). (2011). Cognitive fatigue: Multidisciplinary perspectives on current research and future applications. American Psychological Association.

- Van der Linden, D. (2011). The urge to stop: The cognitive and biological nature of acute mental fatigue. In Cognitive fatigue: Multidisciplinary perspectives on current research and future applications (pp. 149–164). American Psychological Association.

- Anund, A., Kecklund, G., Vadeby, A., Hjälmdahl, M., & Åkerstedt, T. (2008). The alerting effect of hitting a rumble strip—A simulator study with sleepy drivers. Accident Analysis & Prevention, 40(6), 1970-1976.

- Ingre, M., Åkerstedt, T., Peters, B., Anund, A., Kecklund, G., & Pickles, A. (2006). Subjective sleepiness and accident risk avoiding the ecological fallacy. Journal of sleep research, 15(2), 142-148.