Development of an Attention Test using the VitalSky

Himanshu Shah

Technische Universiteit Delft

TU Delft

Development of an Attention Test using the VitalSky

by

Himanshu Shah

to obtain the degree of Master of Science

in the
Embedded and Networked Systems & Internet of Things Group

at the
Delft University of Technology

defended
publicly on Monday September 30, 2019 at 10:00 AM.

Student number:	4739779	
Project duration:	December 3, 2018 – September 30, 2019	
Thesis committee:	Dr. ir. E. M. van der Heide,	Senior Scientist, Philips Research
	Prof. dr. R. R. V. Prasad,	Associate Professor, TU Delft
	Prof. dr. G. Kortuem,	Full Professor, TU Delft
	Prof. dr. C. Lofi,	Assistant Professor, TU Delft

This thesis is confidential and cannot be made public until September 30, 2021.

TU Delft PHILIPS

Acknowledgment

I would like to take this chance to express my gratitude to all the people who I was involved with during my thesis and who kept motivating me through the good and bad phases. I would like to thank my daily supervisor Esther van der Heide, for patiently helping me not only academically but also with the administrative issues faced during my thesis, for always providing me with opportunities to showcase my work to the entire department and for motivating me to go the extra mile.

Furthermore, I would like to thank my supervisors from the university, Prof. Gerd Kortuem and Prof. R.R. Venkatesh Prasad, for continuously evaluating my work critically, for actively discussing my progress and making sure that I stay on the track.

I am grateful to Thomas Falck, for always assisting me with whatever I needed and making me feel a part of the team ever since I joined Philips. Many thanks to Stef Helling, for being available to help me with the software issues related to VitalSky, and Ton Akkermans for educating me about the technical details of the VitalSky.

I would specially like to thank Jacco Eerden and Henk van der Put for sparring time from their busy routine and letting me bounce my design related questions over them, and Mine Danisman-Tasar for providing vital insights about the user experience principles and designing visual contents. I am thankful to Raymond van Ee, for reviewing and guiding me with the psychological aspects of the attention tests used in this thesis.

This acknowledgement would not be complete without the mention of all my crazy friends Francesca, Irene, Esther, Carla, Margarida, Saharsh, Claudia, Sevgi, Sinduja, Atilla, Tim, Jose and all amazing interns I met during my thesis at Philips, thanks for making my thesis fun and memorable at Philips.

Thanks to Mohit, Shivanand, Nilay, Neel, Mihir, Sukanya, Arundhati, Chinmay, Satish, Dhaval, Saumil, Shubham, Priyanka and all the other friends at TU Delft for inspiring and supporting me as a friend and family during my time in Delft.

Finally and most importantly, to my family for the never-ending support and love, for believing in me and providing me with this opportunity and always guiding and inspiring me regardless of being near to me or far away.

Himanshu Shah
Delft, September 2019

Abstract

Delirium occurs in upto 80% of the ICU patients and is one of the most common source of morbidity and mortality in the ICU. Inattention is one of the core features and affects 97%-100% of the delirium patients. This indicates that attention can be a good parameter for detecting delirium. In literature, it has been found that the use of visual attention tests have been successful in detecting delirium among the ICU patients. The VitalSky is a luminous ceiling used for providing light therapy to the ICU patients. Since, the VitalSky is already available in the ICU and allows to render visual contents, it can be used for detecting delirium in the ICU patients by performing visual attention tests. In this thesis, visual attention tests are developed for the VitalSky. Moreover, different interaction devices are explored that can be used by the ICU patients to respond to the tasks of the attention test and help in automation of the test. The technical feasibility and usability of the automated setup is also evaluated in this thesis. The validation of the performance of the attention tests and user preference of the response devices is assessed by conducting a pilot study on 10 healthy volunteers and collecting their subjective feedback.

Contents

List of Figures

List of Tables

1 Introduction

Being admitted to a hospital is a dreadful experience. Apart from the wounds or illness that a person suffers, the environment and conditions in the hospital makes it more difficult for the person to relax and recover at full pace. The factors like continuous noise from the monitoring devices, conversations of the nearby patients, absence of natural lighting which affects the circadian rhythm of the patients have a negative influence on the well being of the patients. The conditions gets more intense in an Intensive care unit (ICU) where patients often suffer from cognitive disorders during or even after their stay in the ICU. A stay in ICU increases the risk of mortality by two times and a risk of permanent cognitive impairment by nine times [8, 9]. One of the most common cognitive disorder during the hospitalisation is delirium.

The term delirium refers to a severe neuropsychiatric disorder involving motor abnormalities and is commonly observed in acute medical patients, following surgery or trauma [10]. It is a common, serious, and potentially preventable source of morbidity and mortality for medically ill or aging patients [11]. It occurs in up to 80% of ventilated patients in the ICU and 60% of older hospitalized patients [12, 13]. It has been estimated that if the length-of-stay for each acutely confused, elderly hospitalised patient is reduced by just 1 day, it will result in a saving of $1 to $2 billion per year for the Medicare [14]. Therefore, an early detection and treatment of delirium is vital for improving the patient recovery and experience during and after their stay in hospital.

The Philips venture VitalMinds works towards delirium management in the ICU and has developed an innovative luminous ceiling named the VitalSky. It improves the environment of the ICU by creating a healing environment for the patients. It simulates natural lighting conditions and creates a calm and comforting environment. Next to light therapy, the VitalSky also permits to render different visual content. It renders content that helps the patient with the circadian rhythm by simulating day and night lighting conditions.

Timely diagnosis of delirium can not only reduce the mortality rate associated with delirium but also the length-of-stay of patients in the hospital. Even though the severity, frequency and consequences of delirium are known, it is still substantially under-detected in hospitals and clinical settings [15]. It has been found that nearly 75% of delirium patients go undiagnosed [16]. In a study [15], it was found that the nurses in non-ICU ward identified delirium in less than one out of three patients. These nurses showed very good negative discriminatory assessment but a very poor positive discriminatory assessment. So, what could be the reason that delirium is still substantially under-detected?

There are a number of factors that could be possible for the underdiagnosis of delirium. Firstly, most of the tests used to detect delirium in clinical settings are subjective i.e they rely on the experience of the tester [17]. So they are suited for use by experts but not for untrained physicians and thus training of the physicians and nurses becomes an important aspect in these tests. Secondly, some of the tests contains similar question and thus the patients get used to the test over time and can learn to answer the questions. Another factor although not a major factor is that the symptoms of delirium and other cognitive disorders like dementia overlap to a high degree [18] which makes the diagnosis of the delirium difficult.

In previous research, it has been found that there exists many symptoms that can be linked as a cause of the

delirium. However, inattention, sleep-wake cycle disturbances, disorientation in time, place and identity are the core symptoms that are frequently observed in the patients. Inattention is one of the core features for the diagnosis of delirium and is observed in 97%-100% of the delirium patients [19]. Hence, a measure of inattention in ICU patients can provide a good way of diagnosing delirium.

This research aims to study existing tests of sustained attention used for the diagnosis of delirium and research the ways to implement them on the VitalSky. Since, the VitalSky is a very unique device, the different methods by which the test can be conducted with the VitalSky needs to be analyzed. The goal of this thesis is to design and develop an attention test for the VitalSky and benchmark it's performance with the reference test which was used for the development of the attention test. In second phase of the study, various devices will be explored to automate the attention test developed for the VitalSky such that it can be performed with no or minimal intervention of the nurse or experimenter. In the later phases, a pilot study will be performed on 10 healthy volunteers to benchmark the performance of the developed tests with the reference test and to collect the subjective feedback of the participants. In the pilot study the participants will undergo five tests of sustained attention including the reference test and the performance of the participants in the tests will be used to benchmark the performance of the developed test with the reference test.

The relevant literature and the tests studied during the research are discussed in Chapter 2. Chapter 3 focuses on the design approach adapted for development of the visual content for the VitalSky and the different designs of the attention test that were developed during this thesis. Chapter 4 will explores the automated setup of the test and describes various devices used in the automated setup, how the devices were selected and how the entire system is integrated. The fifth chapter of the thesis explains the attention tests that will be used in the pilot study. The objective, protocol and results of the pilot study are explained in Chapter 6. In this chapter, the scores of the participants from different attention tests are compared and their subjective feedback are used to understand the o views of the participants about the visual content of the test and preference of the response device. Chapter 7 concludes the research and provides suggestions for the future work.

2

Literature Review

In this chapter, we discuss the symptoms of delirium which is followed by how delirium is diagnosed in the ICU and the problems with those methods. Then, we look for methods which have been useful to detect delirium and can help to overcome the problems associated with the methods used to diagnose delirium in the ICU. Lastly, critical analysis and comparison of the performance of the discussed tests are performed to figure out which test was better than the rest in diagnosing delirium in the ICU.

2.1. Symptoms of delirium

There are a wide range of symptoms that are associated with delirium. These symptoms include cognitive disturbances such as attentional deficits as well as non-cognitive neuropsychiatric symptoms like disturbed sleep-wake cycle, increased or reduced motor behaviors, disorganized thinking and others. Attention is one of the core diagnostic features of delirium and is impaired to an extent that the delirious patients struggle to mobilize and maintain sustained attention. It is prevalent in 97%-100% of the delirium cases [20].

Since attention is such an important aspect for diagnosing delirium, it indicates that inattention can be used to detect delirium. There are many tests which detect delirium by assessing the sustained attention ability of the patients. Hence, it is important to understand how the existing tests of delirium detection are performed on the patients. So, let us explore how is delirium detected in hospitals, what are the problems associated with it and why is there a need for new or innovative methods for detection of delirium.

2.2. Delirium detection in practice

There is no consensus about which is the best test for delirium detection in clinical settings[21]. However, CAM-ICU and ICDSC are the two most commonly used methods to detect delirium in the ICU [22]. In this section, we will discuss about these two tests and the problems associated with these tests.

2.2.1. Confusion Assessment Method-Intensive Care unit (CAM-ICU)

CAM-ICU is a brief, accurate and reliable tool to detect delirium in the ICU patients [23]. It is a quick test for delirium diagnosis and is based on the DSM-III diagnostic criteria. It can be used for nonverbal mechanically ventilated patients. It can be completed in 2-3 minutes and can be incorporated into daily routines of nurses and patients. It has the following four features:

1. An acute onset of mental status or a fluctuating course.
2. Inattention
3. Disorganized thinking

4. An altered level of consciousness.

The patient is said to be delirious if he or she has feature 1 and feature 2 and either of feature 3 or 4. The initial validation study displayed great promise to detect delirium using the CAM-ICU [23]. However, other studies addressed that it has moderate sensitivity due to which milder cases of delirium could be missed and it was also shown that the available CAM-ICU training program[24] is essential for the accurate judgement of relevant diagnostic domains[17].

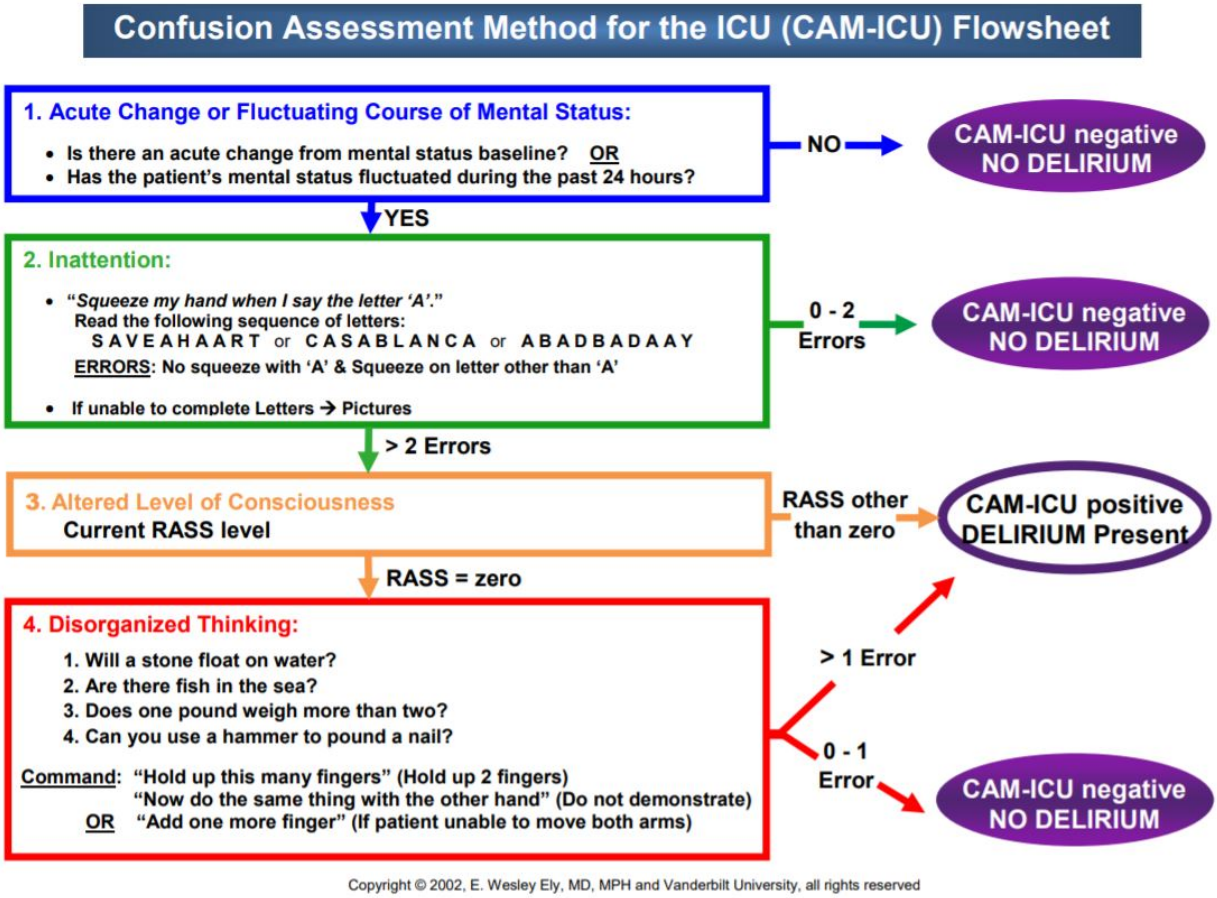
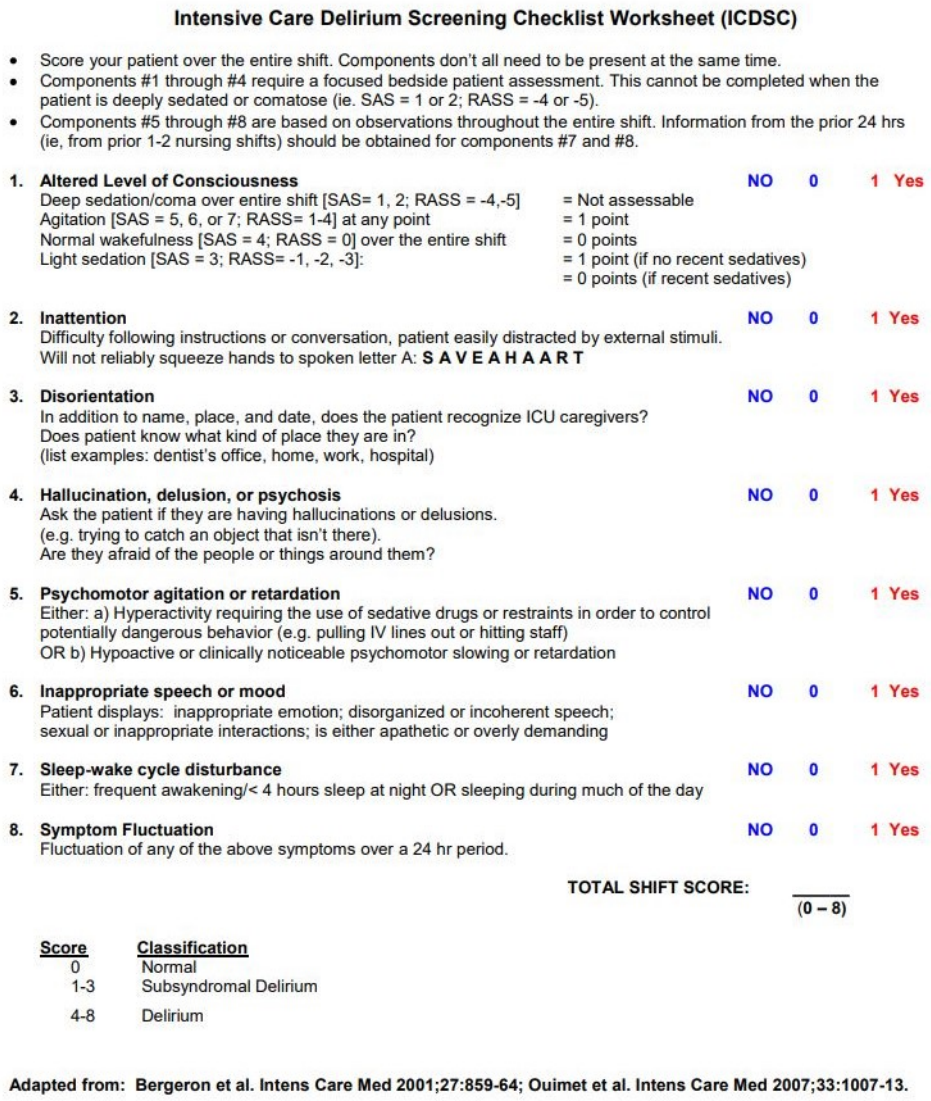


Figure 2.1: Different features of the CAM-ICU are described in the image. If the feature 1 and 2 and either 3 or 4 are present, patient is assessed as delirious. Image taken from: E, Wesley Ely, MD, MPH and Vanderbilt University. The image is copyrighted and all the rights are reserved by the owner.

In CAM-ICU, the nurse assesses if the patient has experienced fluctuations in the mental status in the past 24 hours. If the nurse observes fluctuation in the mental status of the patient then they proceed to examine inattention in the patients. Inattention is assessed by asking the patient to squeeze the nurse's hand whenever they hear a letter 'A'. However, if the patients are unable to complete the letter task, then inattention is examined using Attention Screening Examination(ASE) or a visual task. The ASE or a visual task is performed generally in <5% of the cases [25]. If mental state fluctuation and inattention are both present in the patient then the nurse assesses the altered level of consciousness and disorganized thinking in the patients. If either disorganized thinking or altered level of consciousness is present in the patient, the patient is diagnosed with delirium. This process is illustrated in figure 2.1. Although CAM-ICU has nice sensitivity and specificity, some studies have shown it to be partly dependent on the experience of the tester [26, 27] and have moderate sensitivity in routine assessments [28]. Thus, CAM-ICU is good in diagnosing delirium but not very good in terms of being an objective test for delirium diagnosis.

2.2.2. Intensive Care Delirium Screening Checklist (ICDSC)

ICDSC [29] is a quick and easy tool for delirium classification but can also be used for delirium diagnosis [30]. It consists of eight features which are assessed by the nurses. Each feature accounts for one point and the patient can have a score from zero to eight. The assessment is done based on the patient's behaviour in the previous 24 hours. In ICDSC, if a patient obtains a score of four or more, the patient is said to be suffering from delirium. The checklist for the ICDSC is shown in figure 2.2.



Intensive Care Delirium Screening Checklist Worksheet (ICDSC)

- Score your patient over the entire shift. Components don't all need to be present at the same time.
- Components #1 through #4 require a focused bedside patient assessment. This cannot be completed when the patient is deeply sedated or comatose (ie. SAS = 1 or 2; RASS = -4 or -5).
- Components #5 through #8 are based on observations throughout the entire shift. Information from the prior 24 hrs (ie, from prior 1-2 nursing shifts) should be obtained for components #7 and #8.

1. **Altered Level of Consciousness** NO 0 1 Yes
   Deep sedation/coma over entire shift [SAS= 1, 2; RASS = -4,-5] = Not assessable
   Agitation [SAS = 5, 6, or 7; RASS= 1-4] at any point = 1 point
   Normal wakefulness [SAS = 4; RASS = 0] over the entire shift = 0 points
   Light sedation [SAS = 3; RASS= -1, -2, -3]: = 1 point (if no recent sedatives)
   = 0 points (if recent sedatives)
2. **Inattention** NO 0 1 Yes
   Difficulty following instructions or conversation, patient easily distracted by external stimuli.
   Will not reliably squeeze hands to spoken letter A: **S A V E A H A A R T**
3. **Disorientation** NO 0 1 Yes
   In addition to name, place, and date, does the patient recognize ICU caregivers?
   Does patient know what kind of place they are in?
   (list examples: dentist's office, home, work, hospital)
4. **Hallucination, delusion, or psychosis** NO 0 1 Yes
   Ask the patient if they are having hallucinations or delusions.
   (e.g. trying to catch an object that isn't there).
   Are they afraid of the people or things around them?
5. **Psychomotor agitation or retardation** NO 0 1 Yes
   Either: a) Hyperactivity requiring the use of sedative drugs or restraints in order to control potentially dangerous behavior (e.g. pulling IV lines out or hitting staff)
   OR b) Hypoactive or clinically noticeable psychomotor slowing or retardation
6. **Inappropriate speech or mood** NO 0 1 Yes
   Patient displays: inappropriate emotion; disorganized or incoherent speech; sexual or inappropriate interactions; is either apathetic or overly demanding
7. **Sleep-wake cycle disturbance** NO 0 1 Yes
   Either: frequent awakening/< 4 hours sleep at night OR sleeping during much of the day
8. **Symptom Fluctuation** NO 0 1 Yes
   Fluctuation of any of the above symptoms over a 24 hr period.

TOTAL SHIFT SCORE: ____ (0 – 8)

Score	Classification
0	Normal
1-3	Subsyndromal Delirium
4-8	Delirium

Adapted from: Bergeron et al. Intens Care Med 2001;27:859-64; Ouimet et al. Intens Care Med 2007;33:1007-13.

Figure 2.2: The ICDSC flowchart describing the features of the ICDSC screening. The eight features listed are assessed by the nurse and some features requires the nurse to assess the behavior of the patient in the past 24 hours. Image taken from: [1]. All rights of the image are reserved by the owner.

The initial validation study of the ICDSC obtained a sensitivity of 99% and specificity of 64%. Some of the features of the ICDSC were such that the nurses could monitor them in their daily routine. Hence, the patients were indirectly involved in the ICDSC [31, 32]. In addition, the specificity of the ICDSC was also considered to be low and the sensitivity was not as good as the CAM-ICU [33, 34].

2.2.3. Problems with existing tests

As discussed earlier, delirium is seen in upto 80% of the ICU population [12] and despite the severity of the delirium it is still under-detected in the ICU [9]. There are a number of factors that could be possible for

underdiagnosis of delirium. Firstly, most of the tests used to detect delirium in clinical settings are subjective tests relying on the experience of the tester[17]. So they are suited for use by experts but not for untrained nurses or physicians. This might be one of the factor which reduce the possibility of early detection of delirium which is a vital element for the prevention of delirium. Secondly, the patients tend to get used to the questions of the CAM-ICU over time and can learn how to answer the questions of the test. Also, the symptoms of delirium and other cognitive disorders like dementia overlap to a high degree [18].

Hence, there is a need for an objective test that can be used by clinicians with no or little experience of diagnosing delirium. The test should be sensitive enough to detect mild forms of delirium so that it can help with the detection of delirium in early stages and at the same time should display high specificity to differentiate between delirium and other cognitive disorders with overlapping symptoms like dementia.

2.3. Objective tests of attention

Various objective attention tests have been developed for the detection of delirium. The tests vary from being pen and paper based to smartphone-based tests. They span across the auditory and visual domains while testing the ability of the patients to sustain their attention on some pre-defined target. There are also tests which do not rely exclusively on the sustained attention ability of the participants and also require them to exhibit other abilities like psychomotor abilities, memory retention, speeded response formulation etc. We will be going through some of the popular tests of attention in this section.

2.3.1. Trail Making Test (TMT)

TMT [35] is a popular neuropsychological test. It assess the selective, divided and sustained attention along with visual searching abilities of the patient. Apart from the attentional domain it also relies on the psychomotor speed abilities of the patients. There are two types of TMT test: TMT-A and TMT-B. In the TMT-A test, the participants are presented with a sheet of paper consisting of 25 encircled numbers and the participants are required to connect the numbers with a line in an ascending order as shown in figure 2.3. In TMT-B, there are also alphabet in addition to the numbers and the participants should connect the circles in ascending order alternating between numbers and letters (e.g 1, A, 2, B, 3 ,C and so on).

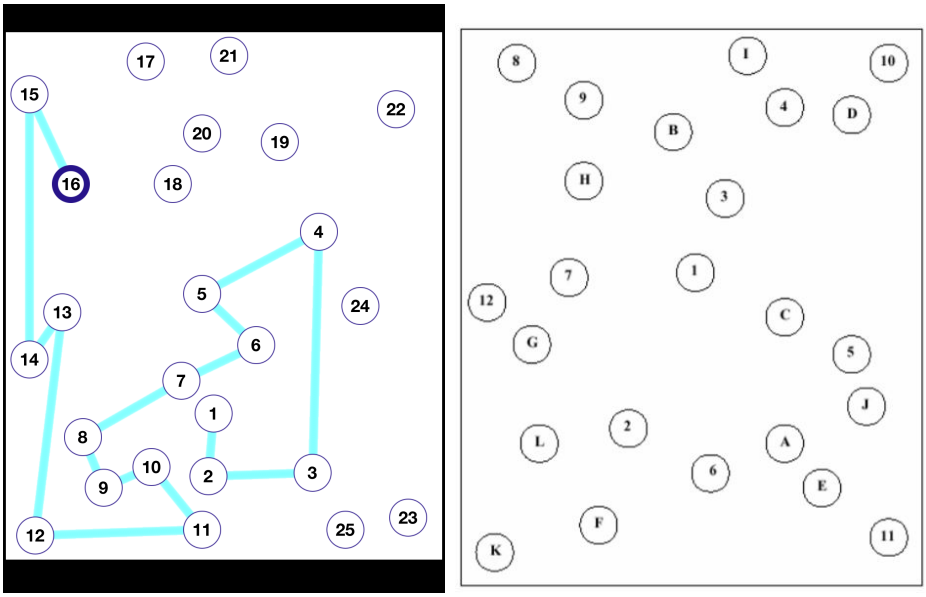


Figure 2.3: TMT-A(left) [2] and TMT-B(right) [3] tests. TMT-A consists of only numbers and participants need to connect them in an ascending order. TMT-B consists of numbers as well as letters and the participants must alternate between letter and number when connecting the circles in ascending order.

In a study[36] which compared the performance of 20 delirious and nine schizophrenic hospitalized patients undergoing the TMT-B test, it was found that both groups were severely impaired in the TMT-B. A significant

correlation was observed between TMT-B scores and Delirium Rating Sacle(DRS) scores(r=0.66)[37]. However, only nine out of 20 patients were able to attempt the test. Another study [38] evaluated the performance of TMT-A and TMT-B on 52 stem cell transplantation patients and 10 healthy volunteers. During their hospital stay 19 patients developed delirium and there was a decrement in the post-operative performance on the test. However, there was no difference noted in the performance between patients with and without delirium.

2.3.2. Digit Span Test (DST)

The DST requires the patients to listen to a sequence of digits, beginning with two-digit numbers, and repeat them. The DST has two variants: Digit Span Forwards(DSF) and Digit Span Backwards(DSB). In DSF, the nurse recites a sequence of numbers to the patients and the patients have to repeat the sequence in the same order as the nurse [39]. The nurse starts with a two-digit number and on every correct response increases the next sequence by one digit. For ex. The nurse recites a two-digit number and if the patient responds correctly then the nurse recites a three-digit number and so on. The participants must be able to repeat at least five-digit number else the score is considered to be abnormal. The test scores of the patient are assigned based on the maximum number of digits they could repeat consistently without error.

DIGITS FORWARDS

Item	First Trial	Result	Second Trial	Result	Total
A	56		32		
B	325		386		
C	1835		5963		
D	25186		15192		
E	786932		493217		
F	8539652		2865752		
				Forwards Score:	

DIGITS BACKWARDS

Item	First Trial	Result	Second Trial	Result	Total
A	23		17		
B	381		956		
C	2371		6843		
D	23156		17863		
E	764237		265479		
F	5476876		2953816		
				Backwards Score:	

Figure 2.4: The DSF and DSB versions of the DST. There is increase of a digit after every correct response. A wrong response leads to second trial of the same digit. A maximum of five digits is used for delirium diagnosis using DSF and three digits for DSB.

In case of DSB, the task is very similar to the DSF except the patients had to repeat the sequence in reverse of the original order [40]. The nurse starts with a two-digit number and continues upto a maximum of three-digit numbers. Similar to DSF, the score is assigned based on the sequence of numbers that the patients can repeat consistently without errors. For DSB, if the patients could not repeat three-digit numbers, the DSB score is considered abnormal.

O'Keefe et al performed a study[41] on 87 acute geriatric patients, out of which 18 had delirium, 17 had dementia and 52 had neither dementia nor delirium. The patients had to undergo four tests of attentiveness, two of which were DSB and DSF. The study found that there was no difference in the results of delirious and demented patients for DSF. However, the difference in the performance of delirious and control and demented and control were significant. Hence, DSF was sensitive to delirium but specificity of DSF for delirium was really bad. On the other hand, patients with delirium, dementia and neither displayed significantly different performances for DSB. DSB was able to differentiate between delirium and dementia and had a sensitivity of 83% and specificity of 88%.

Pompei et al [39] tested DSF on medical and surgical patients(n = 419, delirium = 56, control = 363) and observed a sensitivity of 34% while a specificity of 90% for the test. In the DSF, a performance improvement over time was observed in the patients [40, 42] . In a study conducted by Brown et al [43], 37 patients undergoing

cardiac surgery were tested on the DSB. Out of 37 patients, 9 patients developed delirium during the hospital stay. The patients took the DSB test before and after the surgery. There was a decrement in the pre-operative to post-operative performance of the patients who developed delirium.

2.3.3. Spatial Span Test (SST)

The SST [44] is similar to the DST but includes visual stimuli rather than auditory stimuli. The SST assess the spatial attention span of the patient and visuospatial memory abilities. The SST has two variations namely: Spatial Span Forward(SSF) and Spatial Span Backward(SSB). The test consists of a sequence of blocks arranged in a grid or any other pattern. The nurse or the examiner taps the blocks in a certain sequence. The participant must tap the blocks in the same sequence for the SSF and in the reverse sequence for the SSB. The participant is scored based on the maximum length of sequence the participant consistently reproduces without error.

In a study[21], 265 general hospital adult patients were asked to do the spatial span forward test. Out of the 265 patients 48 were delirious and remaining patients belonged to the control group. With a cut-off of 5 correct box tappings, spatial span forward obtained a sensitivity of 92% and specificity of 69%. The test missed a quarter of delirium cases if the cut-off level was set to 4 blocks. In a study by Meagher et al [45] involving 140 palliative care patients(40 delirium, 40 delirium + dementia, 20 dementia and 40 control), SSF was able to distinguish control group from delirium, dementia and delirium superimposed dementia groups. SSF was also found to have significant correlation with DRS-R-98 [46, 47]. However, Rajlakshmi et al [48] did not find a correlation between SSF and DRS-R-98 in their study.

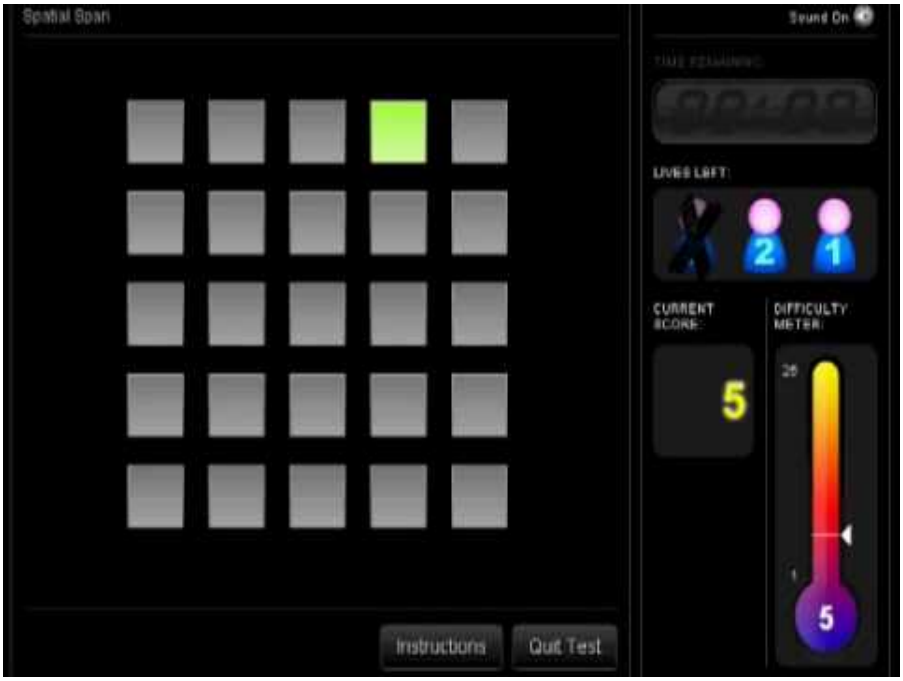


Figure 2.5: A computerized version of the SST. The squares light up in a random pattern and the participant must reproduce the same pattern for SSF and reverse pattern for SSB. The participant can register his or her response by using a mouse. [4]

2.3.4. Vigilance Test

Vigilance test [49] helps to assess an individual's attention abilities by requiring them to detect one or more pre-specified target among a range of other stimuli. It assess the sustained and divided attentional ability of the patients. There exists many variations of the vigilance test and depending on the variation it may or may not involve psychomotor abilities of the patients. Most common variation is the "Vigilance A" test in which 'A' is the pre-specified target and the patient need to perform an action whenever an 'A' appears and do nothing for the other stimuli. Vigilance test can be performed with an audio or video stimuli, in each of the

case the task remains the same i.e the patients need to identify the target stimuli and perform some action like tapping a pen, squeezing nurse's hand, pressing a button, etc.

D Adamis et al[50], conducted the vigilance A test on acute medical patients aged 70 years and older. There were 200 participants, out of which 34 had delirium, 95 had dementia and 71 had neither dementia nor delirium. A list of 29 letters was presented to the patients and 'A' occured for 11 times in the list. The patients had to indicate each time they saw letter 'A'. If the patient indicated any letter other than 'A' or if the patient failed to indicate letter 'A', it was considered an error. If the patient completed the test with two or less errors they got a score of 1 otherwise they were assigned a score of 0. Vigilance test displayed a sensitivity of 82% and specificity of 60%. It was found out that the vigilance test had a high negative predicted value and a low positive predicted value.

In another study[41], vigilance A test was used and consisted of a list of 60 letters, 18 of which among them were letter 'A'. This test was performed on the acute geriatric patients. The patients were asked to tap on the table with a pen on hearing the letter 'A'. The list of letters were read out to the patient at the speed of 1 letter per second. The criteria for an error was the same as in previously mentioned study. In this study, Vigilance test obtained an overall sensitivity and specificity of 83% when considering all the patients. However, the specificity dropped to 35% when considering only the patients with cognitive impairment.

However, Hart et al 1996 [51] performed vigilance 'A' test and SST on 22 delirious ICU patients, 26 dementia, 30 depression and 25 schizophrenic psychiatry referred patients. But, no correlation was found between the performance of patients in vigilance 'A' test and their DRS-R98 score. Similar results were obtained by Rajalakshmi et al [48] when testing the vigilance 'A' test on 84 psychiatry referred delirium patients.

2.3.5. Clock Drawing Test (CDT)

The CDT is a test used to measure cognitive impairments and assess the visuo-constructive abilities of the patients [52]. In the CDT, the participants are required to draw a clock and put the hands of the clock at the mentioned time (see figure 2.6. For example, if the participant is asked to draw fifteen past 9, they need to draw the position of the hands as they would be at fifteen past 9. 1 point is awarded for drawing the clock, 1 for numbers and 1 for hands. So, in total the participant can have a maximum of three points.

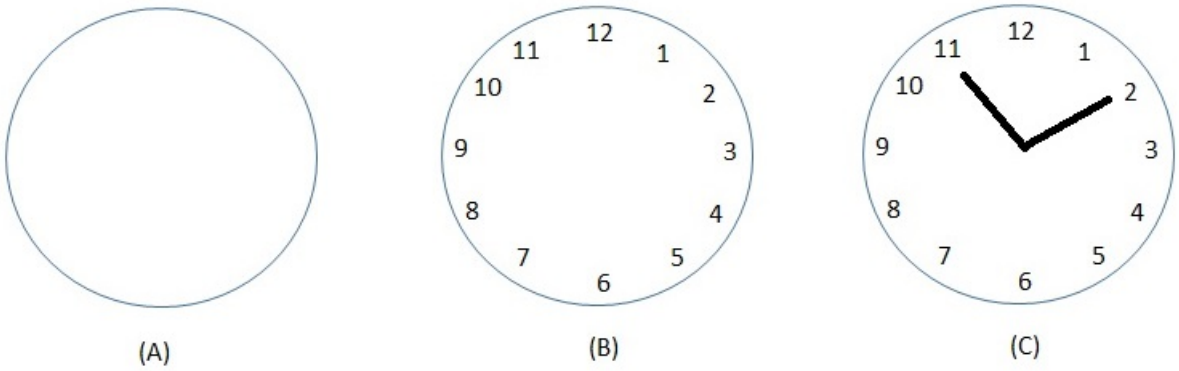


Figure 2.6: Different stages of the clock drawing test. (A) Drawing the boundary (B) Putting the numbers (C) Drawing the hands for eleven past ten. Each step accounted for 1 point

In a study[53] conducted to evaluate the usefulness of the CDT for the detection of delirium a total of 200 acute medical patients took part out of which 46 patients were diagnosed with delirium. The patients being delirious or not was determined using CAM. The results of the study indicated that the performance of the patients in the CDT was not an indication of the presence or absence of delirium. The performance of CDT was mainly a prediction of the general cognitive ability and not a diagnosis for delirium. The CDT could provide a simple and time-efficient solution but at its current state there is no substantial proof to support that CDT can be used as a diagnostic tool for delirium.

Bryson et al[54] also concluded that CDT was not a good test for screening delirium. However, in a study [55] testing 90 urological surgery patients with CDT. The patients had to take CDT before and after the surgery. A

significant difference was observed in the pre-operative and post-operative performance of the patients who developed post-operative delirium.

2.3.6. Serial Sevens Subtraction Test (SSST)

In SSST [56], participants start from 100 and sequentially subtract seven from it. The participant can obtain a score ranging from zero to a maximum of three. If the participant do not make any correct subtraction they get a score of zero and a score of one if there is one correct subtraction. The score for two or three correct submissions is two and a score of three is obtained on performing four or five correct subtractions [50].

D Adamis et al [50] performed a study involving 200 acute medical patients out of which 34 were diagnosed with delirium using the CAM. The participants were asked to perform four tests one of which was the SSST. The SSST obtained the sensitivity and specificity of 62% and 75% respectively when cut-off was set to the score of one. For a cut-off score of three serial sevens displayed a sensitivity of 91% and a specificity of 46%. Hence, the serial sevens subtraction test was good at detecting delirium but performed poorly in differentiating delirium from other cognitive disorders.

2.3.7. Month Of The Year Backwards (MOTYB)

The MOTYB [57] test is also known as months backwards test. The MOTYB test assess attention, however other studies have also found it to be assessing concentration [58], working memory, executive function [59], cognitive flexibility [60] and central proessing speed [61]. As the name suggests, the participants of the test need to recite the months of the year in reverse order. The participants who could recite the months upto July without any error are considered to pass the test. However, if the participant could not recite the months till July, the participant is said to have inattention. The score is either 0 (inattention present) or 1 (attention intact).

In a study [62], 265 general hospital patients were tested with MOTYB test. In the total of 265 patients, 48 patients were found to be delirious. MOYTB displayed a high sensitivity and a modest specificity. MOTYB had a sensitivity of 82% and specificity of 66%. It was also observed that MOTYB had a high negative predicted value but a low positive predicted value. Another study [63] conducted MOTYB test on 234 acute geriatric ward patients. Among the 234 patients, 29 were diagnosed with delirium using DSM-IV criteria by a geriatrician. The MOTYB showed a good sensitivity but the specificity was low in the case of delirium.

2.3.8. Digit Cancellation test (DCT)

This test focused on the sustained attention, visual search and divided attention abilities of a patient in the attentional domain. The test determined the participants ability to search for a pre-defined target among other stimuli. The participants were given 1 or 2 target letters and they needed to cross out the target letters among a list of letters. The test is scored based on the number of letters crossed out. If the participant misses a target letter or crosses out a letter other than the target letter, it was considered as an error. The participants get 1 point for each correct cancellation and a negative score for wrong cancellation or missing the cancellation of a target letter.

DCT was tested on 87 patients of whom 18 were delirious, 17 had dementia and 52 had neither delirium nor dementia [41]. For a cut-off score of less than 8, DCT achieved a sensitivity of 78% and a specificity of 76%. But for a cut-off score of less than 9, DCT had a sensitivity of 94% and specificity of 87%. DCT was sensitive to delirium and also able to successfully distinguish delirium from dementia.

2.3.9. Edinburgh Delirium Test Box(EDTB)

Edinburgh Delirium Test Box(EDTB) [6] is a custom-made hand-held device for bedside detection of delirium. It assess the sustained attention abilities of the patients. The test box is made to be portable, robust and also have large illuminable response buttons as shown in figure 2.8. The test box also has distracting LEDs which is used to distract patients during the test. The patients must focus on one or both the response buttons and count the number of illuminated of the button and in some levels ignore the distracting LEDs.

Figure 1: The six-letter cancellation test

SIX-LETTER CANCELLATION TEST (SLCT)

Name:_____ Course:_____ Date:_____ Session:_____

Target letters: R L F T O Z

Cancel the above six letters with a slash in the given text, at your maximum speed.
Time allowed: 90 seconds

Q W E R T A Y U I O P A S D F G H L J K M Z X C V B N T M Y S Q C
B M C Z K L J G D A P I Y B R W Q E T U O L N J G D X V N O M J B
P H L O K M I J N U H B I Y G V T F C R D X E S Z C W A Q W I X E
L P M K O N J R I B H U V G Y C F T X D N Z S E A Q P W K E V F K
P O I D U Y T R E V W Q Z X C V B N M L K J H G F D S F A B P G O
A M P N O B I V U C Y X T I Z G R E W Q L K J H F X D S E R U C H
A D S F G J H K L Z C X V B W M N Q J E W R T U Y I O P M H N F Z
Z P L K O M J I N Y H U B G Y V F T C D R X S E Z A Q W Y A Z G V
Q P W O E I R U T Y A L S K D J F H G Z M X N C B V D H U P A R W
Z X C V B N M A S D F G H J K L Q W E R T Y U I O P Q L R F S U X
S Q A Z W S X O E D C R F V T G B Y H Q N U J M I K L O P K L T S T
M N B V C X Z L K J H G F D S A P O I U Y T R E W QA E I O U R A

Total attempted:
Incorrect:
Net score:

Figure 2.7: A paper based six-letter DCT. The participants must scan through the list of letters and cancel out all the letters except the target letters. The one-letter DCT and two-letter DCT are the most common DCT[5].

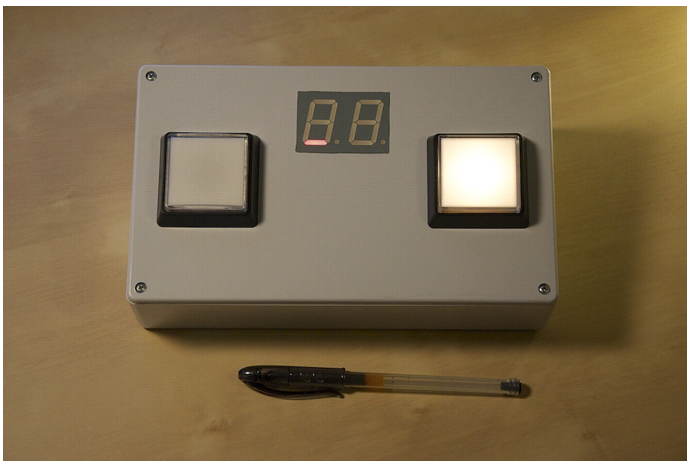


Figure 2.8: Edinburgh Delirium Test Box, a custom-made hardware device for detecting delirium. It consists of two illuminable buttons and distracting LEDs in between the two buttons. The patients should count the number of illuminations of the buttons while ignoring the distracting patterns from the LED(if any) [6]

The test consists of eight sustained attention tasks of differing complexity. The complexity was altered by adding distracting patterns by the central LEDs or by asking the participant to focus on two instead of one response button. The first four tasks requires the patients to count the number of blinks of the response button and respond it verbally whereas in tasks 5-8 the patients had to press a button each time a response button is illuminated.

In task 1, the test box is placed in portrait orientation and the patient must count the illumination of the response button closest to them. Task 2 is similar to task 1 but also consists of distracting stimuli on the central LEDs. In task 3, the box is kept in landscape orientation and the patient must count the illumination of both the response buttons. No distraction appears in task 3. In task 4, the task for the patient remained same but with added distraction from the central LEDs. The tasks 5-8 are exactly similar to tasks 1-4 but the patient must respond by pressing a button each time a response button illuminates.

The performance of EDTB was analyzed by performing a clinical study with general ward patients[6]. The study consisted of 20 patients with delirium, 18 with dementia and 20 control group patients. The score of delirium patients was the lowest in all the eight sustained attention tasks. The scores of patients with dementia and control group patients was similar except on task 4 where patients with dementia scored less than control patients.

The study demonstrated that EDTB was capable of detecting delirium and also able to distinguish it with dementia. However, the study had some drawbacks. Firstly, some patients failed to complete all of the eight tests as the the total test took around 35-40 minutes to administer. Secondly, it could not determine how EDTB would perform for varying severities of delirium and dementia(eg. mild delirium versus severe dementia).

2.3.10. EDTB-ICU

The EDTB-ICU [64] was developed to facilitate the EDTB to detect delirium for intubated ICU patients. As a result, it included non-verbal response methods like pointing, blinking, nodding, sticking tongue out and squeezing the examiner's hands. Similar to the EDTB, EDTB-ICU also assesses the sustained attentional abilites of the patients. The EDTB-ICU consists of two tasks:

1. behavioral assessment task

2. attention task

The behavioral assessment task assesses the participant's level of arousal and basic orienting response. The participants need to respond to their name with eye contact and then if track the examiner's moving finger for five seconds. The patients who successfully complete the behavioral assessment then proceed to the attention task otherwise the test is terminated.

The attention task of the EDTB-ICU consists of one practice trial and nine test trials. In each trial the patient must count the number of illuminations of the target button and ignore the distracting LEDs that might flash along with the target. For each correct response, the patient gets one point and no points if the response is wrong. The target lights blinks between three to six times in each trial and the difficulty of the tasks increases progressively.

A study [64] was performed to test the EDTB-ICU over the ICU patients. The study included 30 ICU patients out of which 15 had delirium. The maximum score possible for the test was 11 i.e two for behavioral assessment and nine for the attention task(one for each trial). For detection of delirium, the EDTB-ICU score of five or less was 100% sensitive and 92% specific whereas a score of four or less was 100% sensitive and 93.7% specific. There was a substantial difference noted between the scores of delirious and non-delirious people.

The results showed that the EDTB-ICU is a promising and objective tool for the detection of delirium with excellent sensitivity and specificity. It was also much faster than the EDTB and took three to seven minutes to administer the test and hence, it was easier for the patients with delirium to complete this test.

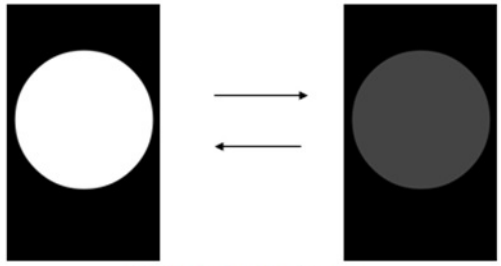
2.3.11. DelApp

DelApp [7] is a software application based on the EDTB for detecting delirium by assessing impairment in the visual sustained attention of the ICU patients. DelApp consists of two tasks:

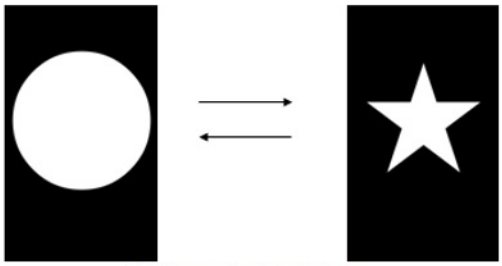
1. visual acuity task
2. visual sustained attention task

The visual acuity task tests the ability of the patients to distinguish the color and shape of the stimuli. It consists of six short trials where the participants have to identify a change in color of the stimuli (figure 2.9(a)), identify a change in shape of the stimuli (figure 2.9(b)) and identify the alphabet presented on the screen(figure 2.10). Only if the participant successfully completes the visual acuity task, they can attempt the visual sustained attention task.

In the visual sustained attention task, the participants need to count the number of blinks of the circle and ignore distracting triangles which may appear randomly around the circles. The visual sustained attention task consists of three levels and seven trials of counting task. In level 1 (trial 1-3), the participants are presented with only blinking circles and no distractor appears. From level 2(trial 4-5) onwards, the distracting triangles are introduced and the number of triangles might vary between one to four. The number of distractors increases in level 3 and vary between five to eight. The participant gets one point for each trial in which they responded the number of blinks of the circle correctly otherwise zero points.



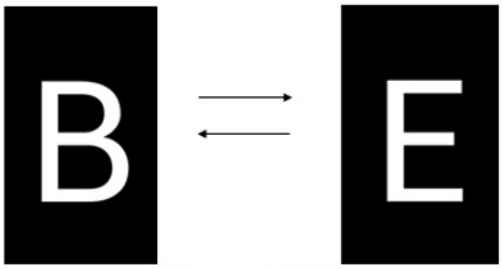
(a) Change in stimuli color



(b) Change in stimuli shape

Figure 2.9: The visual task in DelApp. The patients should be able to recognize the change in stimuli color (figure 2.9(a)) and change in stimuli shape (figure 2.9(b)). If the patients are not able to complete these tests they are declared as unfit for the visual sustained attention task of the test[7].

To assess the performance of DelApp in geriatric delirium patients and to evaluate if DelApp was able to distinguish between delirium and dementia, a total of 156 patients were recruited from orthopedic and acute and rehabilitation wards. [7]. Among the total population of 156 patients, 50 had delirium, 52 had dementia and 54 were control group patients and did not suffer from delirium or dementia. The DelApp scores for the group with delirium was found to be significantly lower than the patients with dementia and control group and the patients with dementia scored significantly lower than the control group. Hence, DelApp was not only able to detect delirium but also distinguished delirium from dementia. It had a sensitivity of 98% when detecting delirium and a specificity of 93%.



(c) Change of alphabets

Figure 2.10: Alphabet change recognition task in DelApp Visual task where the patients need to recognize the alphabet shown by the nurse. This part of the task was discontinued later because the nurse already used to assess the same in behavioural assessment[7].

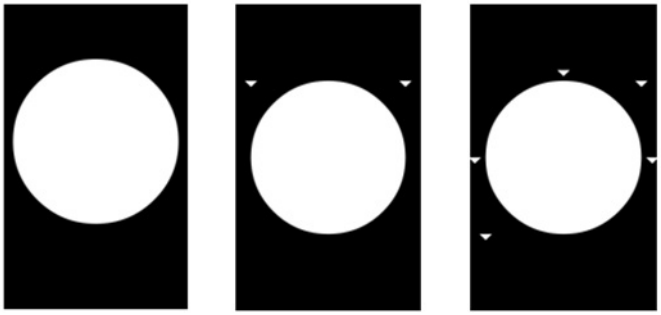


Figure 2.11: Three levels of the visual sustained attention task of the DelApp. Level 1(left) involves just the target stimulus. The distractors are introduced from level 2(center) and the number of distractors increases in level 3(right). In all the levels, the patients must count the number of blinks of the circle and tell it to the nurse[7].

2.3.12. DelApp-ICU

DelApp-ICU[65] is similar to DelApp and was developed such that it can be used with ICU patients. It assess the sustained attention abilities of the participant and includes behavioral assessment and also incorporates non-verbal response methods. The test consists of three tasks namely:

1. behavioral assessment
2. visual task and
3. visual sustained attention task

The participant could proceed to the next task only if they completed the previous task successfully(figure 2.12). The behavioral assessment in DelApp-ICU is similar to EDTB-ICU with only difference being that behavioral assessment in DelApp-ICU is a 3 point scale rather than a 2-point scale. The visual task of DelApp-ICU is similar to the visual task of DelApp and do not account for any points in the final score.

The visual sustained attention task in DelApp-ICU (figure 2.11) was similar to DelApp and the participants need to count the blinks of the circle while ignoring the distracting triangles. The patients committed more mistakes in level 3 of the visual sustained attention task compared to level 2. This indicated a performance difference between the levels. However, this performance difference did not reach significance as a result two extra trials were added in the visual sustained attention task of the DelApp-ICU to make the performance difference substantial [7]. So, DelApp-ICU consisted of nine trials instead of seven. In the DelApp-ICU, a patient can score a minimum of zero point and a maximum of 12 points. The behavioral assessment accounts for three points and the visual sustained attention task has a maximum of nine points. At the end of each trial, the patient is asked to tell the number of times the circle blinked, if the patient tells the right count he or she gets one point otherwise zero points .So, the patient can get one point for each trial totalling to a maximum of nine points in the visual sustained attention task.

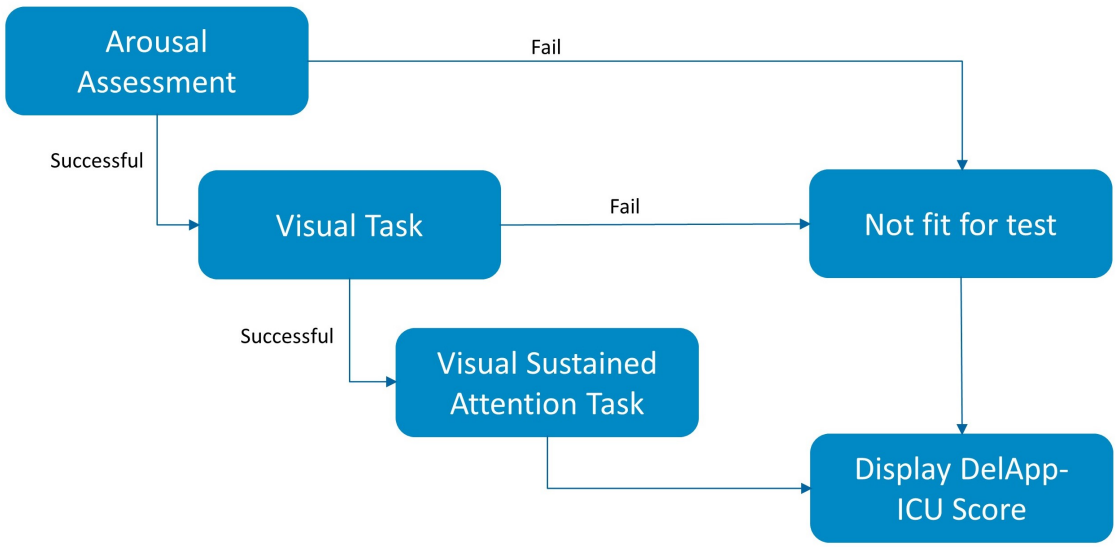


Figure 2.12: Flowchart of the DelApp-ICU test. The participants can only proceed to a next task upon successfully completing the previous task.

Tang et al conducted a study [65] which involved 47 ICU patients involving 20 delirious and 26 non-delirious patients undergoing DelApp-ICU assessment. The DelApp-ICU took approximately three to eight minutes to complete. The study found that patients with delirium scored significantly lower than the patients without delirium. A DelApp-ICU score of six or less was 100% sensitive and 96% specific for delirium whereas a score of five or less had a sensitivity of 95% but had a specificity of 100%. The positive and negative predicted values for the DelApp-ICU were 91% and 100% respectively.

2.4. Comparison of attention tests

There are a vast range of attention tests that have been discussed in the previous section. The tests vary from being paper-pen based to using smartphones. Some of the tests rely only on the core attentional abilities of the patients whereas other tests also required rapid perceptual processing of information, speeded response formation or execution of motor responses. Although the nature of the tests vary largely from one another, a thorough comparison between the different tests is essential to come up with the test that displayed the best promise to be used as a diagnostic test for delirium. The performance and other details about the various attention tests is summarized in the Table 2.1. CDT and TMT are not mentioned in Table 2.1, because the studies did not calculate the sensitivity and specificity for the tests and only observed a change in the pre-operative to post-operative performance on the test.

Table 2.1: Summary of the performance and the type of population for the different tests.

Test	Sensitivity	Specificity	Validation	Patients	Number of Patients		
					Delirium	Dementia	None
Vigilance 'A'	72%	71%*	DSM-III	Acute geriatric	18	17	52
SSST	77%	61%	CAM, DRS-R98	Geriatric	34	-	166
DSB	83%	88%	DSM-III	Acute geriatric	34	-	166
MOTYB	83%	91%	CAM, DRS-R98	General hospital	48	-	217
SSF	92%	69%**	CAM, DRS-R98	General hospital	48	-	217
DCT	94%	87%***	DSM-III	Acute geriatric	18	17	52
DelApp	98%	93%	CAM-ICU	Geriatric	50	52	54
EDTB-ICU	100%	92%	CAM-ICU	ICU patients	30	-	30
DelApp-ICU	100%	96%	CAM-ICU	ICU patients	20	-	26

Among all the tests discussed the tests DCT, EDTB, EDTB-ICU, DelApp and DelApp-ICU displayed really high

sensitivity for delirium. The sensitivity of the tests was important because having false positive were more preferred in case of delirium rather than having false negatives. Another factor was the ability of the test to be performed using the visual stimuli as the test was to be used as a reference test for the development of an attention test for the VitalSky.

The DCT was a quick and paper-based test showing a good sensitivity and specificity. But, it also relied on the ability of the patient to read and cross out the target letters and hence included a dependency on motor responses. Although EDTB and EDTB-ICU displayed excellent performance as a delirium diagnostic test, the use custom-built hardware reduces the scope for the application device. The hardware needs to be carried to each and every patient or made available in all the facilities. Moreover, the authors of EDTB and EDTB-ICU also reported that the nurses had a concern with the infection control of the device. Hence, EDTB and EDTB-ICU showed a great potential as a diagnostic tool for delirium but a solution implemented on universally available device like smartphones would be more readily acceptable by clinicians and in research.

The DelApp and DelApp-ICU are the successors of EDTB and EDTB-ICU developed by the same group and very identical to each other. They are developed taking into account the learnings from the EDTB and EDTB-ICU respectively. However, DelApp-ICU incorporated non-verbal responses as well which meant the test could also be used for the intubated ICU patients which was lacking in the case of DelApp. Also, in DelApp the performance difference of the patients in the levels of DelApp did not reach significance which was improved in DelApp-ICU. This is also evident by the performance of DelApp and DelApp-ICU as the sensitivity of DelApp increased from 98% to 100% in DelApp-ICU and specificity from 93% to 96%. So, the DelApp-ICU

1. only focuses on the sustained attention abilities of the patients.
2. makes use of visual stimuli.
3. can be used with the ICU patients.
4. achieved the best sensitivity and specificity.
5. is a quick and short assessment which increases the probability of patients finishing the complete test.

Considering all the above mentioned points, DelApp-ICU seems to be a good reference test to design and develop an attention test for the VitalSky.

2.5. Research Goal

The research goal of this thesis is to

> To design and develop an attention test for the VitalSky and benchmark the performance of the test with the reference test DelApp-ICU.

Apart from the primary goal, there are a number of secondary goals that we will be aiming to achieve in this study.

1. To develop an automated version of the attention test which can be used without or with minimal nurse intervention.
2. To gain insights on subjective experience of the participants about the used visual content in the different attention tests.
3. To understand the user experience and preference of the different set-ups for automatic execution of the attention test.
4. To assess the usability of the various automated setups.
5. To assess the technical feasibility of the various automated setups.

3

Exploration of the Visual Content

In this chapter, we examine the DelApp-ICU more closely and look into the details of the test and use it as a reference for designing the attention test for the VitalSky. We start by exploring the structure and design parameters of the DelApp-ICU. Later, in the chapter we go through the process of visual content design for the VitalSky. The chapter ends with explanation of different types of targets, distractors and visual designs used for the test.

3.1. DelApp-ICU

The DelApp-ICU performed well in diagnosing delirium and provided the best sensitivity and specificity while also being a quick and objective tool for detecting delirium. Here, we explore the design features of the visual content of the DelApp-ICU, the structure and scoring of the test. We will use this information as reference later to design the test for the VitalSky.

The behavioral assessment and visual task of the DelApp-ICU are used to assess the level of arousal of the participant and their ability to identify changes in shapes and size of the stimuli respectively. The measure of these abilities is important for the ICU patients but not so much for the healthy population. Since, the experiments in this thesis will involve healthy awake volunteers without any visual problems, the behavioral assessment and the visual tasks are out of scope of this thesis. As a result, we will only be focusing on developing the visual sustained attention task of the DelApp-ICU.

3.1.1. Setup of the DelApp-ICU

The visual sustained attention task of the DelApp-ICU consists of three levels which are further divided into trials. Each level consists of three trials which are performed sequentially (see figure 3.1). The visual content in DelApp-ICU consists of a target and a distractor (depending on the level). The patients are shown a visual content in each trial and the patients need to count the number of blinks of the circle and report the count at the end of the trial. For all the trials, the circle blinks between three to six times and is always visible for a duration of 1000ms. All the trials in all the levels are identical except for the interstimulus duration between the circle in the trails. The interstimulus duration is 1000ms in trial 1, 2000ms in trial 2 and 3000ms in trial 3. The number of distractors in the visual content vary with the level. The appearance of distractors is believed to make the task difficult for the patients. The distractors begin to appear in level 2 where the number of distractors varies from one to four. In level 3, the number of distractors increases and varies between five to eight. The distractors are always shown for a duration between 300ms - 700ms. An example image of the different levels is shown in figure 3.1.

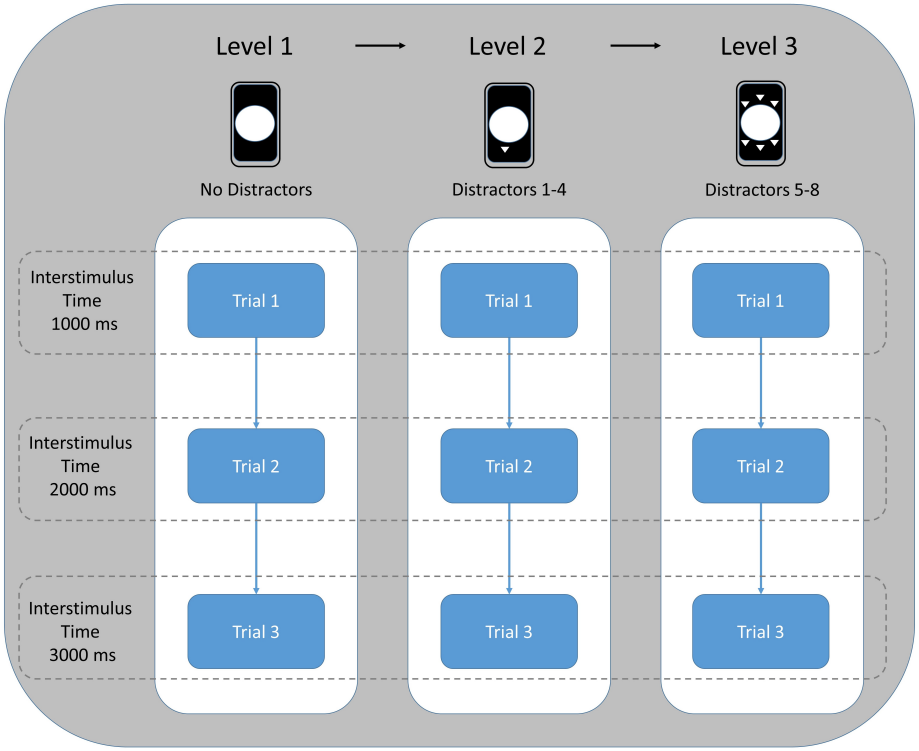


Figure 3.1: Division of the levels of the DelApp-ICU. Each level contains three trails which are performed sequentially. The participants are presented with one visual content in each trial and go through a total of nine visual contents in the test. Level 1 consists of just the target, i.e., white circle against a black background. The small downward pointing triangle distractors are introduced in level 2 and level 3.

3.1.2. Visual content of DelApp-ICU

The DelApp-ICU is an application on a Samsung Galaxy S2 smartphone. The smartphone has a screen size of 9.2 cm x 5.9 cm and is held at a distance of approximately 30 cm from the patients. The target is a large white circle of 5 cm diameter whereas the distractors are small downward pointing white triangles of dimensions 0.3 cm x 0.3 cm x 0.4 cm. The target and distractors are presented against a black background (see figure **??**). The circle is positioned at the center of the screen and always appears at the same position. The distractors appear around the circle and the number of distractors depends on the level of the test.

3.2. Visual content design for the VitalSky

The approach while designing the content for the VitalSky was to keep the visual content close to the visual content of the DelApp-ICU. However, the VitalSky being a unique device posed quite some challenges for designing the visual content:

1. **Size and field of view**: The DelApp-ICU was performed on a handheld smartphone device. This allowed the patients to move the device and adjust the it such that the visual content was in their field of view. However, as the VitalSky is a fixed luminous ceiling it is not possible for the patient to adjust the field of view. Moreover, the entire smartphone was in the field of view of the participant when the participant was looking at it. But, the VitalSky being a huge luminous ceiling, the participant will only have a certain section of the VitalSky in their field of view at a given time. Hence, the section of the VitalSky needs to be determined which is in the field of view of the patients.

2. **Resolution**: The VitalSky is a very low resolution device. The maximum possible resolution in the VitalSky is 40 x 100 LEDs, which corresponds to a resolution of 432px x 1080px, i.e., approximately 10 pixels are represented by 1 LED. The resolution of the VitalSky in market varies from 32 x 50 to 32 x 80 LEDs, which is quite low and designing visual content for such low resolution is quite challenging.The

VitalSky used in this thesis had a resolution of 32 X 70 LEDs or 346px x 756px.

The LEDs of the VitalSky are covered by a diffuser. The diffuser allows the light from the LEDs to spread uniformly and prevents the glare on the patients eye. However, because of this the visual content loses its sharpness and details and this gives the visual content a blurry look and feel.

3. **Position of the visual content**: In the VitalSky, the position of the elements in the visual content is very important as the patient cannot see the entire device in one sight. If the element is not in the right position, the patient will need to move his or her head to see the visual element which could result in discomfort for the ICU patients or making the test difficult.

4. **Size of the visual content**: Another important factor is the size of the elements of the visual content. If the position of the content is appropriate, but if it's size is not correct, the patient might have difficulty in seeing the content. If the size of the visual element is too big it might be difficult for the patient to focus on the element. On the other hand, if the content is too small the content might not be visible at all due to the blurred nature of the visual content on the VitalSky. Moreover, as the VitalSky is a luminous device, the size of the objects will also affect the luminosity of the environment. Hence, a too big of visual element may suddenly increase the brightness of the room and could lead to stroboscopic effect.

These factors make it challenging to design the visual content for the VitalSky while keeping the content close to the DelApp-ICU. The following sections describes how different aspects of the visual content was developed and the various design decisions that were made while designing the visual content for the VitalSky.

3.2.1. Scaled version

The idea was to scale the visual content of the DelApp-ICU for the size of the VitalSky. The dimensions of the VitalSky and the smartphone were compared and the percentage of area occupied by the target and distractors in the DelApp-ICU was calculated and the same percentage of area was allocated to the target and distractors for the VitalSky 3.2.

This provided a lot of insights that were necessary to design the visual content for the VitalSky. The insights are listed below:

1. **Target and distractors out of proportion** :The circle when scaled to the VitalSky dimensions was too big and focusing on the circle was difficult. The distractors on the other hand were very small and difficult to notice because of the big target.

2. **Color scheme** : The visual content had a black background and the target and distractors were white in color. This created a large fluctuation in the luminosity because of two reasons:

   - The circle being too big was the dominant and only source of light. So, whenever the circle appeared there was a bright light in the room and once the circle disappeared there was no light.
   - As the background of the visual content being black, there was no other source of light other than the target and distractors.

3. **Too bright visual content**: The white content on the VitalSky was too bright. This may have been because of the black background and sudden flashing of white visual elements, but also the brightness levels of the VitalSky were at the maximum. So the brightness of the VitalSky was then brought down to 30% of the maximum brightness of the VitalSky.

These learnings helped to understand that the visual elements of the DelApp-ICU needs to be altered for using it on the VitalSky. So, different parameters of the visual content was explored individually. These parameters are discussed in the sections below.

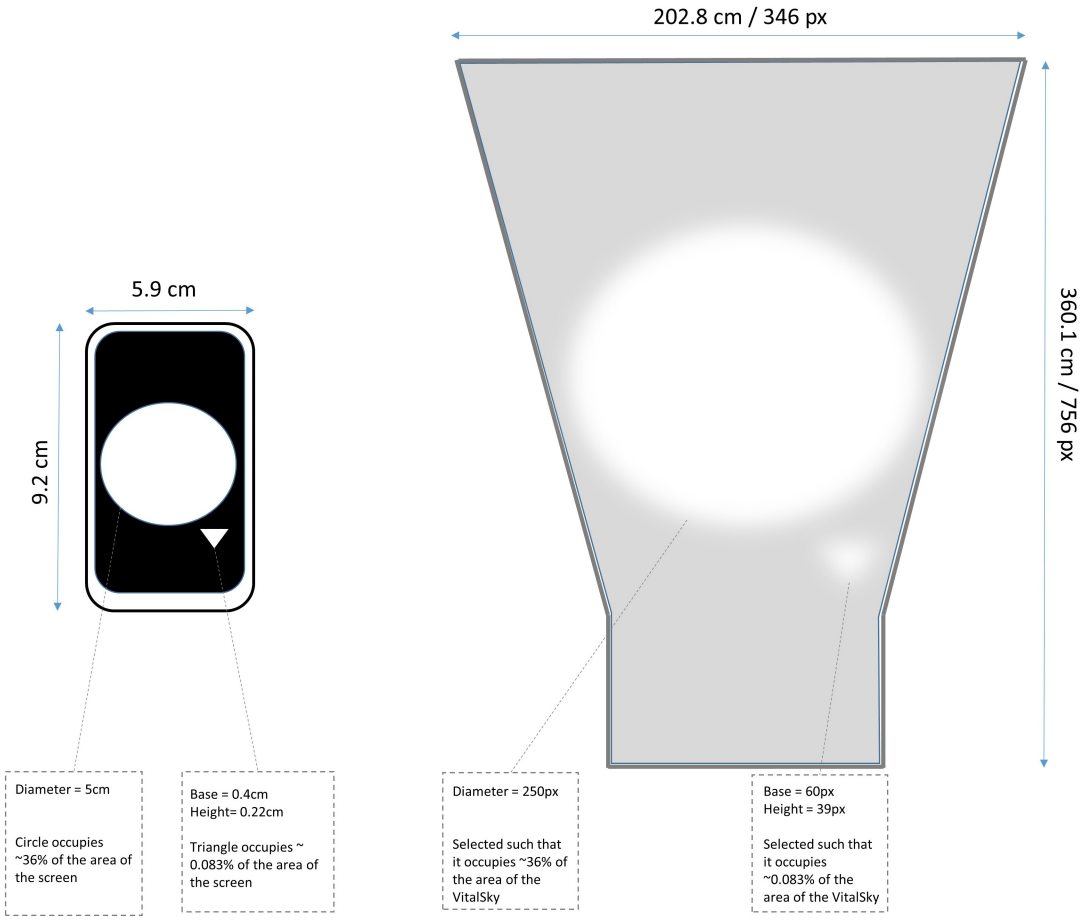


Figure 3.2: The scaled version tested on the VitalSky. The visual content was proportionally scaled to the size of the VitalSky. The content was observed to be discomforting when tested on the VitalSky.

3.2.2. Background color

It was clear that the black background might not be the best choice for the VitalSky attention test. If the visual content consists of black color, the VitalSky renders it by turning off the LEDs corresponding to the black region in the visual content. This creates a large illuminance difference on the blinking of circle. When the circle appears, the patients are exposed to bright light from the VitalSky whereas absolutely nothing is emitted by the VitalSky when the circle disappears. This creates a very discomforting experience for the patients.

As a result, different colors were tried for the background of the visual content. Due to the size of the VitalSky, the background color of the visual content affected the overall color of the lighting of the room. So, for any color that was used as the background of the VitalSky, the lighting of the room was influenced by the same color (see figure 3.3). This was not a good idea because the lighting of the room could influence the mood of the patients and may have undesired effects on the patient.

However, when using white color as the background the lighting in the room had a natural feeling. Using white background for the visual content reduced the change in lux or intensity of the light because of the consistent light from the background. Finally, it also provided nice contrast for almost every other color when used against a white background as the color for the target.

3.2.3. Shapes of target and distractors

In the DelApp-ICU test, a circle was used as the target stimulus. Different shapes were tried to observe if other shapes might have a better appearance on the VitalSky. This included shapes like square, rectangle, triangle, circle, star and hexagon. Due to the content being blurred and diffused when presented on the VitalSky, the

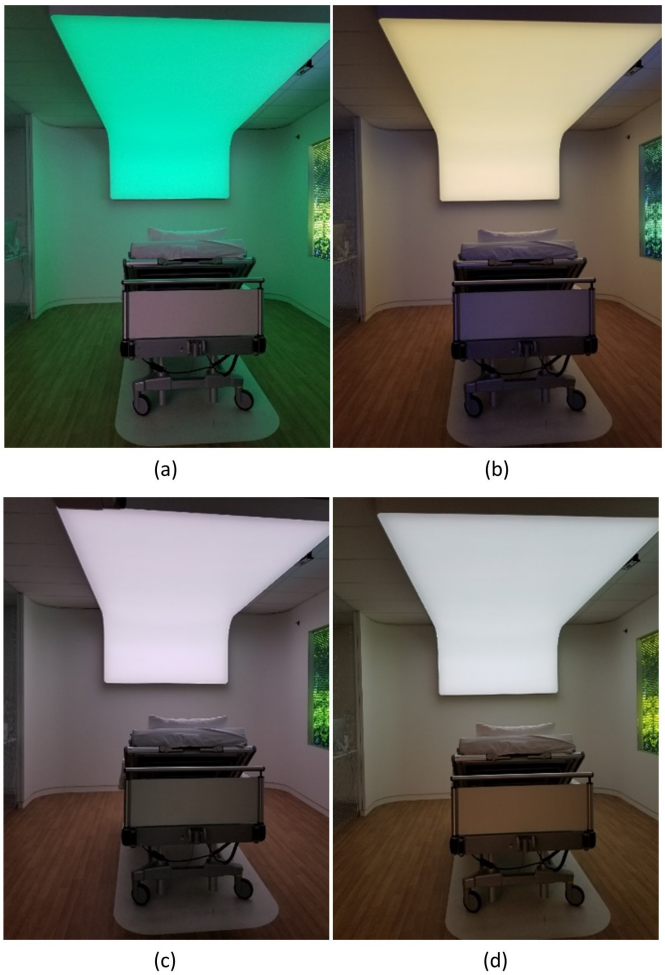


Figure 3.3: Effect of background color on the room environment. In figure 3.3(a) green was used as the background color and the room lighting also had the influence of green. Similar effect was seen for yellow and blue in (b) and (c) respectively. But, this was not in the case when white background was used and the lighting of the room felt natural.

shapes lost the detail and at many occasions were difficult to recognize and distinguish from each other. The edges of the shapes were diffused and this made the shapes difficult to identify.

The visual elements of the form of squares, hexagon and star were difficult to distinguish from a circle because the detail of the edges were lost. The rectangular shape was better distinguished but the rectangle needed to be sufficiently longer in one dimension to differ from a square or circle. Due to uniform circumference and absence of edges in circle, the diffusion spread uniformly and it made it easier to recognize a circle than other shapes. Hence, circle was chosen as the target for the visual content of the VitalSky (figure 3.4).

The distractor was kept as triangle because the other shapes when smaller were difficult to recognize and were confused with the circle. Thus, circle was used as target and triangles were used as distractors. This was also consistent with the target and distractors used in the DelApp-ICU.

3.2.4. Size of target and distractors

The size of the target and distractors was another important aspect while designing the visual content for the VitalSky. The size of the circle was varied from 50 px to 250 px and the diameter values tried for the circle were 50 px, 75 px, 100 px, 135 px, 150 px, 170 px, 210 px and 250 px (figure 3.5). After the trials, the circle of diameter 135 px and 150 px was considered to be the best for visibility and focus of the patients. The difference between the sizes of the 135px and 150px circles was very small. Five individuals were shown the two circles and asked for their preference and four out of five people chose the 150px circle. So, the circle with diameter of 150px

Figure 3.4: Appearance of different shapes on the VitalSky. The shapes of square and star were difficult to distinguish from the circle when the size of the shapes were similar but rectangle and triangles were distinguished from circle better than other shapes. The color of the shapes also had an effect and the shapes were more difficult to recognize when in yellow.

was selected for designing the visual content.

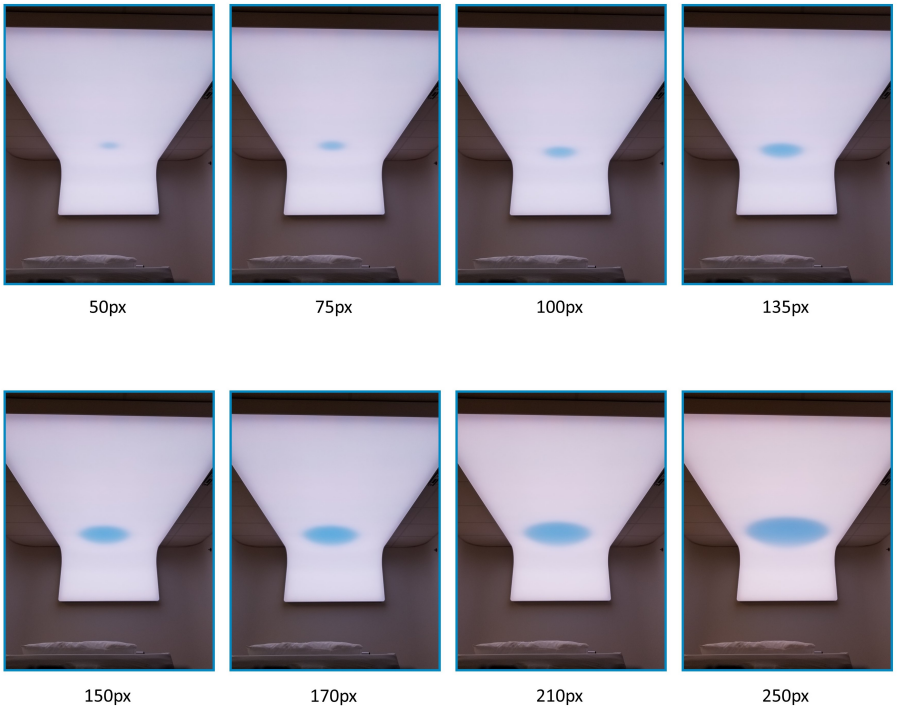


Figure 3.5: The visibility of the target circle on the VitalSky with different values of the diameter. The circles with diameter 135px and 150px were shortlisted. Later, the 150px circle was chosen for designing the visual content.

A Similar process was employed for the distractors. The distractors were downward pointing triangles appearing around the target or circle. The size of the distractor was varied until a size was reached which was considered to be easy to focus on and had good visibility. A feedback was taken from five individuals and all the people felt the triangles had good visibility and were easy to focus on. The triangles had a height of 39 px

and base of 60 px.

3.2.5. Positioning of the content and distractors

The positions of the target and distractors was a very important aspect of the design phase. As discussed earlier, even if the target and distractors are of the right size but not at the right position, the patients might not be able to properly see the target. So, to determine the appropriate position for the target and distractors different possible position combinations were tried and tested on the VitalSky.

The positions of the target was tested by developing visual content with the target at different positions on the VitalSky. The tested combinations of the visual content are illustrated in the figure 3.6.

Figure 3.6: Testing the position of the circle at different places in the VitalSky. In the left figure it can be seen that the content on the curve is much more blurred than the rest. The circle was placed below the curve on the figure on the right. The circle below the curve is still blurred but less than the circle at the curve.

The VitalSky has a curve which helps it to extend to the wall in front of the patient. This is a unique feature of the VitalSky and this provides a better experience for the patient. However, an observation made during the experimentation was that the content on the curve is blurrier than the content above or below the curve of the VitalSky. Hence, it was more difficult to recognize or focus on the content if the content was on the curve of the VitalSky.

The blurrier content on the curve was caused due to the fact that the distance between the LEDs and the diffuser of the VitalSky was greater on the curve than the distance between LEDs and the diffuser at the other areas of the VitalSky (figure 3.7). As the distance between the LEDs and the diffuser was greater at the curve, the light from the LEDs spread to a greater area before reaching the diffuser and were then diffused which made the content become blurrier when compared to the content from other areas of the VitalSky.

Another learning from this process was that it was very difficult for the patient to focus on any content in the top 50% area of the VitalSky. It required frequent head movements to see the content which could be really strenuous if the patient is intubated or not well. Hence, the target and also preferably the distractors must not be placed on the curve or top half of the VitalSky.

From the figure 3.8, it can be noticed that the best place for the circle was just above or below the curve, i.e, regions C and E. When the circle was in the region C or E, the circle was the easiest to focus on and was perfectly in the field of view for the patient. This also reduced the head movement required by the patient

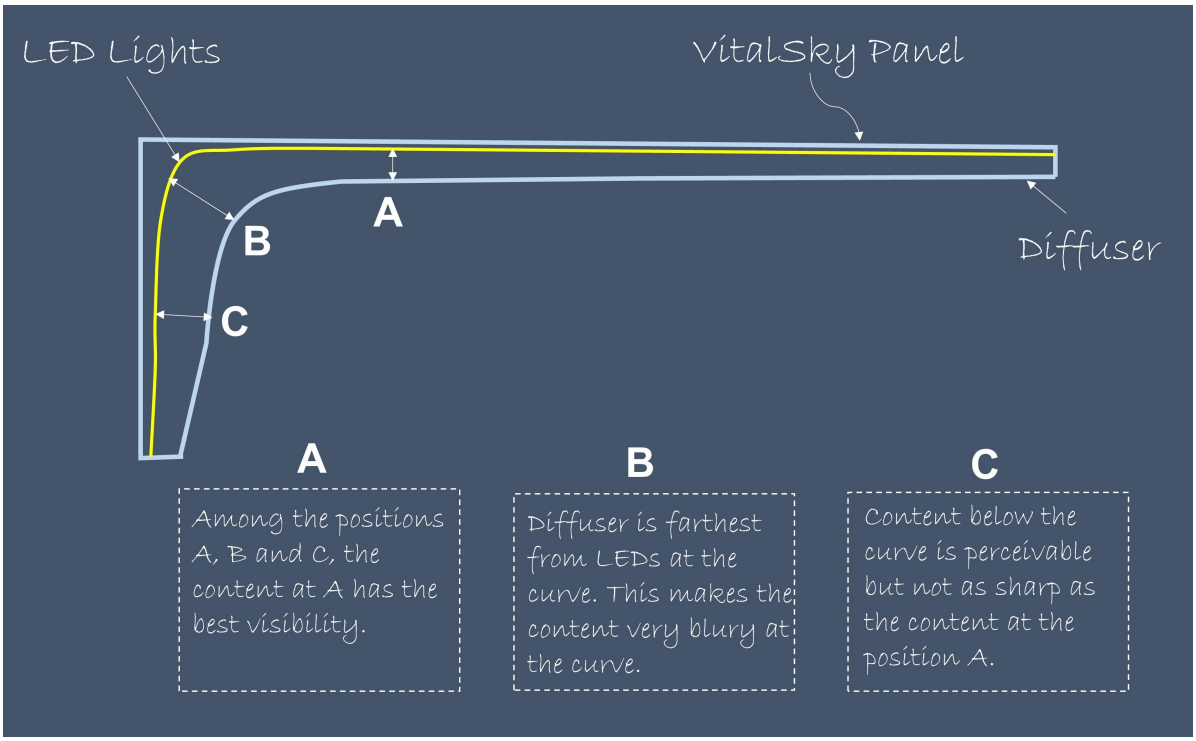


Figure 3.7: SideView cross-section of the VitalSky depicting that the distance between the LEDs and diffuser is largest at the curve.

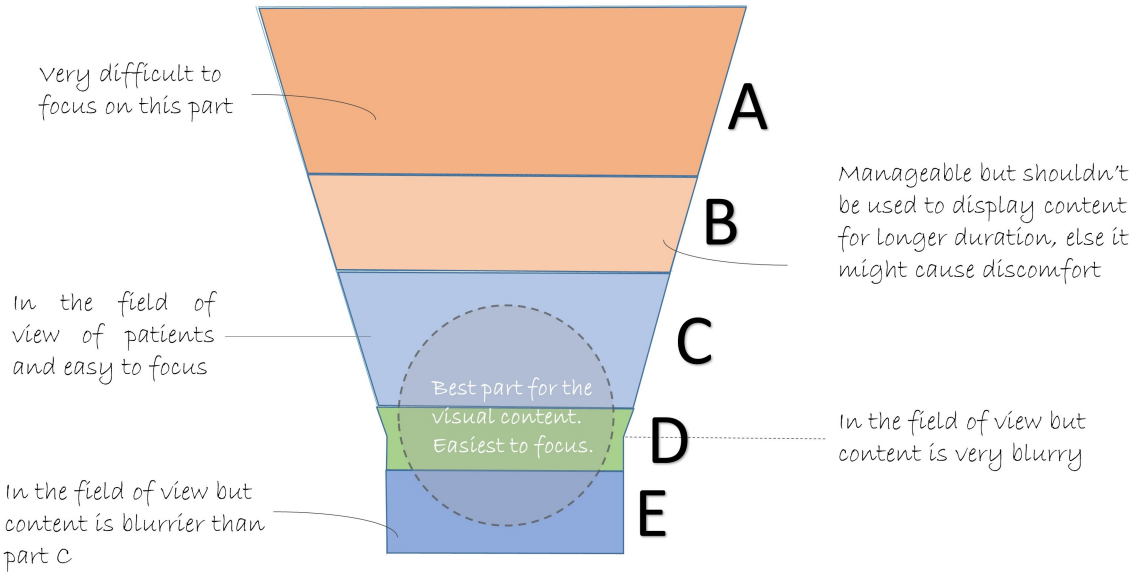


Figure 3.8: Segmentation of the VitalSky into different regions based on the field of view of the patient. The region A is very difficult to look at continuously and requires a lot of head movement. The region B was difficult as well and excessive use of region B for placing content would be discomforting for the patient. Regions C,D,E were easy to look at and were in the field of view of the patient. But, the content in region D was blurrier than the content in regions C and E.

to see the target. However, the position preferred was region C because the content in region C was sharper compared to the content in region E. Moreover, the curve was not present in all the variants of the VitalSky, so keeping the content in region C ensured that the content can be used in all the variants of the VitalSky.

To have a distracting effect on the patient, the distractors need to be in the field of view of the patient. As a result, the distractors cannot be placed in the upper half of the VitalSky as it is outside the field of view of the patient. The distractors are smaller than the target and putting it on the curve greatly reduces the visibility of the distractors. Thus, the only possible area of the VitalSky for the distractors was the region C (3.8) Since, the target is also being displayed in the same area, it was decided to place the distractors around the circle.

3.3. Alternative design exploration for the VitalSky attention test

Other design options were explored deviating from the DelApp-ICU. The different distractors and target that are used in these designs and how they are used is illustrated in figure 3.9 and figure 3.10. In all the design, circle was used as the target because of its uniformity and easy identification on the VitalSky when compared to other shapes. Below we will look into the various designs that were brainstormed in this thesis.

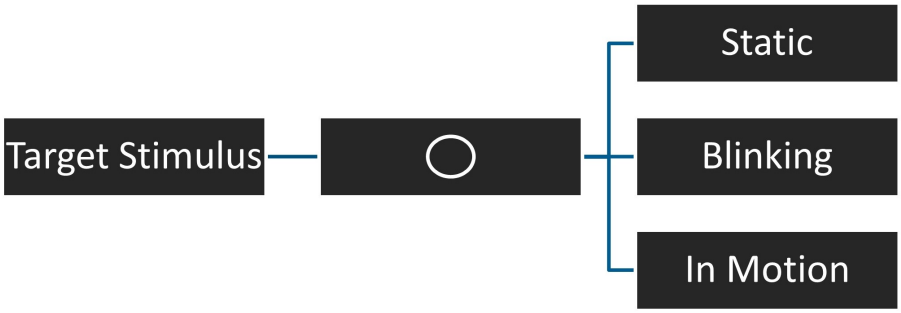


Figure 3.9: Circle being used as the only target in all the tests. The circle varies from being static to blinking or moving around the VitalSky.

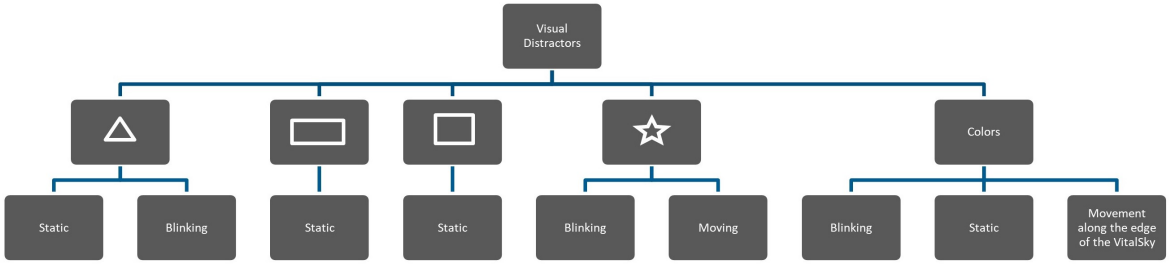


Figure 3.10: Different distractors used and the states in which they appear in the various test designs.

3.3.1. Colored edges

This design explored the use of colored edges as the distractors instead of the triangles. In the level 1 of the test, only a blinking circle was presented and the task was to count how many times the circle blinked. The difficulty was increased in level 2 by introducing the distractors. In level 2, a strip of different colors mixed together was used as a distractor on the two vertical sides of the VitalSky. In level 3, the vertical sides as well as the area below the curve of the VitalSky was used to display the distractors. The images for the three levels are shown in figure 3.11.

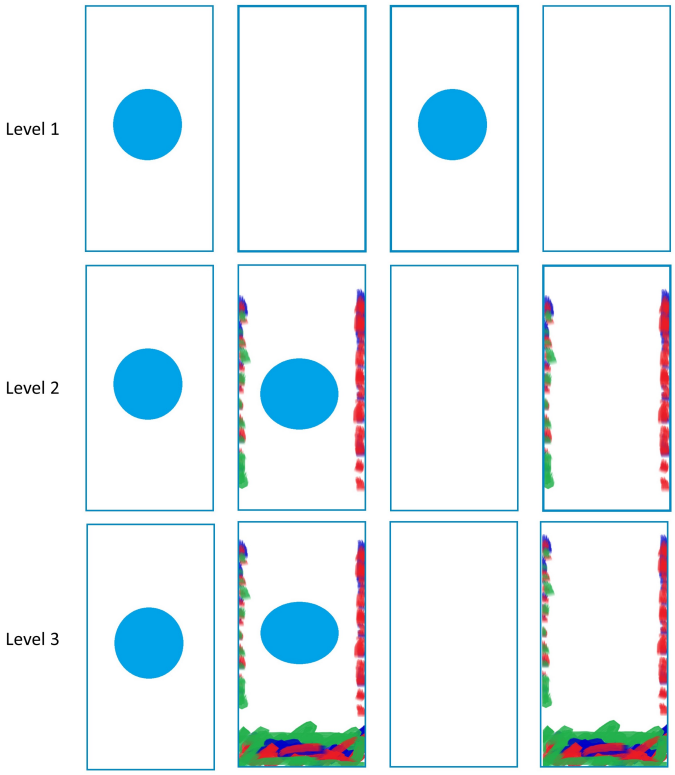


Figure 3.11: The images show the different instances of the levels (from left to right) of the colored edges test design. In level 1, the circle blinks with no distracting patterns. In level 2, colored patches appear as distractors on the edges of the VitalSky. In level 3, the colored patches also appear below the curve of the VitalSky.

3.3.2. Distracting star

The level 1 of this design is same as the previous design where the number of blinks of the circle needs to be counted without the distractors. In level 2, a star is introduced as distractor. This star moves around the VitalSky while the circle keeps blinking. The task was to count the number of blinks of the circle and ignore the distractor. The only difference in level 3 and level 2 of the test is the number of distractors. The number of moving stars increases from one to two in level 3. The three levels of the distracting star design are illustrated in figure 3.12.

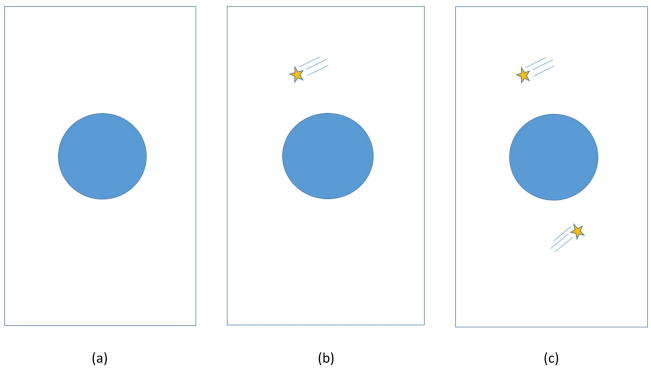


Figure 3.12: Three levels of the distracting star design. In level 1 (3.12 (a)), the participants only see the circle and count the number of blinks. In level 2 (3.13 (b)), a distracting start is introduced and the number of star increases to two (3.13 (c)) in level 3. The task for the participant remains the same to count the blinks of the circle while ignoring the distracting star.

3.3.3. Moving target

In level 1 of this design, blinking circles had to be counted and no distractors appeared. In level 2, the target moves while blinking, so the task was to focus on the moving target and count the number of blinks. No distractors appeared in level 2. In level 3, while the target is moving and blinking, more shapes randomly appear and disappear on the VitalSky as distractors. However, still the number of blinks of the moving circle needed to be counted while ignoring the distractors. The visual content for the three levels are illustrated in figure 3.13

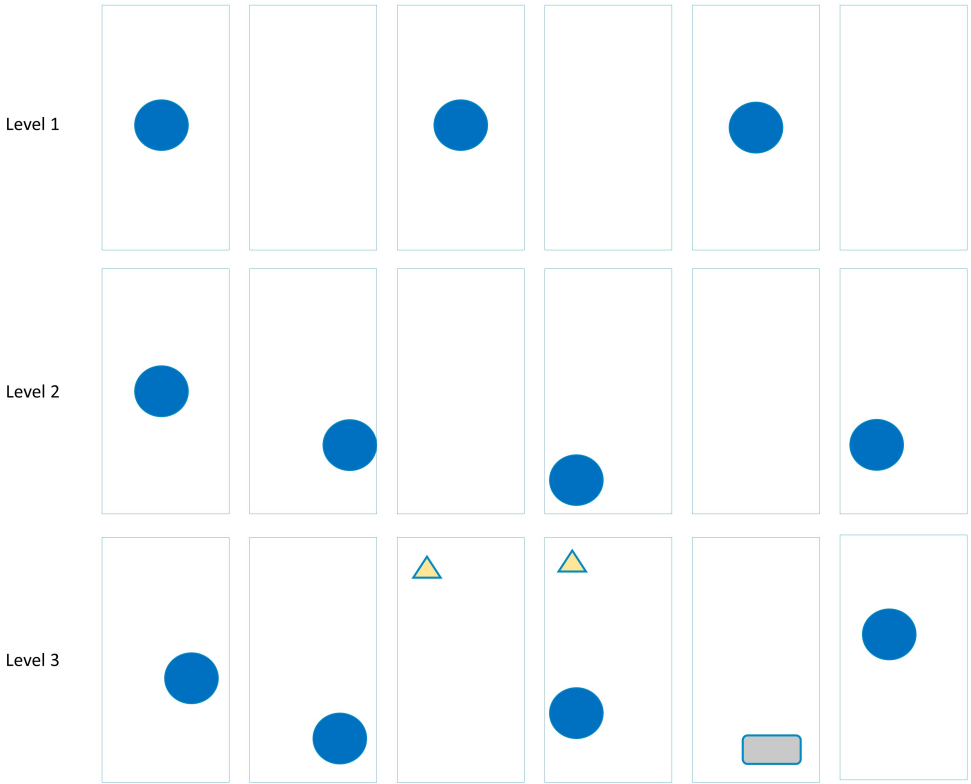


Figure 3.13: The images show the different instances of the levels (from left to right) of the moving target test design. In level 1, the circle blinks and the participant must count the number of blinks of the circle. In level 2, the circle starts to move around the VitalSky while blinking. In level 3, while the circle is moving and blinking, other shapes also appear as distractors. In all the levels, the participants are expected to count the number of blinks of the circle and ignore the distractors, if present.

3.3.4. Find the circle

In this test, level 1 included circle and other shapes like square, rectangle and triangle. All the shapes except the circle were used as distractors. The circle and the other shapes were presented in a static images and the participants must distinguish the circle from other shapes and count the number of circles present in the image.

In level 2, two circles were placed next to each other and the circles blinked at the same pace. Other shapes(rectangle, square and triangle) were introduced as distractors. These shapes were statically present and spread out across the entire VitalSky. The task was to count the number of times each of the two circle blinked. Level 3 was similar to level 2 with the only difference being that the circle may or may not be next to each other.

The blinking circles only appeared above or below the curve of the VitalSky so that it stayed in the field of view. The three levels of the test are illustrated in the figure 3.14

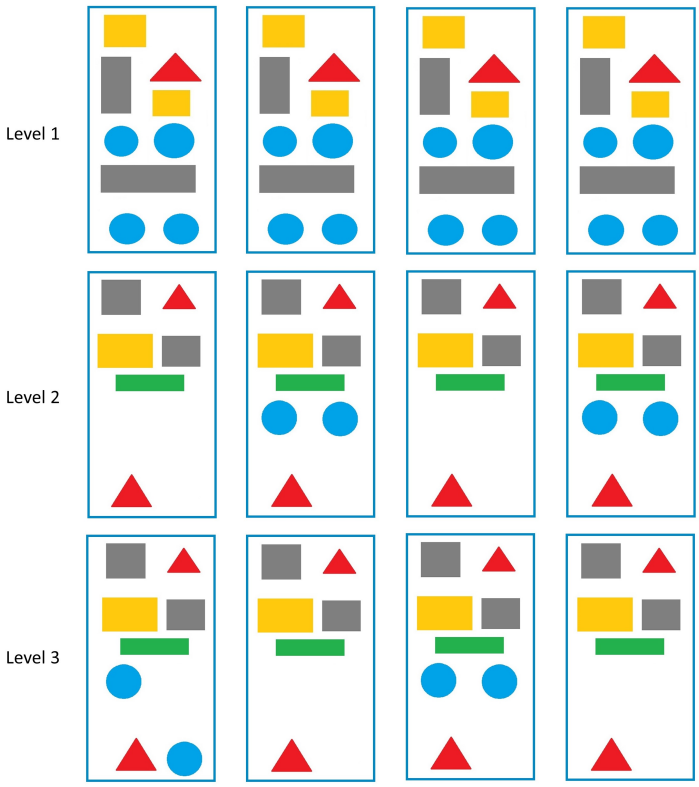


Figure 3.14: The images show the different instances of the levels (from left to right) of the test design. In level 1, a static image is displayed for a duration of 10 seconds and contains a lot of different geometric shapes. In level 2, two circles blink next to each other and the patients must count the number of blinks. The distracting shapes are static and always present. In level 3, the blinking circles may or may not be next to each other.

3.3.5. Shortlisting the design for the VitalSky

Considering the limitations for the duration of the study with the volunteers, it was possible to test only one of the four tests discussed earlier.So, among all of the explored designs only one design was shortlisted to be tested on the 10 healthy volunteers.

The "colored edge" test design was not selected because the blinking of colors caused high fluctuation in the lux levels of the VitalSky and produced a discomforting feel. The designs "distracting star" and "moving target" included movements of the visual content. Since, the static blinking of the target and distractors of the DelApp-ICU was already taxing for the patients, adding movements to them would make it too complicated for the patients. Thus, "moving target" and the "distracting star" designs were not considered for testing on volunteers. The "find the circle" design involved a lot of shapes and colors and was pleasant in appearance. Moreover, the absence of movement and no extreme fluctuations in the luminosity made the design preferable. Hence, out of all the test designs explored in this section "find the circle" design was selected to be tested on the healthy volunteers. A more elaborate description of the attention tests selected to be tested on the volunteers and their setup will be discussed in Chapter 5.

4

Exploration of automation of the test

Apart from the design and development of an attention test for the VitalSky, the secondary goal of this study was to automate this attention test so that it can be performed with minimal or no intervention of the nurse. In this chapter, we will look into different devices that were considered to be used for the automation of the attention tests on the VitalSky and their setup and technical challenges that were faced during the implementation of the entire setup. Lastly, we study the system architecture used to achieve this setup and the flow of data between the various devices of the setup.

4.1. Requirements for the response device

The VitalSky allows rendering of different visual contents but there did not exists any method for the patients to interact with the VitalSky. For the attention test to be automated, it was very important that the participants were able to interact with the VitalSky. This helps the participants to register their responses to the tasks of the attention test. The response devices for the automation of the attention test were selected keeping in mind the state of the patients in the ICU. The requirements put in place for the selection of the response devices were:

1. The response devices should be easy to use for the patients. The response device should be simple and must not be too complicated like requiring multiple authentications or .
2. The response devices must not be error prone. The devices being error prone will require patients to respond multiple times or can record incorrect data and both the cases can result in the patients being irritated.
3. The response devices should be non-invasive. Since, the delirium patients tend to remove and throw the measuring devices connected to their body, the response devices should be able to capture the response with minimal contact possible.
4. The devices should not increase the difficulty of the attention test. The method of responding with the device should not be more difficult than the attention test itself or in no way should increase the difficulty of the attention test.

4.2. Interaction devices exploration

A wide range of devices were explored to facilitate the interaction between the VitalSky and the patients. The devices were considered in the verbal, tactile and gestures domain. The verbal domain devices would be useful if the patient can speak freely and is not intubated whereas considering the scenario where the patient is intubated the devices in tactile and gestures domain would be more useful. The summary of devices considered for the pilot study are shown in figure 4.1

1. **Verbal domain**

   (a) **Respeaker:** Respeaker is a hardware device which is capable of capturing and producing audio signals. It consists of microphones that can capture the audio signals and it can also produce audio signals if a speaker is connected to it. This device was chosen because it opens up the possibility of participants responding verbally to the tasks of the attention test as they do normally in daily conversations.

2. **Tactile domain**

   (a) **Tablet:** A tablet is very commonly used in the medical applications these days. It was chosen as a response device because most of the people are familiar with its usage and it allows for customizations that can be done to the device according to the requirement of the application.

   (b) **Bluetooth Keypad:**Bluetooth keypad was a small numeric keypad that providing the normal numeric operations supported by a general calculator. The device was simple, intuitive and small enough to easily fit in the palm of the participant.

   (c) **Flic button:** Flic button is a small Bluetooth button that provides three option to respond: tap or click, double tap and hold. This device was chosen keeping in mind the simplicity of the device, its ease of use and small size.

   (d) **Squegg:** Squegg is a smart squeeze ball that measures the grip strength of the user. It measures the force that is applied to the device by the user. The Squegg was not programmable and the software state of the device only allowed force output on the Squegg owned Android application. Thus, it was not possible to integrate the device with the VitalSky.

3. **Gesture domain**

   (a) **Leap Motion:** Leap Motion is a hardware device that supports motions of fingers and hands as input. The device consists of an Infrared (IR) camera that captures the images of hands and provides inbuilt APIs to track the movement of the hands. The device was chosen considering in mind the intubated patients who might not be able to speak and hence can respond using gestures.

   (b) **Camera:** A camera was considered during the initial exploration but was found to be redundant with the Leap Motion as the same tasks of recognizing gestures could easily be performed with the Leap Motion. Hence, camera was not considered further for the response devices selection.

   (c) **Myo Band:** Myo Band uses Electromayogram (EMG) signals to measure electrical impulses from the muscles and could have been used to detect gestures from the patients but the device is not being manufactured anymore.

   (d) **Kine Kiso:** Kine measuring devices are small electrodes that measure EMG signals from the body. It provided a wireless method for measuring the muscle movements of the participants but the electrodes must be attached to the body parts of the patient. Since, the study were exploring non-invasive methods of collecting response from the patients, the kine measuring devices were not considered further.

   (e) **GeneActive accelerometers:** GeneActive accelerometers can be used as a watch or attached to the body and it tracks the movement of the device along the three axes. The implementation for the collection of real-time data from the GeneActiv watches could not be completed in the required time frame and also the device was not considered to be non-invasive, so the devices were not used in the pilot study with the volunteers.

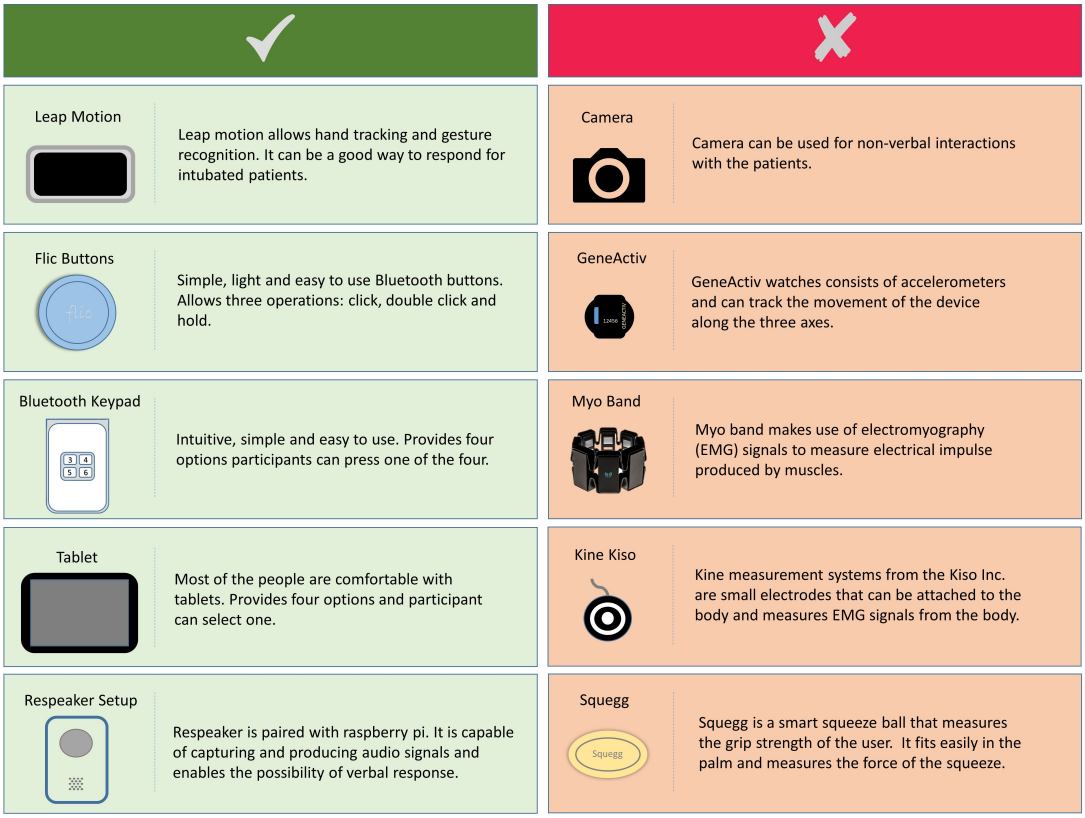


Figure 4.1: The list of devices that were considered to be used as a response device in the ICU. The devices were considered keeping in mind the ease of use of the device and state of the patient in the ICU.

4.3. System setup

4.3.1. Respeaker

With the help of respeaker, the participant could verbally give their response and it is processed using the respeaker. Snips platform is an online platform that allows to create customizable speech recognition systems which can then be deployed locally. Snips platform was used with the respeaker to train the system to recognise and process the voice input from the participant. The snips platform enabled the system to recognize the specific phrases or words and what action to be performed when they are detected.

Figure 4.2: The respeaker and Raspberry Pi setup. The setup was enclosed in a 3D designed box to prevent contact with the components and mishandling and damage of the components.

To respond using the respeaker, the participant had to first speak a keyword called "Hey Snips", which activates the respeaker, and then the participants can say their response. The voice sample is processed by the respeaker and at the end a numeric response is obtained. This response like all the previous responses is sent to the node server and then saved into a response log by the Raspberry Pi.

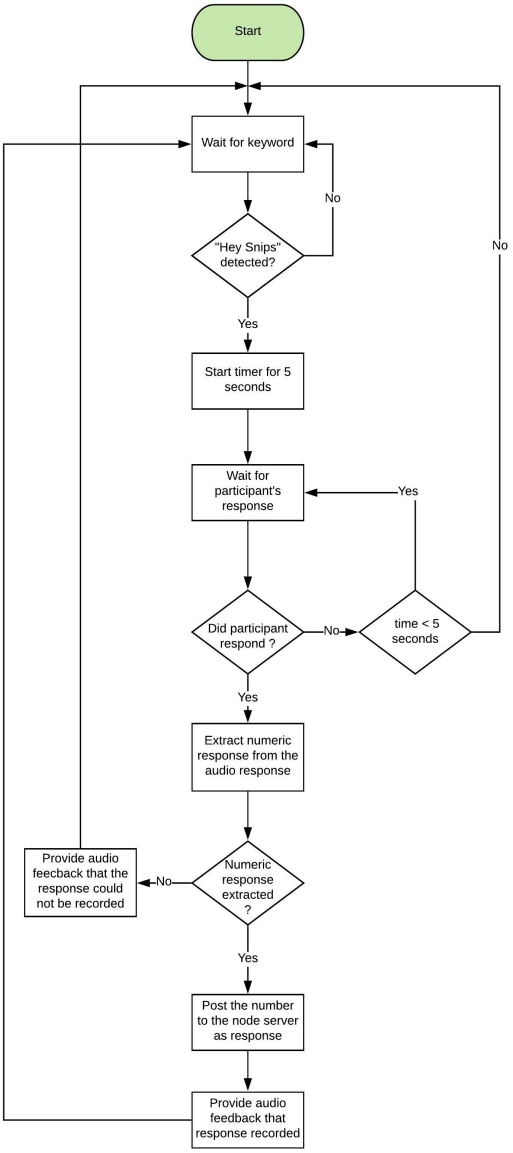


Figure 4.3: The algorithm and working of the respeaker. The respeaker employs snips platform to process the audio samples and convert them into a text response which was then used as the participant's response. The audio samples were not saved, only the text responses were saved.

4.3.2. Tablet

A Samsung Galaxy Tab S4 tablet was used to present four options to the participant and the participant can choose the correct response by tapping on the response shown in tablet. The response screen is shown in figure 4.4.

Figure 4.4: The response screen used for the Tablet. The participant can respond by simply tapping on the response he/she thinks is the correct response.

The four response choices are put together on a webpage which was developed using React.js and the webpage was hosted on a laptop over the local network. When the user tapped on any of the option of the webpage, the webpage displayed a pop-up to confirm the response. The response was then recorded and sent to the node server. This response was then read by the Raspberry Pi and stored into the response log.

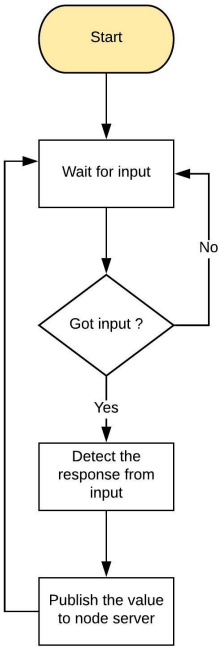


Figure 4.5: The flowchart explaining the working algorithm of Tablet and Bluetooth keypad.

4.3.3. Bluetooth Keypad

In this method, a Bluetooth based numeric keypad was taken and modified to present only four buttons to the participant and rest of the buttons were hidden using a 3D printed case (figure 4.6). Also some of the buttons were altered so that the response options can be presented to the participants in the center of the keypad. The participant can press one of the options which they consider to be the right response.

The Bluetooth keypad was controlled using a Python script which read the key pressed by the participant and forwarded the key press to the node server. The participants got an audio feedback indicating that their response is recorded and stating their response.

Figure 4.6: The original Bluetooth keypad (left) and the modified version of the Bluetooth keypad (right) used in the study. The keypad was modified such that it only allows relevant operation and options for the participant and all the other keys were hidden.

4.3.4. Flic buttons

Flic buttons are connected to the Raspberry Pi using Bluetooth and allowed the participants to tap it to register their response. The participant responded by pressing the flic button as many times as their response. For example, if the participant's response was five, he or she presses the flic button five times. After pressing the flic button, the response was shown as a visual feedback on the VitalSky. This visual feedback represented the number of times the participant pressed the flic button with the same number of circles. An illustration of this is shown in figure 4.7.

Figure 4.7: The visual feedback provided to the participants on the VitalSky when they pressed the flic button. The number of circles on the screen are a representation of the number of times the participant pressed the flic button.

A Python script was used to program the flic button. Every time the flic button was pressed a timer of 5 seconds was initiated and a count variable was incremented. If the flic button was not pressed again within the 5 seconds duration, the value of the count variable was considered as the participant's response. If a participant pressed the button again within the 5 seconds interval the timer was restarted. The value of 5 seconds was chosen so that the participants had sufficient time to look at the visual feedback and change their response if required. This response was then sent to the node server from where it was read by the Raspberry Pi. The algorithm for the flic buttons is shown in figure 4.8.

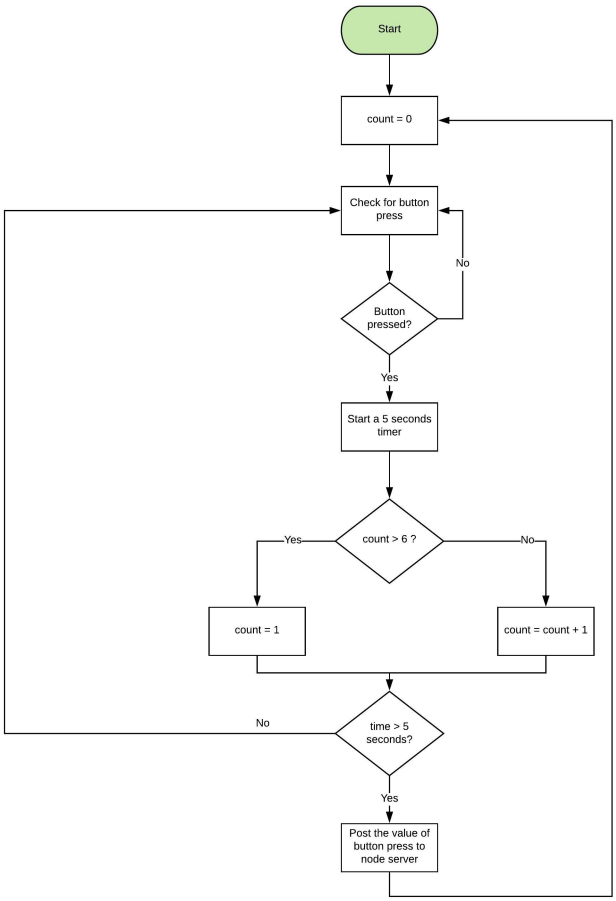


Figure 4.8: The flowchart explains the working of the flic buttons and the algorithm used for it.

4.3.5. Leap Motion

The Leap Motion processed the hand gestures of the participant and converted it into a numeric response. This numeric response was then sent to the node server using a HTTP POST request. The leap motion detected the number of fingers of the participants that are extended and processed that number as the participant's response. For example, for stating three as response, the participant had to keep three fingers above the Leap Motion. The response method for six was chosen as fist because otherwise the participants would have to use two hands and it would have been inconvenient. The gestures are shown in figure 4.9

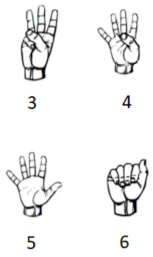


Figure 4.9: The gestures which the participants can use to respond non-verbally using the leap motion.

The participants were asked to make the gesture and hold it above the Leap Motion until they hear a beep (2-3 seconds), which indicated that their response is recorded. The leap motion continuously records the response and only produces a beep if the previous three consecutive responses were the same. The algorithm for the leap motion is shown in figure 4.10.

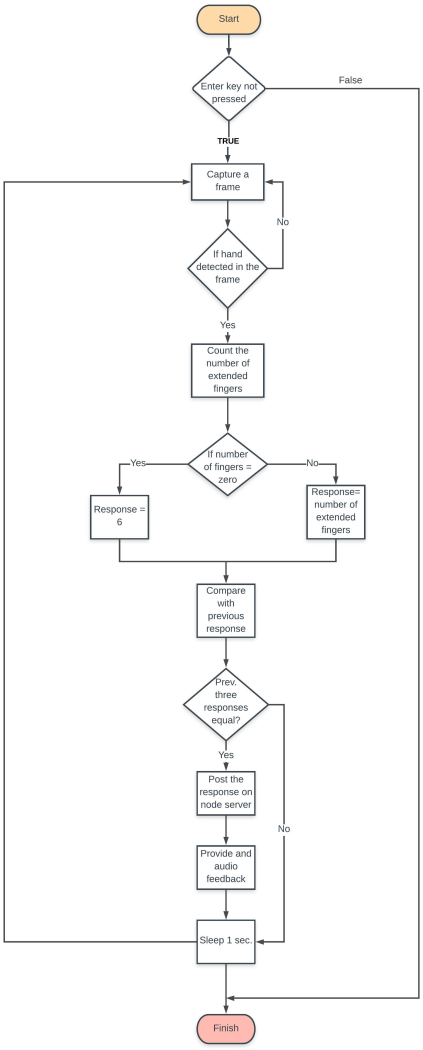


Figure 4.10: The flowchart describes the working of the leap motion control script. This algorithm detects and counts the number of the fingers of the participant and convert that into a numeric response.

4.4. Data collection and data flow

Apart from the devices, the main components of the automated setup were

1. **Raspberry Pi:**Raspberry Pi was used to control the visual content being played on the VitalSky, integrate different response devices with the VitalSky and to gather and process the data from the response devices. All the response devices sent the response data to a node server. The data from the server was gathered by the Raspberry Pi. The Raspberry Pi then saved this response in a file which contained the responses from all the devices and levels of the test. The Raspberry Pi also controlled the visual content of the VitalSky in the automated version of the test. The Raspberry Pi was programmed to play the video in a pre-defined sequence which was based on the participant number and was randomized for each participant. The Raspberry Pi sent a HTTP POST request to the VitalSky server which instructed the VitalSky to play the video mentioned in the request.
2. **VitalSky server**: The VitalSky server provides access to different functions possible with the VitalSky like controlling the brightness of the VitalSky, uploading videos to the VitalSky, starting and stopping a particular video. All the visual content used for the attention tests was uploaded to the VitalSky server and these visual contents were then played by the sending a HTTP POST request to the VitalSky server.
3. **Node server**: The node server was created to host the data from the different response devices. The

server had a dedicated URL for each of the device which was used to store and read the response from the response devices. The scripts of each of the devices published the response from the participant to this server using a HTTP POST request. These responses were then read by the Raspberry Pi using a HTTP GET request and saved into a log file.

4. **Webpage server**: The webpage server was used to host the webpage which was used by the participants to respond using the tablet. The server was hosted on a laptop. Whenever the participants tapped any of the response on the webpage, the response was sent to the node server.

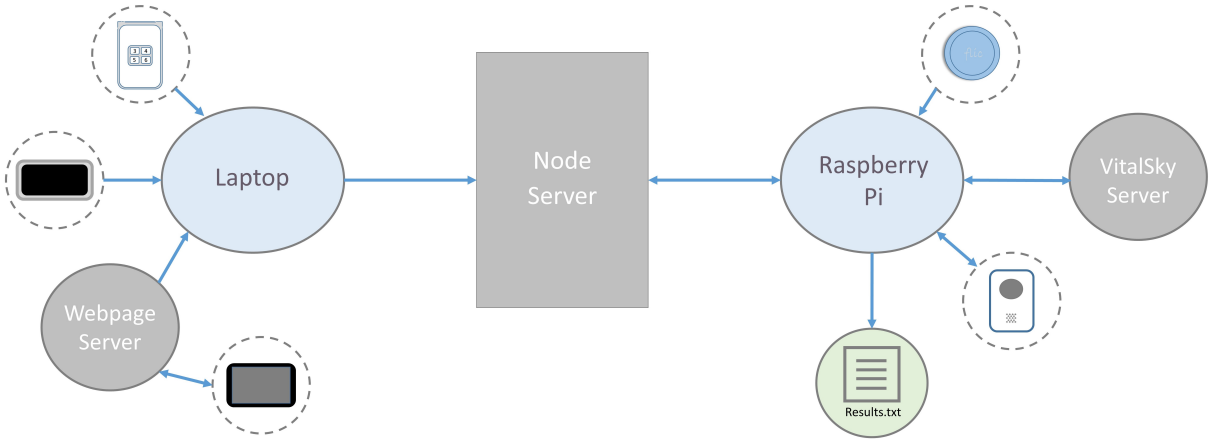


Figure 4.11: The figure presents the components of the automated setup and the data flow between them.The response devices are displayed in dotted circles. All the data from the devices was routed to the node server. The Raspberry Pi then read the data from the node server and stored it into a local results log file.

Before the test was started, all the videos were uploaded on the VitalSky. Next, the scripts for webpage and node server were executed and both of these servers were hosted on a laptop. The scripts for the Bluetooth keypad and leap motion were activated from the laptop whereas the scripts for the flic buttons and the re-speaker were activated from the Raspberry Pi. The webpage was displayed using the tablet by entering the webpage URL generated by the webpage server. The responses from all of these devices were sent to the node server from their respective scripts by using the POST request. The data from the node server was pulled by the Raspberry Pi and compiled into a single response log file. This file was later analyzed to evaluate the performance of the participant in the test.

5

Methodology

This chapter describes the structure and composition of the attention tests that were developed in this thesis. The chapter starts with the description of the test followed by the illustrations of the visual contents of the attention tests developed for the VitalSky. The chapter ends with the explanation of the study setup and protocol.

5.1. Development of the attention tests for the VitalSky

5.1.1. DelSky

DelSky attention test was similar to the DelApp-ICU but this test was designed with the help of learnings obtained in Chapter 3. The DelSky made use of colors for the visual elements to reduce the fluctuations in the luminosity. The structure of DelSky was similar to the DelApp-ICU. It consisted of three levels with three trials in each level. The levels of the Delsky are illustrated in figure 5.1.

- **Level 1:** The participants saw a blue blinking circle against a white background and the participant should count the number of blinks.
- **Level 2:** The green downward pointing triangles were introduced as distractors and the number of distractors ranged between 1 and 4.
- **Level 3:** The number of distractors increased proceeding into level 3 and five to eight distractors were visible to the participant.

5.1.2. DelSky-Smart

DelSky-Smart is the proof of concept of an automated test that can be used to autonomously perform the attention test without or with minimal intervention of a nurse. The visual content of the test are exactly similar to the DelSky test. DelSky-Smart included various devices that were discussed in Chapter 4 for the patients to interact with the VitalSky and respond using the device. The idea of this test was to understand the technical feasibility and usability of the automated setup and the user preference for the response device.

In DelSky-Smart, the participants had two levels instead of three and each level had two trials. The levels 1 and 3 of the DelSky were selected to be used in DelSky-Smart. The participants repeated this for each device. Thus, with five devices and four visual content for each device, the participants saw a total of 20 videos in DelSky-Smart.

The participants were asked to respond with each of the response device one at a time. The participant had to count the number of blinks of the circle and then respond by using the provided device.

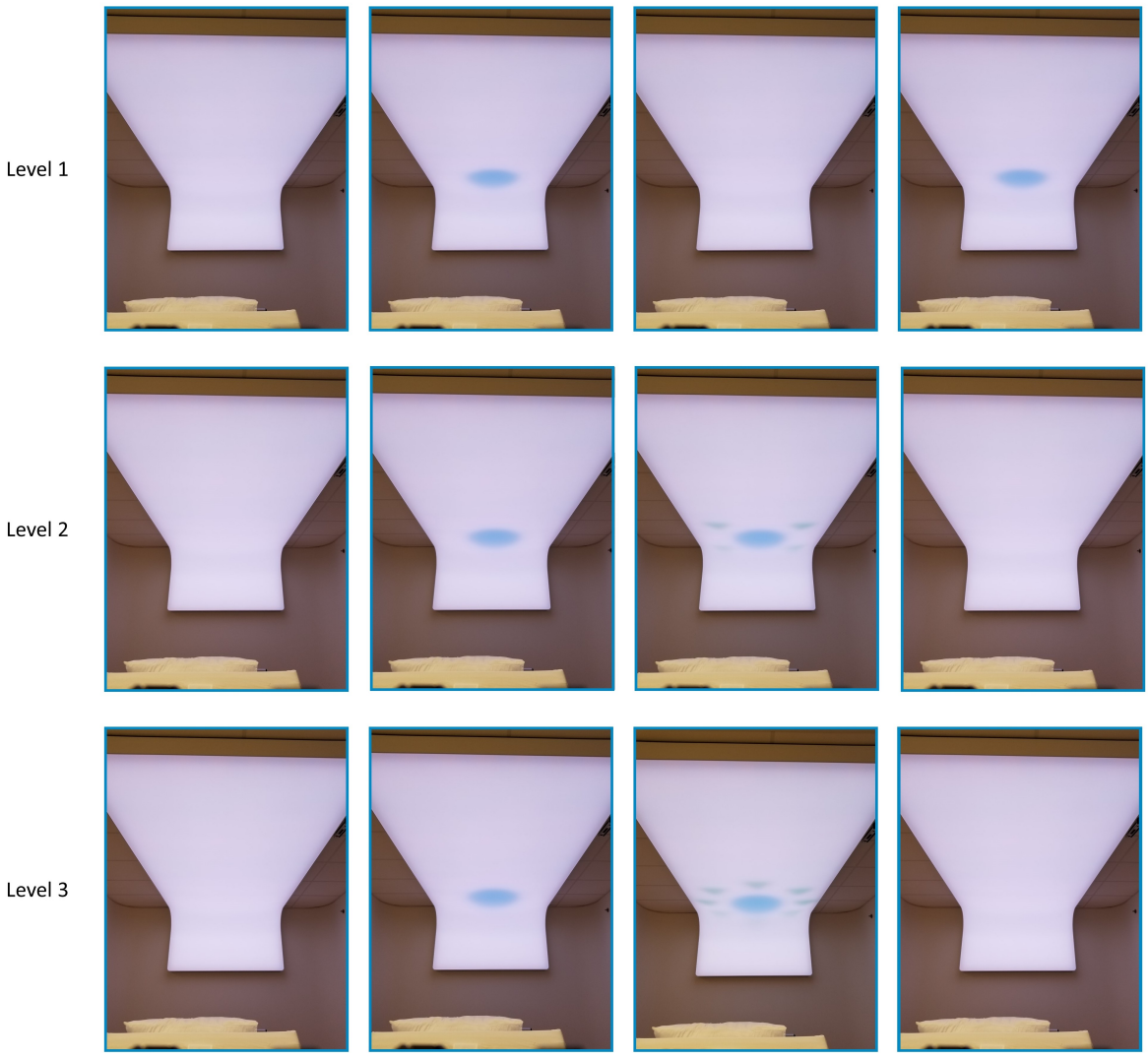


Figure 5.1: The images show the different instances of the levels (from left to right) of the DelSky. In level 1, the participant see a blue blinking circle against a white background. In level 2, along with the circle green triangular distractors (one to four in quantity) are shown to the participant. In level 3, the number of distractors are increased (five to eight). In all the levels the participant must count the number of blinks of the circle and ignore distractors, if visible.

5.1.3. DelApp for VitalSky (DAViS)

The DAViS is a translation of the DelApp-ICU for the VitalSky with similar design of visual content and color scheme. The main difference is the display, instead of smartphone the test is performed on the VitalSky. All the other characteristics like color (black background and white target stimuli and distractors), visual elements used in the levels, the interstimulus time in the trials and the time that the target stimulus and distractors are displayed for are similar to the DelApp-ICU (See figure 3.1 in Chapter 3). The visual elements were scaled according to the VitalSky for better appearance on the VitalSky. The levels of the DAViS are illustrated in the figure 5.2.

- **Level 1:** Consisted of just white blinking circle against the black background and the participants must count the number of blinks of the circle.
- **Level 2:** The small downward pointing distracting triangles were introduced as distractors in level 2. The number of distractors vary between one and four. The participants should count the number of circle blinks while ignoring the distractors.
- **Level 3:** In level 3, the number of distractors increased and varied between five to eight. The task for

the participants remained the same.

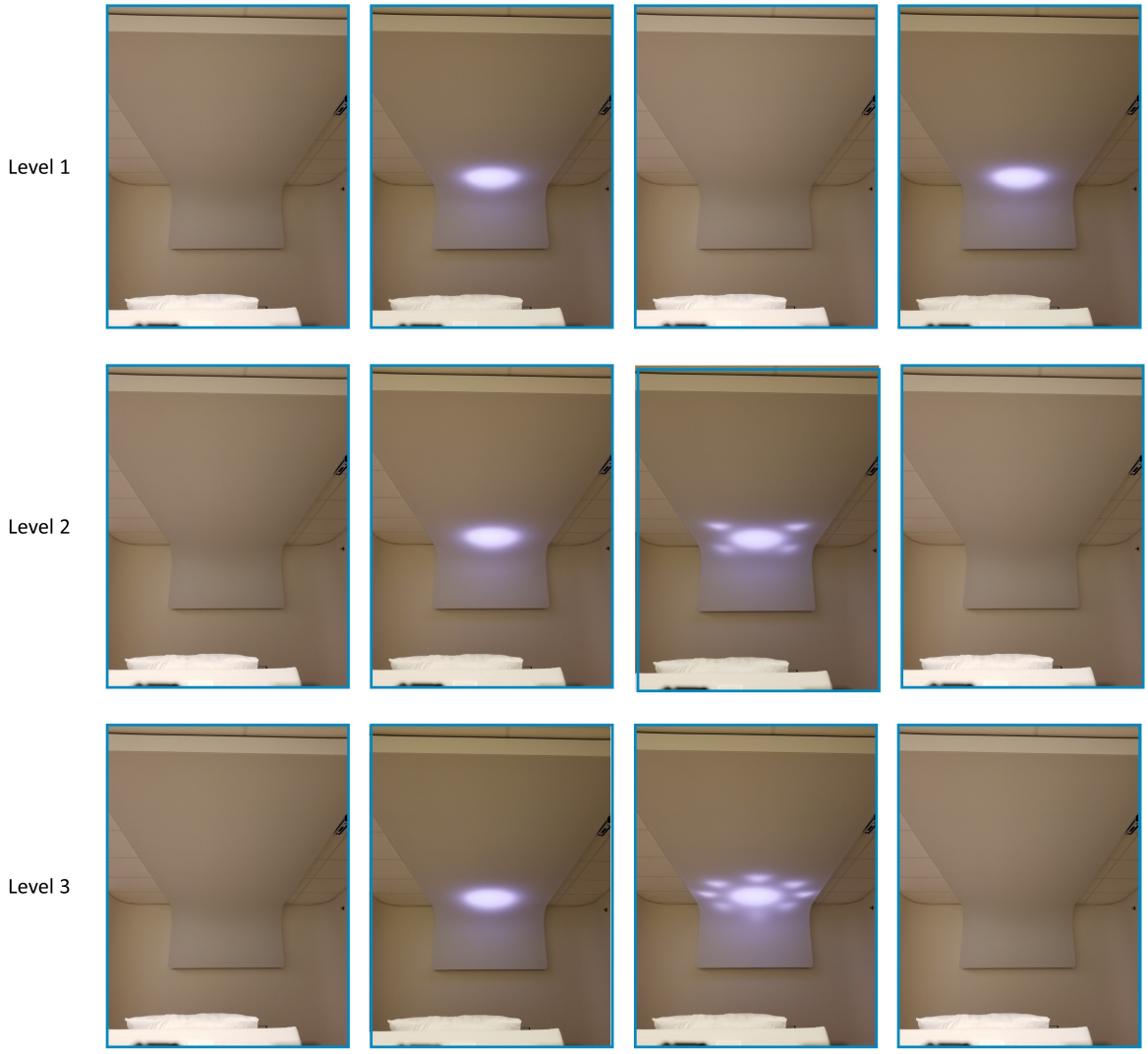


Figure 5.2: The images show the different instances of the levels (from left to right) of the DAViS test. In level 1, a blinking circle is shown to the participants and they need to count the number of blinks. In level 2, distractors are introduced in the form of small downward pointing triangles. The distractors appear around the circle and the participant must ignore the distractors while counting the blinks of the circle. In level 3, the number of distractors increases.

5.1.4. VitalSky Delirium Test (VSDT)

VSDT was the name given to the "find the circle" attention test design explored in Chapter 3. It is different from the four other attention tests in terms of the visual content. The VSDT was designed to explore different possibilities of the visual content with the VitalSky. It consists of static or blinking circles as target stimulus and static geometric shapes as distractors. The distractors remain static in all the levels.The VSDT consists of three levels and each level consists of three trials and one visual content is displayed in each trial. The visual content from each level is depicted in figure 5.3.

- **Level 1:** The participants see a static image with lots of geometric shapes and they must identify and count the number of circles in the image. The participants see a total of three static images (one in each trial) in this level and each image is shown for a duration of 10 seconds.
- **Level 2:** The participants are presented with two circles blinking next to each other and the participants should count the number of blinks of both the circles while ignoring the static distractors. The only

difference between the three trials of the level 2 is that the interstimulus time of the circle is 1000ms in trial 1, 2000ms in trial 2 and 3000ms in trial 3.

- **Level 3:** Level 3 is similar to level 2 but the circle may or may not be next to each other and the participants count the number of blinks of each circle.

Figure 5.3: The images show the different instances of the levels (from left to right) of the VSDT. In level 1, a static image is displayed for a duration of 10 seconds and contains a lot of different geometric shapes. In level 2, two circles blink next to each other and the patients must count the number of blinks. The distracting shapes are static and always present. In level 3, the blinking circles may or may not be next to each other.

5.2. Scores for the tests

All the test, except DelSky-Smart, followed 9-point scoring system. The participants were asked to count the number of blinks of the circle or number of circle (in level 1 of VSDT) in each trial. At the end of each trial, the participant responds with the number they counted. If the participant's response is correct, they get one point otherwise they get zero points. Since all the tests (except DelSky-Smart) had nine trials, the possible range of scores for each test varied from zero to nine.

The DelSky-Smart test was not scored. This is because the objective of the DelSky-Smart was to assess the technical feasibility of an automated setup and the user preference for the response device. However, the

responses of the participant in DelSky-Smart were recorded to evaluate the correctness of the response and usability of the device.

5.3. Study Setup

There were certain hypothesis that were formulated before conducting the pilot and study and these hypothesis were tested by the results obtained in the study:

1. The performance of the participant will be relatable in DelApp-ICU, DelSky and DAViS.
2. The visual content of the DAViS will be perceived as bright and uneasy by the participants due to large fluctuations in luminosity.
3. The participants will prefer the verbal response with respeaker method over others as it feels more natural.
4. The participants will agree with the design choices like size and positioning of the visual content.
5. The color scheme of the DelSky will be preferred by the participants over DAViS.

For the pilot study, 10 healthy volunteers were recruited. All the participants were Philips Research employees. The inclusion criteria was that the participants should be 18 years or older . The exclusion criteria for the study included auditory or visual impairments, known history of epilepsy, migraine or any cognitive disorders and unable to read, speak or understand English. All the participants provided a written informed consent about their participation in the study.

The pilot study was performed in a mock-up ICU room at the Philips Research (see Figure 5.4). The setup of the study is shown in figure 5.5. The study consisted of five attention tests. The DelApp-ICU was performed using a smartphone, rest of the tests were performed on the VitalSky. For the DAViS, DelSky and VSDT, the participant's response was manually collected by the experimenter entered in a laptop.

Figure 5.4: The VitalSky lab is a mock-up ICU room with the VitalSky attached on the ceiling. The lab was used for all the experiments performed in this study. The lab is located in the Philips Research facility.

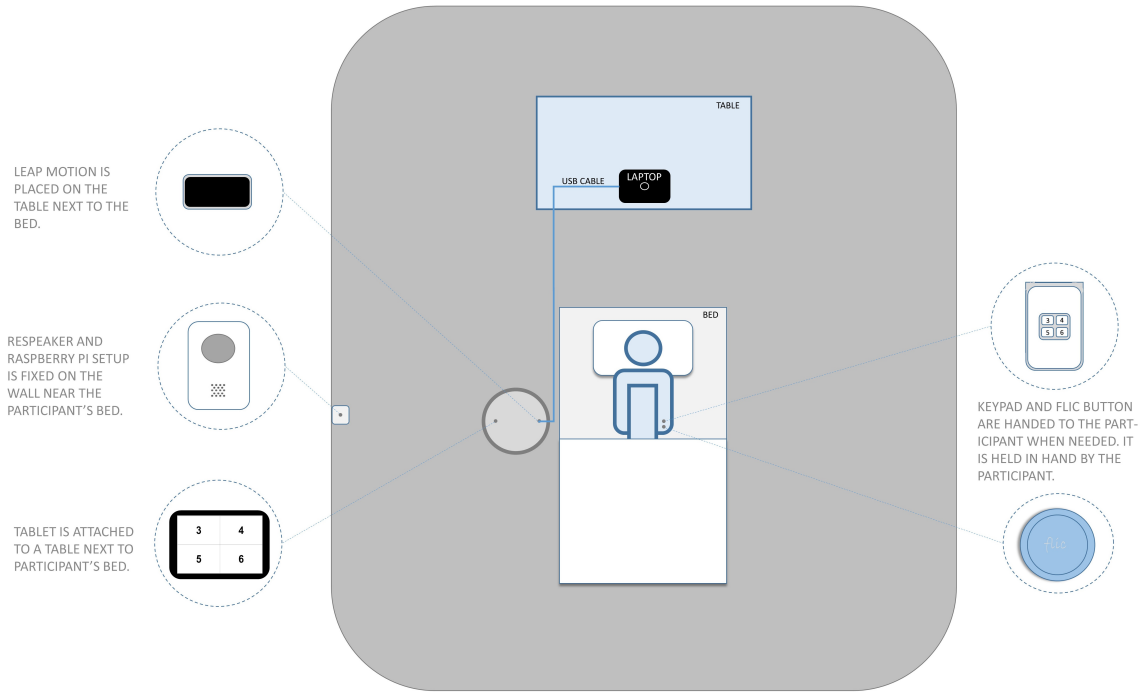


Figure 5.5: The setup for the pilot study and the positions of the different devices. The devices were positioned in a way that the participant had to do least effort to respond with the device.

In case of DelApp-ICU, the participants entered their response in a smartphone. After each trial, the participant got a popup in the smartphone to enter the response. Upon completion of the test, the smartphone displayed a message screen indicating the completion of the test. All the responses were stored in the smartphone and were retrieved after the experiment.

In DelSky-Smart, a variety of devices were used which allow the participant to respond without the need of the experimenter (see Chapter 4). The respeaker with Raspberry Pi was attached to the wall on the right of the bed whereas the tablet and leap motion are placed on a table next to the bed. The tablet is attached to the table using a tablet holder to adjust the tablet to a comfortable position for the participant. The flic button and the Bluetooth keypad were held in hand by the participant and were provided only when the participant needed to use them.

5.4. Protocol

The study started by the experimenter welcoming the participant and briefing the participant about the study protocol. The experimenter explained about the different tests, the tasks for the tests and gave a demo of how to respond using different response devices. The participant was then asked to sign an informed consent form.

The total duration of the study was 1 hour and 15 minutes during which the participant undergoes five sustained attention tests. Each test was approximately 5 minutes long, except DelSky-Smart which lasted approximately 20 minutes (see figure 5.6). At the end of each test, the participant was asked for a subjective feedback on the visual content of the test. The subjective feedback was collected by means of a short questionnaire including questions that range from the color scheme of the test to rating the tests against each other (feedback questionnaire attached as supplementary material). In DelSky-Smart, the subjective feedback was collected after the usage of each response device. At the end, when the participant have completed all the tests, they were asked to compare and rate the different tests and devices.

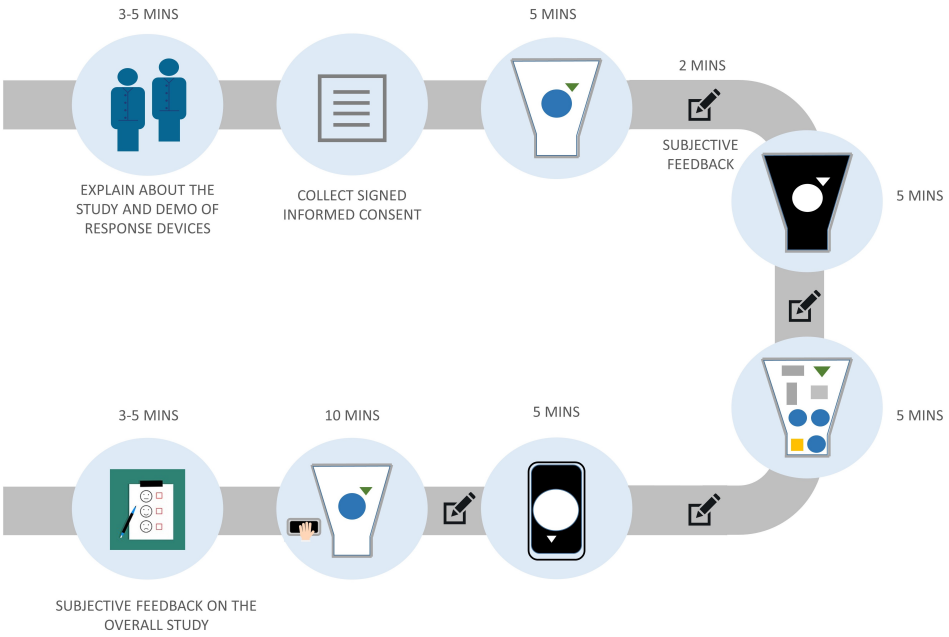


Figure 5.6: The timeline of the pilot study. The participants are initially briefed about the study and asked to complete an informed consent form. The four attention tests except DelSky-Smart are performed in pseudo-randomised order. A subjective feedback about each of the test is collected at the end of the test and an overall subjective feedback is collected at the end of the study.

In the subjective feedback about the tests, the participants were asked to rate the the attention tests against each other based on the distractors used in the test, color scheme of the tests. The participants also rated individual attention tests based on their difficulty and expressed their opinions about the target, distractors and the position of the visual elements. Two of the questions in the subjective feedback about the attention tests were later found to be not useful and were not considered during result analysis. These question were "How was the pace of the test?" and "How would you rate the attention tests based on the ease of use?".

In the subjective feedback about the devices, the participants rated the devices based on the uesr interaction and ease of use of the device. The participants expressed their opinions about using the device as a response device in the ICU. Finally, the participants were asked to rate the devices against each other based on usability and ease of use.

The DelSky-Smart test was always performed as the last test whereas the DelApp-ICU, DAViS, DelSky and VSDT were performed in pseudo-randomized order(Table 5.1). This minimized the learning effect among the participants.

Table 5.1: Order of the attention tests based on the participant number. The tests were pseudo-randomized and were performed in the same order as described in this table.

Participant	Test 1	Test 2	Test 3	Test 4	Test 5
PP1	DelApp-ICU	DelSky	VSDT	DAViS	DelSky-Smart
PP2	DelSky	DAViS	VSDT	DelApp-ICU	DelSky-Smart
PP3	VSDT	DelSky	DAViS	DelApp-ICU	DelSky-Smart
PP4	DelSky	DelApp-ICU	VSDT	DAViS	DelSky-Smart
PP5	VSDT	DAViS	DelSky	DelApp-ICU	DelSky-Smart
PP6	DAViS	DelSky	DelApp-ICU	VSDT	DelSky-Smart
PP7	DelApp-ICU	VSDT	DAViS	DelSky	DelSky-Smart
PP8	DAViS	DelApp-ICU	DelSky	VSDT	DelSky-Smart
PP9	DelSky	VSDT	DelApp-ICU	DAViS	DelSky-Smart
PP10	DelApp-ICU	DAViS	VSDT	DelSky	DelSky-Smart

After the participant know the procedure of the study, the participant is asked to lie down in the bed and the experimenter starts the study. The experimenter once again gives a brief of the test to be conducted to

the participant and upon their consent starts the test. The experimenter operates the tests using a laptop. The experimenter enters the participant number in the laptop and a Python script queues the videos for the mentioned participant. This list is compiled based on the Table 5.2. The list is prepared such that it ensures that all the participants experience the same cognitive load. The participant had the choice of stopping the experiment whenever they wanted.

Table 5.2: The number of times circle blinks in each trial based on participant number. Please note that the length of the videos varies from 122 seconds to 132 seconds due to randomized order.

Participant	Level 1			Level 2			Level 3			Time
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	(sec)
PP 1	3	5	4	5	4	4	6	5	3	123
PP 2	5	3	4	4	5	4	5	6	3	123
PP 3	5	4	3	4	4	5	5	3	6	126
PP 4	4	5	3	5	4	4	6	3	5	123
PP 5	4	3	5	4	5	4	3	6	5	129
PP 6	3	4	5	4	4	5	3	5	6	132
PP 7	3	5	4	4	5	4	6	5	3	124
PP 8	5	3	4	5	4	4	5	6	3	122
PP 9	4	5	3	4	4	5	3	5	6	129
PP 10	4	3	5	4	5	4	6	3	5	126

In DelSky-Smart, the participant responds with each of the response device one by one. The order of the devices are randomized to eliminate the learning effect in the participants. The experimenter explains the usage of the device before starting the DelSky-Smart test and then the participant is allowed to have a few trail responses with the device. Then, the experimenter starts the test and the participant responds using the provided device. The experimenter does not need to gather response in this test and the response is saved in a result log file automatically. The experimenter collects the subjective feedback of the participant after the usage of each device. A timeline of the DelSky-Smart test is shown in figure 5.7.

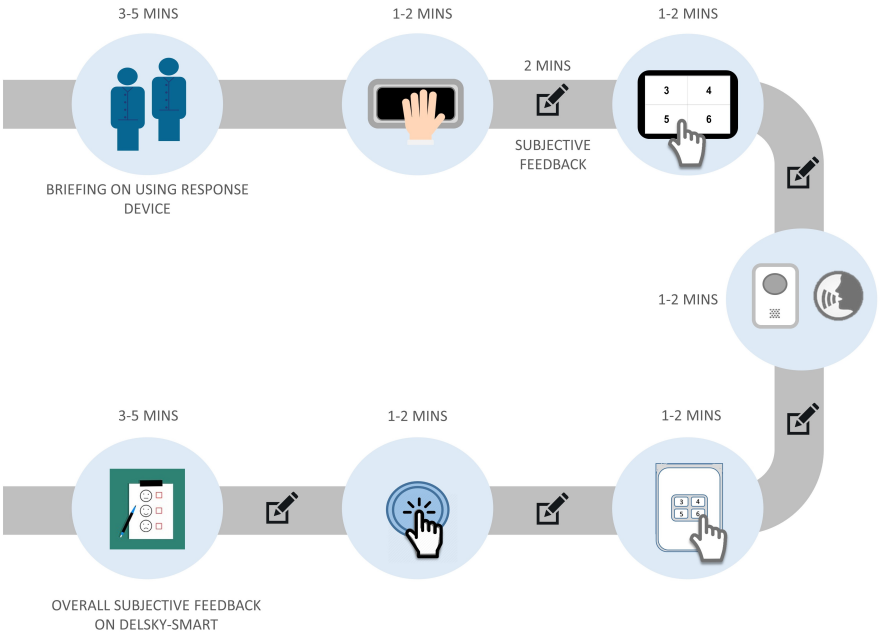


Figure 5.7: The timeline of the DelSky-Smart test. In this test, the participant responds using five response devices. After using each device, the participant is asked to fill a subjective feedback about the device and at the end of the test the participant fills in an overall subjective feedback.

Upon completion of all the tests, the participant is asked some more questions about the overall study like comparing and rating the tests and devices. This concludes the pilot study and the experimenter discusses about the experiences of the participant about the study.

6

Results

The pilot study consisted of five male and five female volunteers. There were four volunteers which belonged to the age group of 26-35 years, another four volunteers were between 36-45 years old. There was one volunteers each in the 46-55 years and 56-65 years age group.

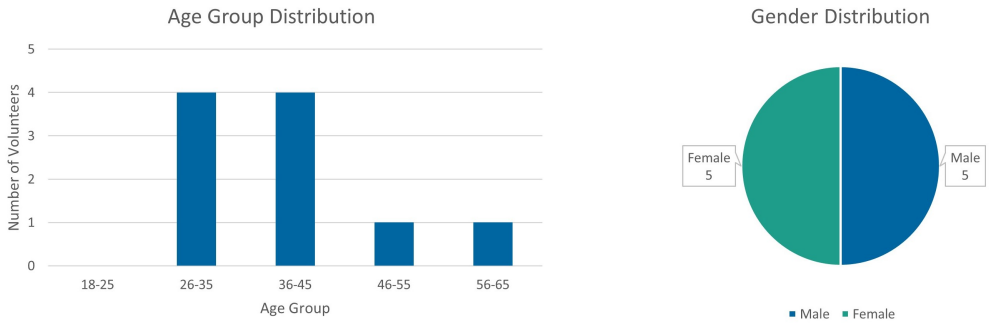


Figure 6.1: The distribution of age and gender among the volunteers. The participants voluntarily participated in the study and also provided an informed consent form before participating in the pilot study.

6.1. Attention Tests

6.1.1. Scores of the participants

The scores of the participants were exactly the same in DelApp-ICU, DAViS and DelSky tests. All the participants completed these three tests without making any errors. However, two out of 10 participants did make mistakes in the VSDT test which was supposed to be more difficult then the DelSky and DAViS. One of the participant gave two incorrect responses in VSDT and achieved a score of 7 whereas the other participant gave one incorrect response and got a score of 8.

In case of DelSky-Smart test, out of all the responses submitted by the participants, mistakes were noticed only in three instances. In all the three instances the devices and the participants were different. The devices with which incorrect responses were provided were respeaker, flic button and keypad.

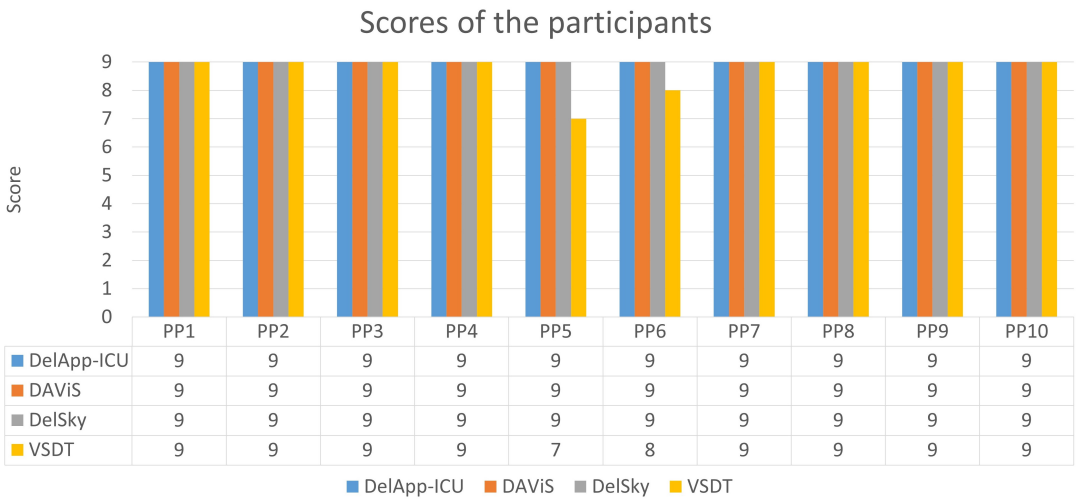


Figure 6.2: The scores of the participants in the different attention tests. The participants got exactly similar scores in DelApp-ICU, DelSky and DAViS tests.

6.1.2. Subjective feedback by the participants

6.2.2.1. DelApp-ICU

The participants rated the DelApp-ICU to be easy to perform. On a scale of five, where one was difficult and five was easy, six participants rated DelApp-ICU five and four participants rated it four. When asked about the sizes of the target and distractors, all the participants felt that the size of the target was appropriate whereas seven participants stated the size of distractor was right and the remaining three participants said that they would have preferred a bigger distractor. Six out of 10 participants mentioned that the distractors were visible to them but did not distract.

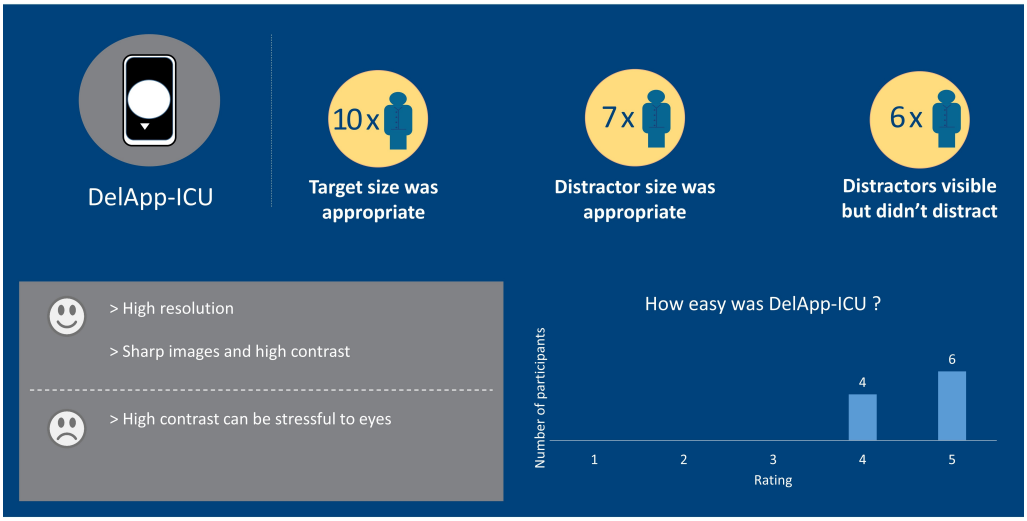


Figure 6.3: Compilation of subjective feedback of the participants for the DelApp-ICU test. For the difficulty of the test, a rating of 5 meant easy whereas a rating of 1 meant difficult.

One of the participant raised the concern that high contrast can be stressful to the eyes when exposed for longer duration. In DelApp-ICU, the participants were not asked if the visual content was in their field of view because the test was performed on a smartphone and participants had the freedom to move and adjust the device such that it was in their field of view.

6.4.2.2. DAViS

The participants rated the difficulty of the DAViS tests very similar to the DelApp-ICU. The participants felt the tests was easy with 9 of the participants rating DAViS either four or five. All the participants agreed that the visual content was in their field of view. The size of the target was appropriate for eight of the 10 volunteers whereas nine of the 10 volunteers felt that distractors were of the right size. Five participants saw the distractors but were not distracted by it, one participant was distracted by the distractors whereas for four participants the distractors were not prominent. When it came to distinguishing the target from the distractors, nine participants said that they were successful in doing it.

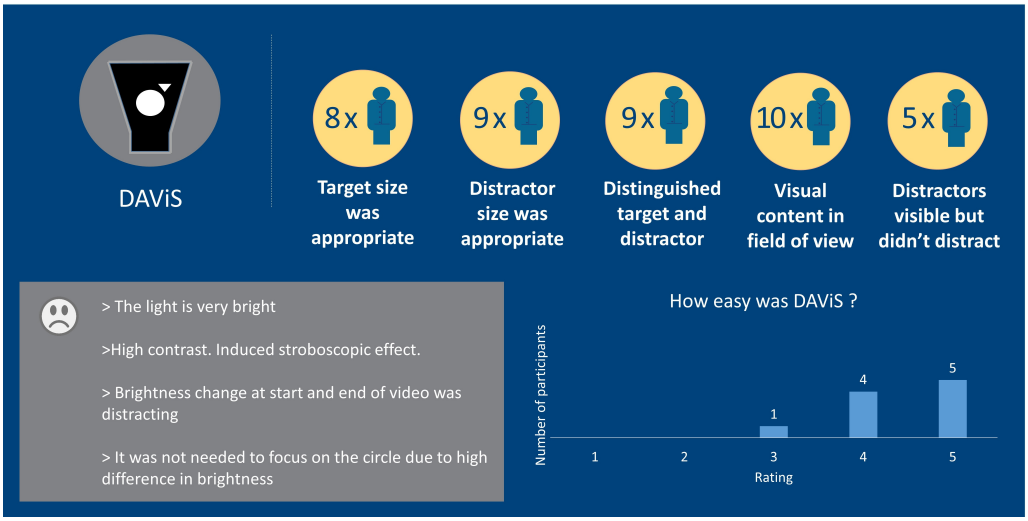


Figure 6.4: Compilation of subjective feedback of the participants for the DAViS test. For the difficulty of the test, a rating of 5 meant easy whereas a rating of 1 meant difficult.

The blinking circle of the DAViS test was found to be unpleasant by eight participants. The participants stated that the light emitted by the circle was very bright and in some occasions it induced stroboscopic effect. Another participant made a remark that because of the high difference of luminosity between the appearing and disappearing of the circle, it was not required to focus on the circle and instead just noticing the change of brightness in the room will do the job.

6.4.2.3. DelSky

All participants except one felt that the distractor and target were appropriate in size and they had no difficulty in distinguishing the target and the distractors. The remaining one participant said that the distinction between the target and distractor was made by looking at the size and color rather than focusing on the shapes. All the participants said the the visual elements were perfectly in the field of view of the participants. Eight participants voted that the distractors were visible to them but did not distract, the remaining two participants felt that the distractors were not prominent and could be easily ignored.

The participants felt that the visual content was pleasing and did not had the extreme fluctuations in the brightness. One of the 10 participant felt that the white background of the DelSky test becomes uncomfortable to look at after a while and hence it becomes difficult to focus on the visual content.

6.4.2.4. VSDT

VSDT was little different from the other tests and helped in proving some of the decisions taken during the design phase of the attention tests. The participants had more difficulty in distinguishing between target and distractors in VSDT test. In Chapter 3, while designing the attention test it was found that some shapes such as squares, circles and hexagons were difficult to differentiate from each other. Since, VSDT uses a combination of shapes such as squares, circles, rectangles and triangles, this would have led to greater number of participants not being able to distinguish target from distractors or vice versa. This also had an impact in the difficulty of the test where we can see slightly more variation in the ratings compared to DelSky and DAViS.

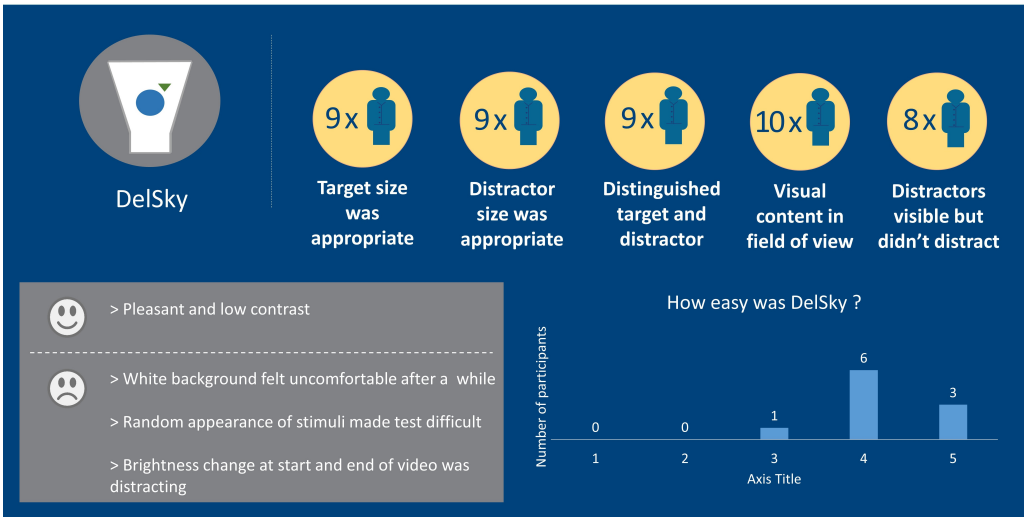


Figure 6.5: Compilation of subjective feedback of the participants for the DelSky test. For the difficulty of the test, a rating of 5 meant easy whereas a rating of 1 meant difficult.

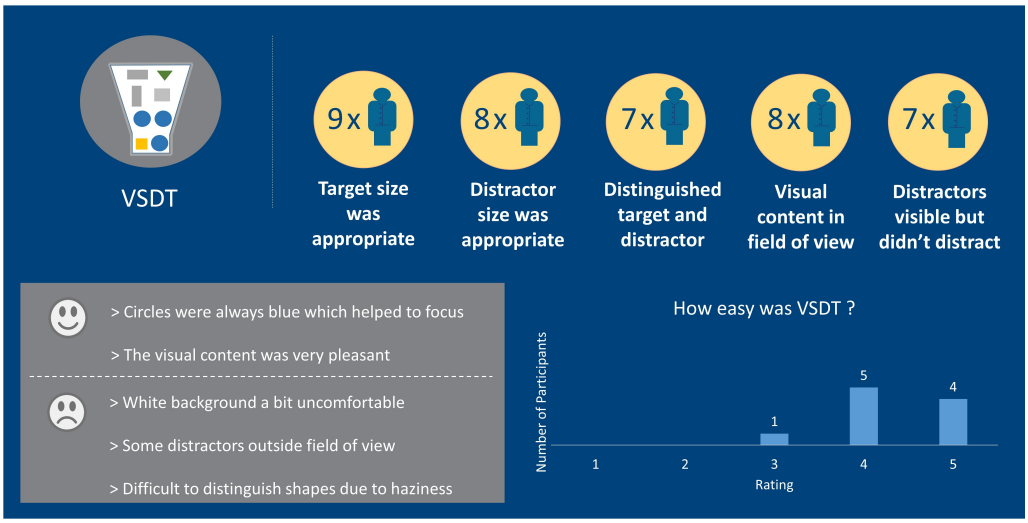


Figure 6.6: Compilation of subjective feedback of the participants for the DelSky test. For the difficulty of the test, a rating of 5 meant easy whereas a rating of 1 meant difficult.

The participants stated the sizes of the target and the distractors was appropriate with nine and eight participants agreeing with the sizes respectively. For eight participants the visual content was in their field of view. Some participants reported that some of the distractors were outside their field of view which was due to the fact that the visual content was spread across the entire area of the VitalSky and hence some of the distractors were then placed outside the field of view of the participants. Three participants believed that the distractors were not prominent and could be easily ignored whereas seven participants stated that the distractors were visible but did not distract.

The participants found the visual content of the VitalSky to be very pleasing. However, one of the 10 participant found the white background to be a bit uncomfortable on looking at it for longer durations.

6.4.2.5. Overall study feedback

Size and position of the visual element in the attention tests

In the subjective feedback of the individual tests, all the participants said that the visual elements (target and distractors) were in their field of view for DelSky and DAViS whereas eight of 10 participants had the same

view for the VSDT. This justifies the decisions taken for the positioning of the visual elements during the design phase of the attention tests. In all the attention tests at least 80% of the participants stated that the size of target and distractors were appropriate and hence verifies that the sizes selected for the target and distractors were appropriate.

Color scheme of attention tests

On a scale of 1 (Most preferred) to 4 (Least preferred), the most preferred test by the participants was DelSky followed by the VSDT (see figure 6.7). The DelApp-ICU was the third preference for the participants where as DAViS was the least preferred test based on the color scheme of the tests.

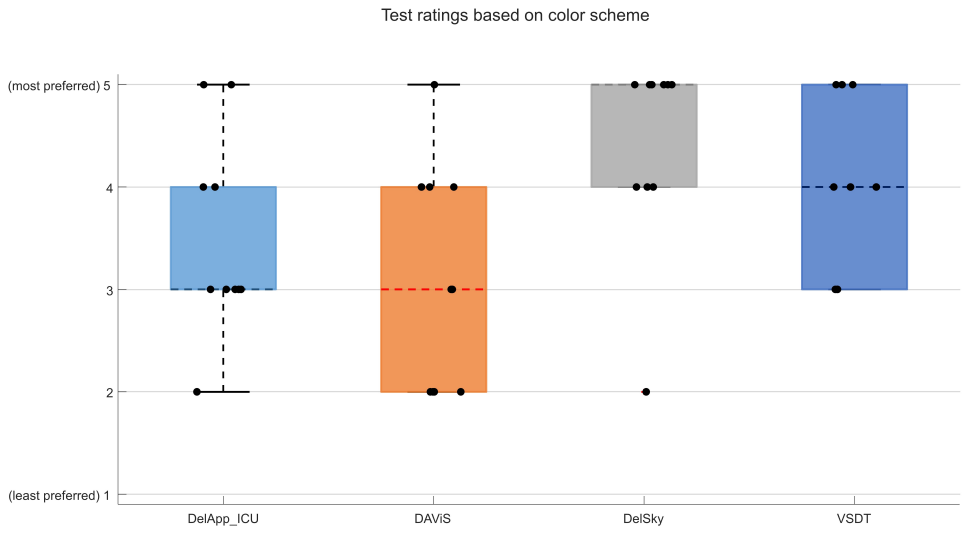


Figure 6.7: The ratings given by the participants when they were asked to rate the attention tests against each other based on the color scheme of the tests. A rating of 1 indicates most preferred test where as a rating of 5 was given to the least preferred test.

Similarity between the attention tests

Based on the distractors of the attention test, there was a lot of similarities observed between the ratings of the DelApp-ICU, DAViS and DelSky tests (see figure 6.9). The participants were asked to rate between one (most distracting distractor) and four (least distracting distractor). The ratings for the distractors of DelApp-ICU, DelSky and DAViS were very similar and in many occasions exactly the same.

Similar results were obtained when we compare the ratings given by the participants for how easy were the individual tests. For DelApp-ICU all of the participants rated DelApp-ICU four or greater where five indicates easy and one indicates difficult. On the other hand, DelSky and DAViS both received a rating of four or above from 9 of the participants. When we look into the average ratings of the individual tests, DelSky was rated 4.2, DAViS was rated 4.4 and DelApp-ICU 4.6.

In the subjective feedback, the participants were asked to rate the similarity between the DelApp-ICU and DAViS and DelApp-ICU and DelSky based on the attentional demands of the tests. The participants had to rate between one and five where one indicates that the tests were not similar whereas five indicates that the tests were exactly similar. The results are shown in Figure 6.10. The average rating for the DelSky was 3.1 whereas for DAViS it was 3.2. When we look into the distribution of the ratings, 70% of the participants rated both DelSky and DAViS 3 or above. This indicates that the tests were not exactly similar but there was some similarity between the tests.

Another important point to mention here is that some of the participants rated the tests based on the color scheme and sharpness of the visual content of the tests rather than the attentional demands. Hence, this factor also had influence in the ratings of similarity of the tests. Also, if the number of participants would have been greater, a more significant difference between the ratings might have been observed.

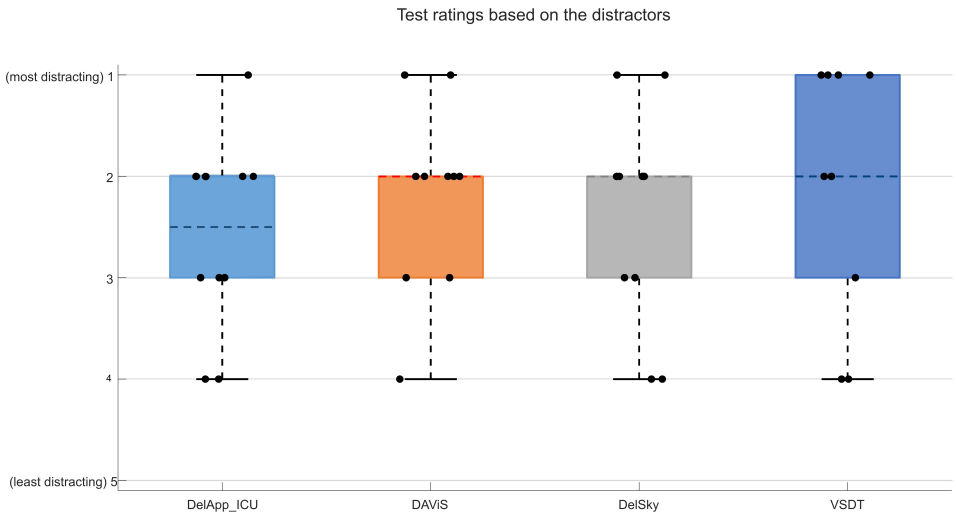


Figure 6.8: The ratings given by the participants when they were asked to rate the attention tests against each other based on the distractors used in the tests. A rating of 1 indicates test with most distracting distractors where as a rating of 4 was given to the test with least distracting distractors.

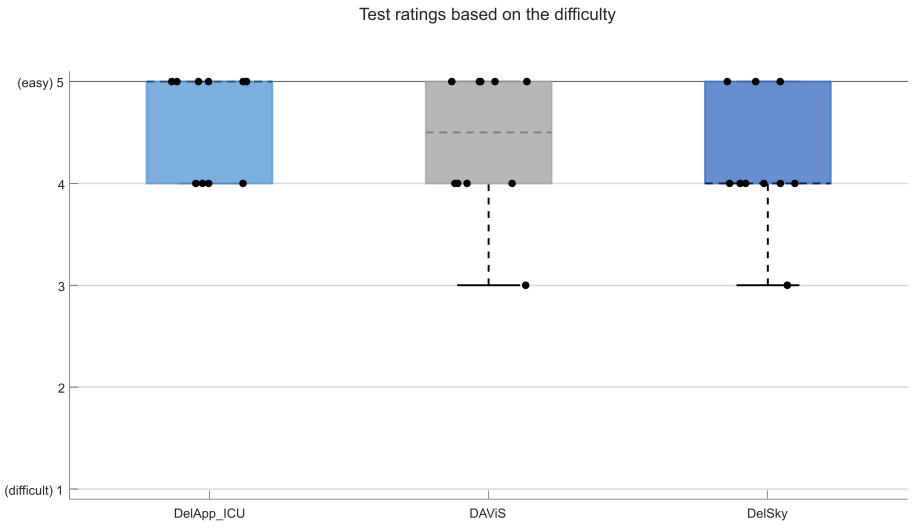


Figure 6.9: Compilation of the ratings given by the participants for the difficulty of the individual tests. A rating of 5 indicates the test was easy whereas 1 indicates that the test was difficult.

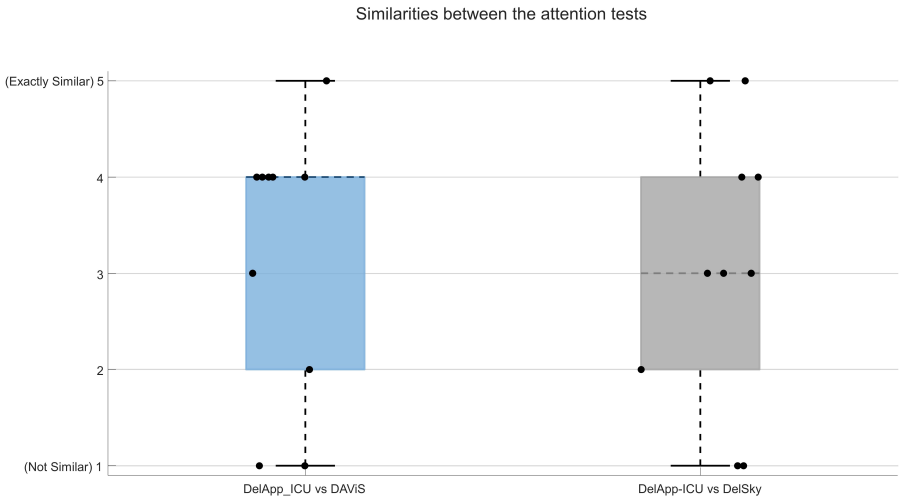


Figure 6.10: The ratings given by the participants when asked for the similarity between the DelApp-ICU and DAViS tests and DelApp-ICU versus DelSky tests. A rating of 5 meant the two tests were exactly similar whereas a rating of 1 meant the tests were not similar at all.

6.2. Response Devices

6.2.1. Leap Motion

In terms of user interaction, three participants liked the device and rated the user interaction as excellent (5 on a scale of 5) whereas one of the participant rated the response as poor (1 on a scale of 5). The user interaction of Leap Motion was rated on average 3.4 out of 5 (see figure 6.11.

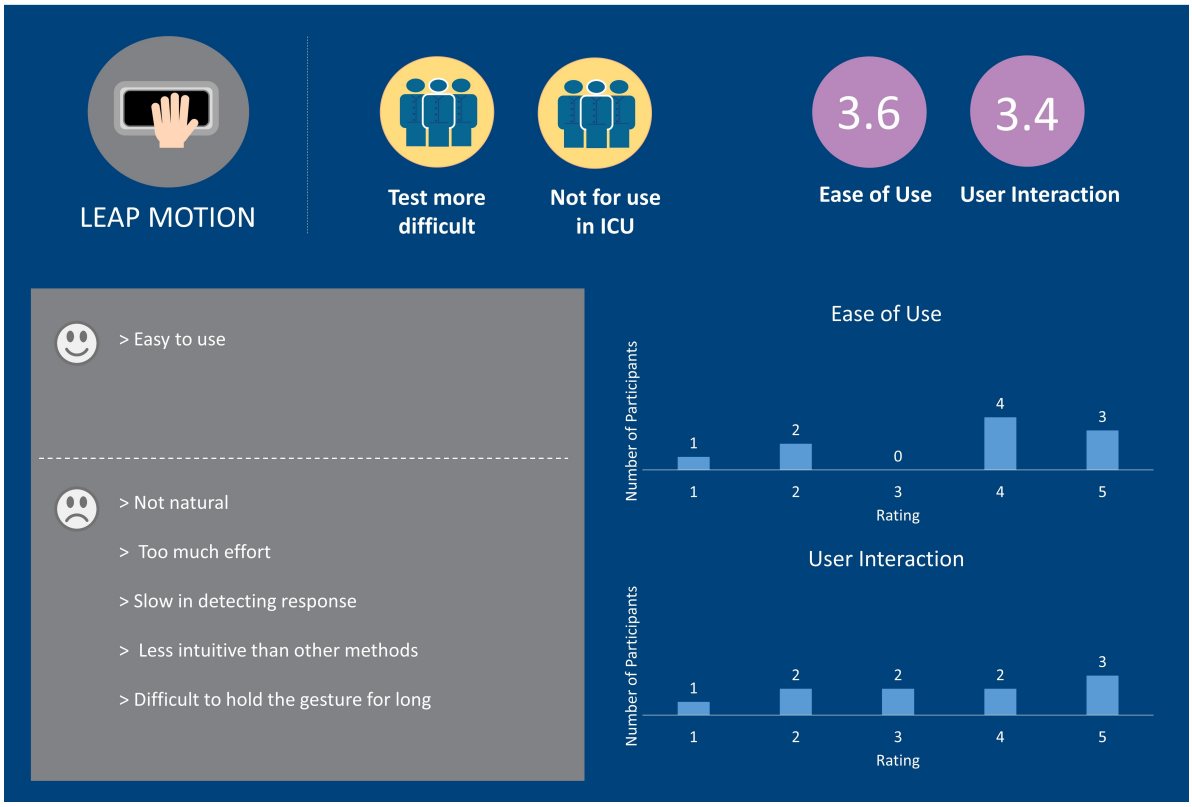


Figure 6.11: The summary of the subjective feedback of the participants for the leap motion device. The leap motion received low rating from the participants for ease of use and user interaction.

In terms of ease of use, the participants had varying opinions about the leap motion. Three participants rated the device 5 on a scale of 5 (which indicates easy) but one of the participant rated the device 1 (which indicates difficult). Four of the participants rated the ease of use as 4 where as the remaining two participants rated 2. The average rating of the leap motion for ease of use was 3.6 out of 5 which was also the lowest among all the devices. Moreover, three of the participants stated that they felt the attention test was more difficult when they had to respond using the leap motion.

When the participants were asked their opinion about using Leap Motion as a response device for the ICU patients, three participants did not recommend the use of the device in the ICU on the other hand one of the participant felt the device can be used in the ICU. The remaining participants did not explicitly state a yes or no but had doubts over certain features of the leap motion. One of the participant stated that responding with the Leap Motion was too much effort as the participant recall the gesture for the response and then hold it over the leap motion. Another participant expressed concern that the patients in the ICU might not have sufficient control over their hand and it might be difficult for them to hold the gesture for long. The participants also stated the device to be slow in detecting the response. Other subjective feedback and remarks made by the participants are illustrated in figure 6.11.

6.2.2. Respeaker

The respeaker achieved an average rating of 3.4 out of 5 for user interaction. The device was rated 5 by two participants whereas two participants rated the device 2 out of 5. Out of the remaining six participants, four participants rated the device 3 and two rated 4 out of 5 (see figure 6.12).

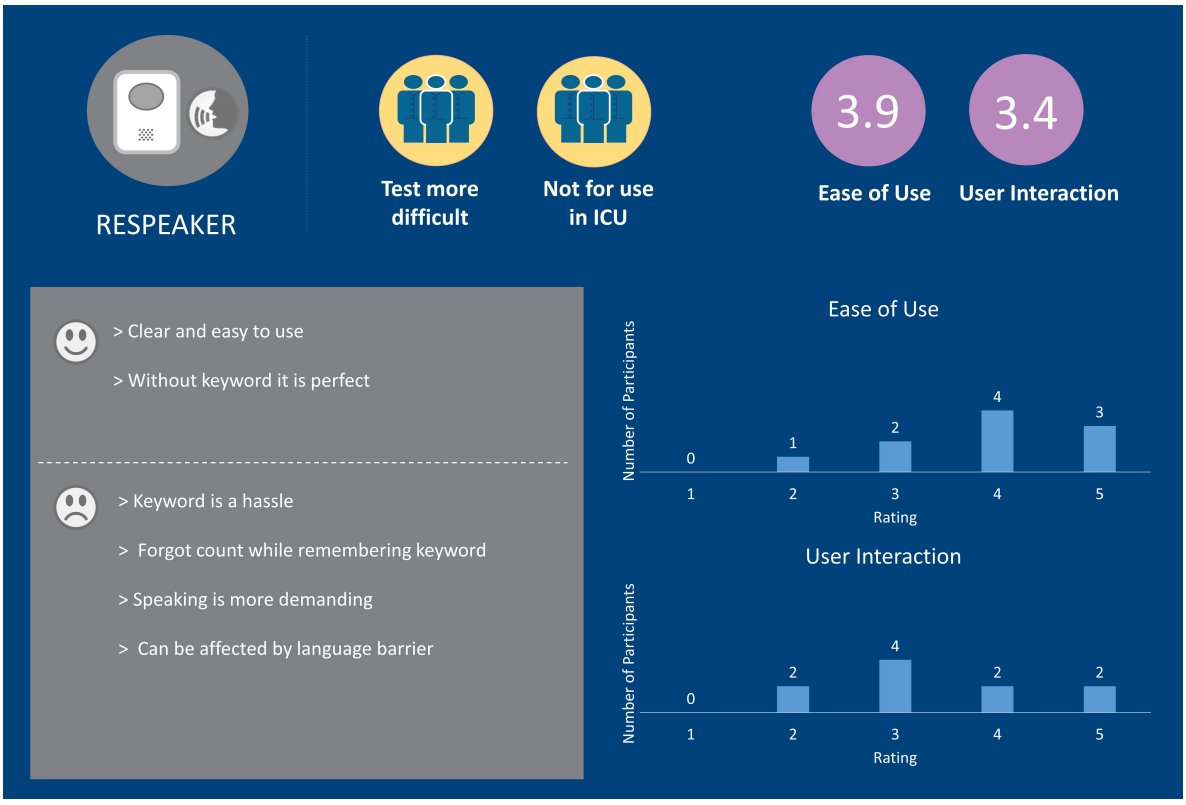


Figure 6.12: Compiled subjective feedback from the participants for respeaker as a response device. Most of the participants felt that respeaker will do better as a response device without the "Hey Snips" keyword.

On terms of ease of use, three participants rated the device 5 and four participants rated 4. There were three participants who rated the device 3 or below, with two participants rating device 3 and one participant rating the device 2. The average rating for the Leap motion for ease of use was 3.9 out of 5.

Most of the participants stated that they did not like the usage of keyword. Some of the participants had problems remembering the keyword whereas some forgot the number of the blinks of the circle while re-

membering the keyword. As a result, three out of 10 participants mentioned that they felt the attention test was more difficult using the respeaker.

When the participants were asked if the device is a good option to use in the ICU, three participants recommended the device whereas three participants did not recommend the device. The remaining participants did not state if the device can be used or not but instead gave suggestions or remarks about the device. The most common remark by the participants was that the use of the keyword before saying the response made it difficult for them to respond with the device. Some participants also raised concern about the language issues that might arise with the usage of the device, as the patient must be able to speak and understand the language that will be used by the respeaker. Other remarks from the patients are illustrated in figure 6.12.

6.2.3. Flic buttons

The user interaction of the flic button received a lot of remarks. The most common remark was about the use of circles as the visual feedback for indicating the number of times flic button was pressed. This confused the participants with the target stimuli and made them think it was part of the attention test and needs to be counted. Another problem faced by the participants was that it was not clear for them that when they should respond. This was because audio feedback was not used in the case of flic buttons and only visual feedback was used. The user interaction of the Flic button was rated 3.8 out of 5.

Nine out of 10 participants rated the ease of use of flic buttons four or above. Five participants rated the device 5, four participants rated the device 4 and one participant rated it as two. The average rating for the flic button for ease of use was 4.3 out of 5 (figure 6.13).

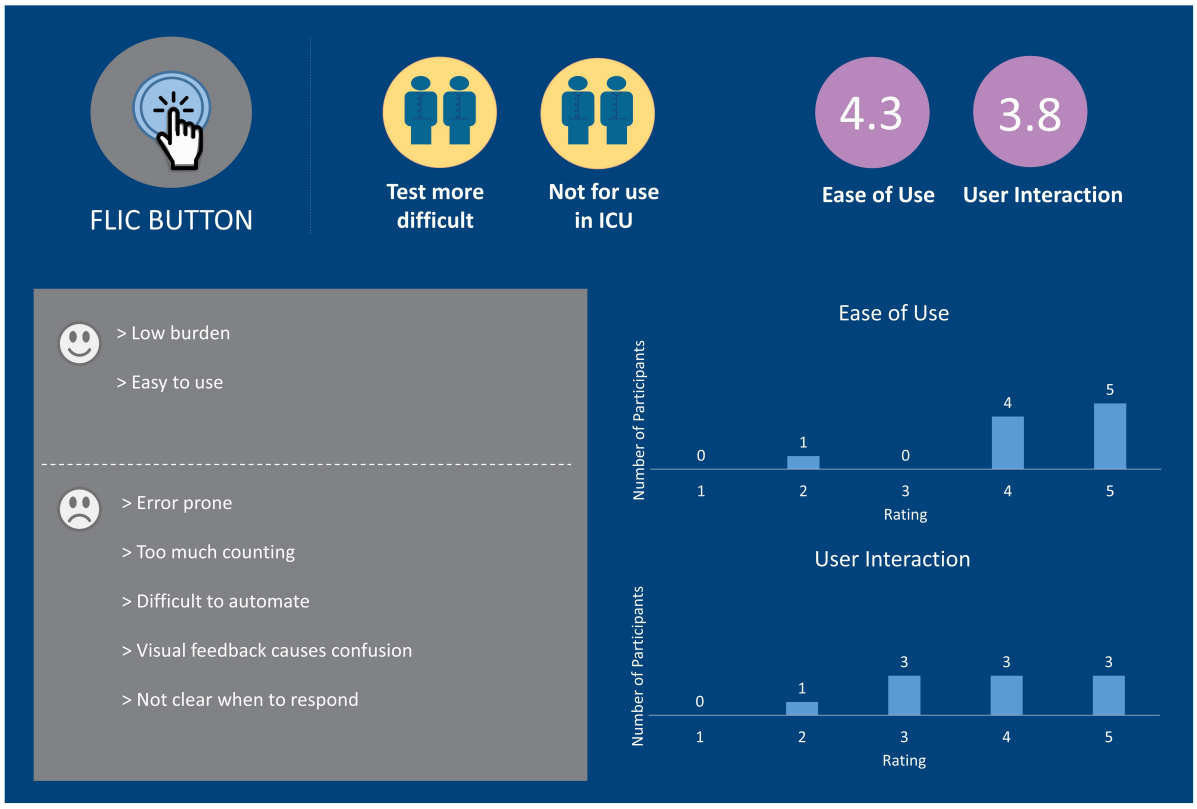


Figure 6.13: Participants feedback about using flic buttons as a response device for ICU patients. The participants liked the simplicity provided by the flic buttons but the visual feedback provided on the VitalSky caused confusion.

Four participants believed that the flic button can be used as the response device for the ICU patients as it is easy to use and low burden device for the patients. Among the remaining participants, two participants did not recommend the use of flic buttons in the ICU whereas rest did not stated an explicit opinion. The negative remarks involved that the patients will need to do too much counting when responding with flic button as

they first need to count the circles and then they need to count again to press the flic button right number of times. Another feedback was that it will be difficult to automate the test with flic button because someone will need to manually handover the button to the patient.

6.2.4. Tablet

The participants liked the user interaction of the tablet and the ease of use it offered. The user interaction for responding with tablet was rated 4.3 out of 5 whereas ease of use was rated 4.4 out of 5 by the participants. The participants felt that the user interaction was good except for the fact that they have to confirm their choice by pressing "Ok". This extra confirmation step was not necessary in the participants view. Apart from the confirmation step, two of the participants had trouble with tablet not detecting their touch input properly. The participants mentioned that they faced this issue with other touch screens as well and could be caused due to the dryness of the skin and hence it should be taken care while using the device in the ICU. Due to this the participants had to press the tablet multiple times and one of the participant among them felt that the test was more difficult by using the tablet as the response device.

For ease of use, participants were happy to respond with the tablet as they have to just tap on the tablet screen and since the tablet was mounted on a holder they didn't had to hold it. However, one participant made the remark that to respond with the tablet, the patient will need to change their focus from one screen to another, i.e, from VitalSky to tablet. This could be straining for the patient. Another participant felt the device was clumsy and cumbersome to be included in the ICU.

Five of the 10 participants felt that tablet as a response device can be used in the ICU and believed the tablet to be fast, intuitive and less error-prone. One of the participant did not recommend usage of the tablet in ICU as it can be straining for the patient to switch between two screens. Four participants neither recommended nor declined the usage of the device in the ICU.

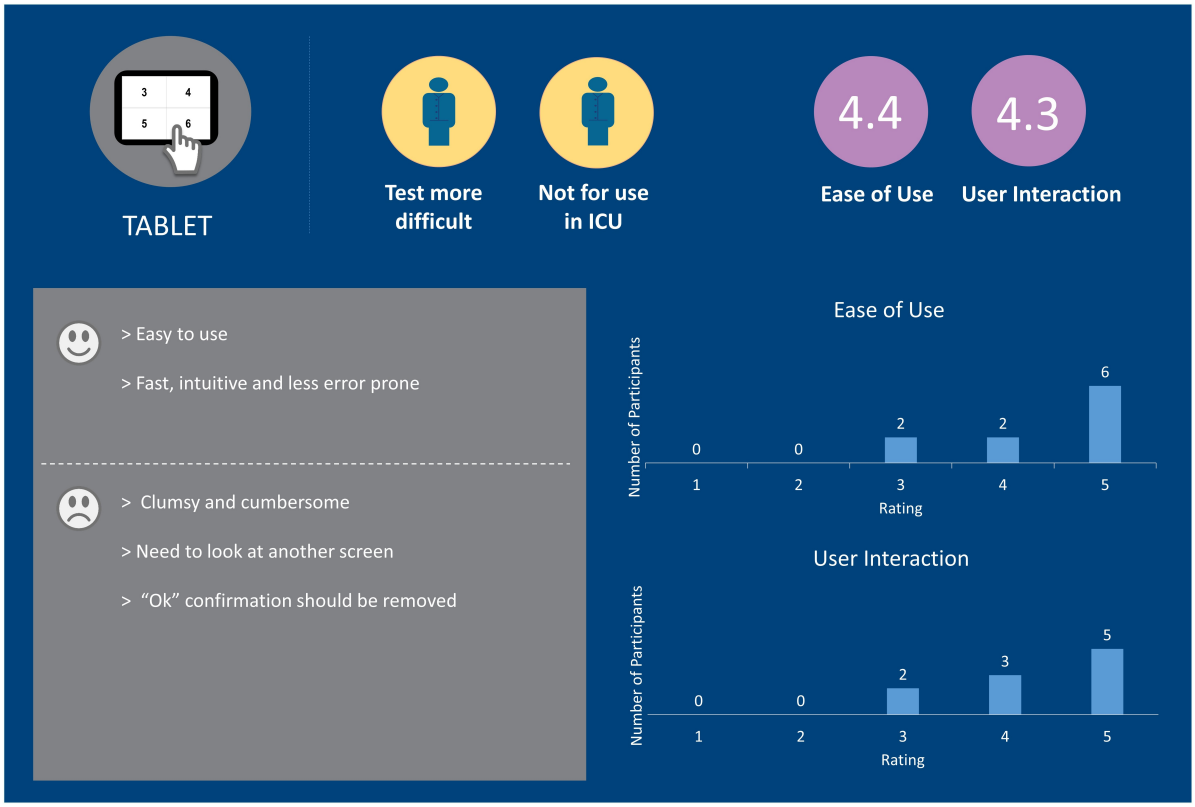


Figure 6.14: The subjective feedback from the participants about the tablet is illustrated in this figure. Most of the participants liked the user interaction and ease of use of the device and felt that the tablet can be used as a response device in the ICU.

6.2.5. Bluetooth keypad

The user interaction of the device was rated very high with seven of 10 participants rating it a 5 out of 5. One participant stated that if the device would have been mounted like tablet it would have been better where as one of the other participant felt that holding the device provided more focus. Only one of the 10 participants felt that the attention test was more difficult with the keypad stating the reason to be that the there was a need to use hands and the participant need to look away from the VitalSky and change focus to see the keypad.

The participants felt that the device was easy to use, simple and fast. For ease of use, nine out of 10 participants gave the keypad a 5 out of 5 rating whereas one participant rated it 4 out of 5 resulting in an average rating of 4.9 out of 5. One of the participant stated that the device could also be used as a help device for the ICU patients where they can tap the button when they needed help.

When the participants were asked if the device is a good option to use in the ICU, the participants were very positive and seven of 10 participants said the device can be used in the ICU. Three participants were unsure and said the usage of device will depend on the state of the patient. None of the participants said the device could not be used in the ICU. One of the participant raised the concern that the keypad might face issues with infection control.

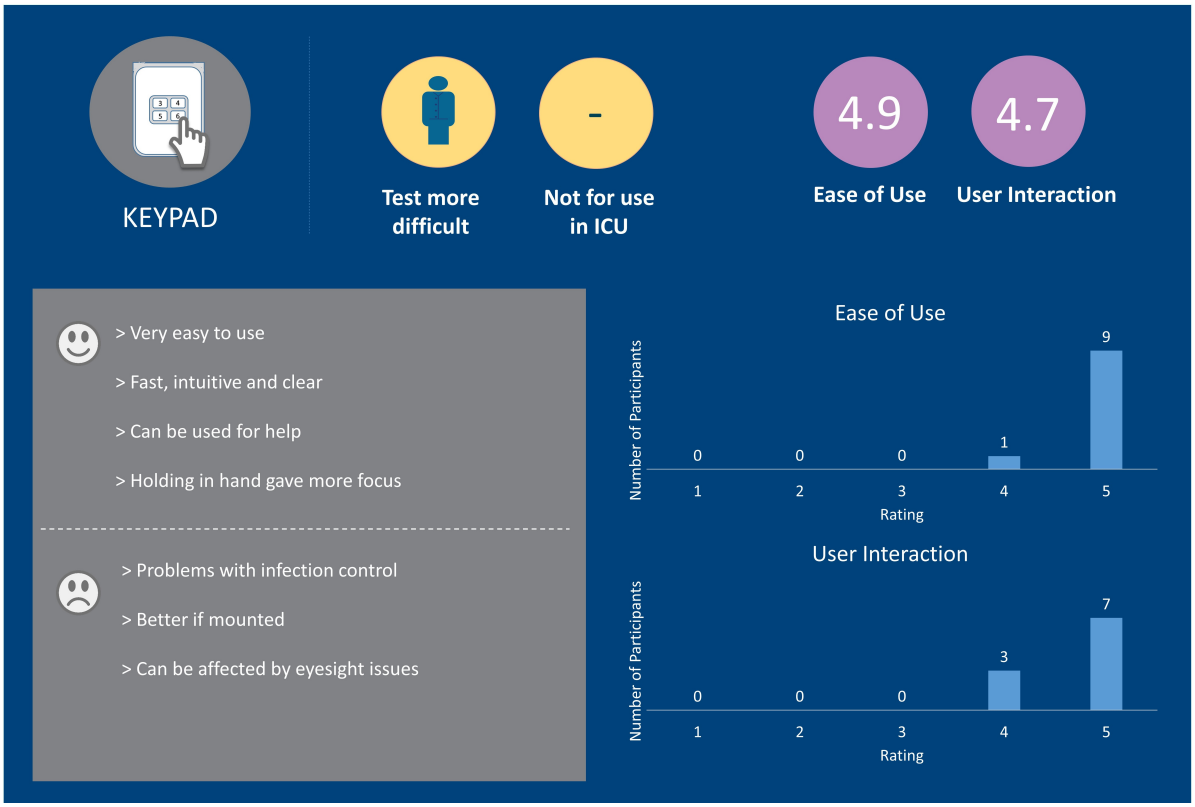


Figure 6.15: The subjective feedback of the participants for the Bluetooth keypad. The keypad was the most preferred device by the participants both in terms of ease of use and user interaction.

6.2.6. Comparison of the response devices

Apart from rating the devices individually, the participants were also asked to rate the devices against each other. The participants had to rate the devices against each other based on the usability and ease of use. The participants should rate the most preferred device as one and least preferred device as five. If the participants felt that two or more devices were exactly similar to each other they were allowed to provide the same rating to these devices.

From the plots in figure 6.17 and figure 6.18, it is clear that keypad was the most preferred device among the participants followed by tablet. The participants had varied opinions about the flic buttons both in terms of

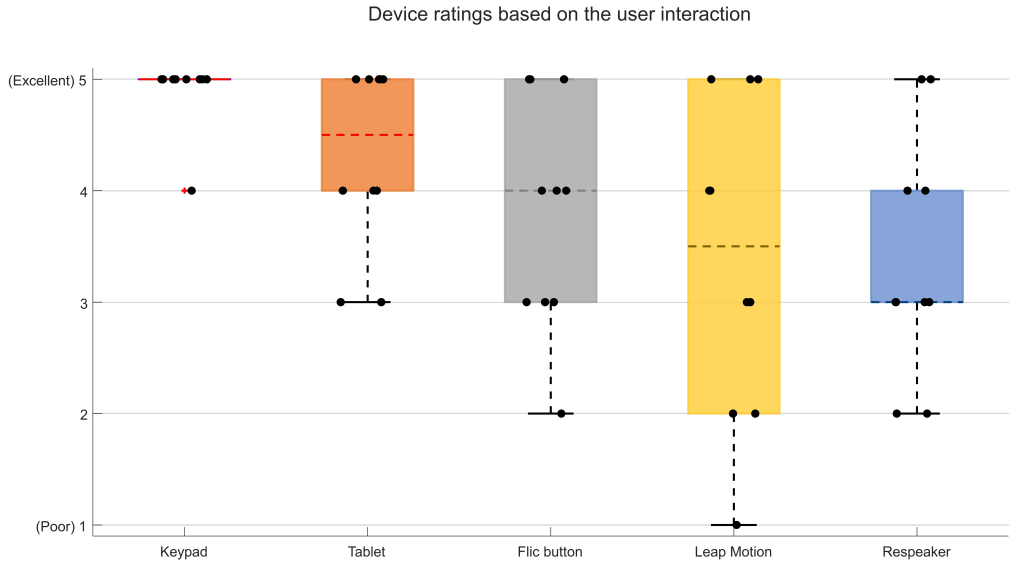


Figure 6.16: The boxplot shows the distribution and comparison of the user interaction ratings given by the participants to each response devices.

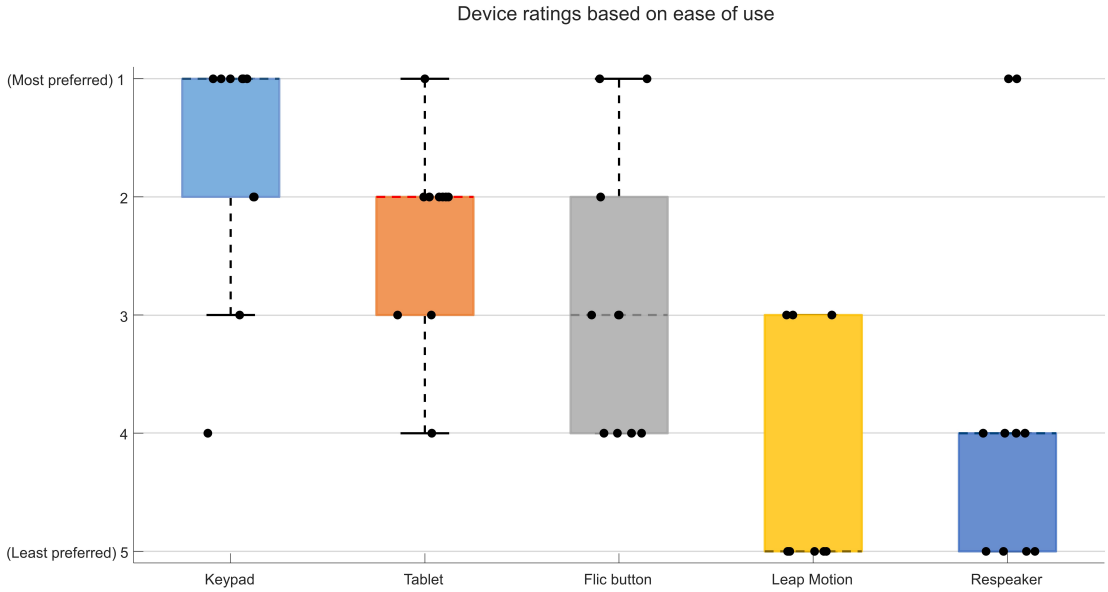


Figure 6.17: The distribution of the ease of use ratings when participants were asked to rate the devices against each other.

ease of use and usability. A similar variation is observed in the ratings of the respeaker, where participants liked the response device in general but were not comfortable with the use of the keyword to respond. The leap motion was the least preferred device by the participants and was constantly rated low by the participants.

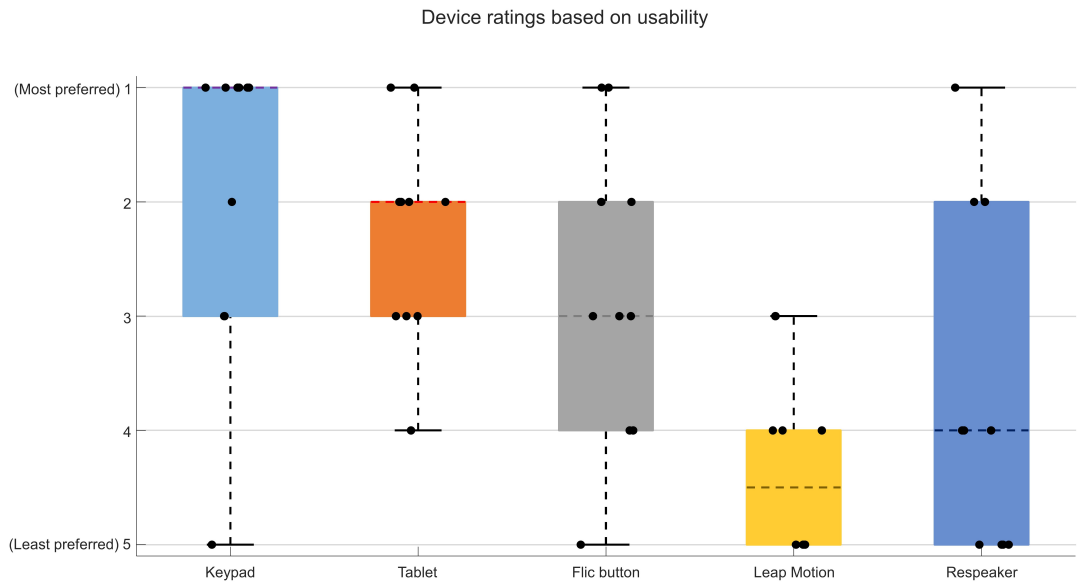


Figure 6.18: The distribution of the usability ratings when participants were asked to rate the devices against each other.

7

Discussion and Conclusion

7.1. Discussion

Similarity between the attention test

The main goal of the study was to develop attention test for the VitalSky and benchmark its performance with the DelApp-ICU. Here, we will compare the performance of the participants in the DelSky and DAViS attention tests with the DelApp-ICU. The test VSDT and DelSky-Smart are not considered in this comparison because VSDT differed from the design of the DelApp-ICU and was designed to explore the possibilities of the attention test with the VitalSky whereas DelSky-Smart was more focused on understanding the user preference of the response devices and technical feasibility of an automated attention test.

It was found out that the scores obtained by the participants in the DelApp-ICU, DAViS and DelSky tests were exactly the same. All the participants completed the test without making any mistakes. In case of the DelApp-ICU study, six out of 10 participants had completed the DelApp-ICU test without any mistake [65]. The difference could have been due to the audience on which the test was conducted. This was expected because all the participants were healthy volunteers without any history of cognitive disorders. The tests were fairly simple for the volunteers and all the volunteers completed DelSky, DAViS and DelApp-ICU tests without any errors. This was expected because the tests are targeted towards delirium patients and are not expected to be challenging for normal healthy population and using more challenging tests for normal population like Sustained Attention to Response Task (SART) [66, 67] would have made the test unfeasible for the delirium patients. The learning from this step was that the performance of the participants were comparable in the DelApp-ICU, DAViS and DelSky tests and it was not the case that there were large differences in the scores of the participants in different tests.

The participants believed that the DAViS and DelSky tests had similarity with the DelApp-ICU test. This can be said because seven out of 10 participants rated the similarity between the DelApp-ICU and DelSky and the similarity between DelApp-ICU and DAViS tests 3 or above where 1 was not similar and 5 was exactly similar.

Another factor confirming the similarity of the test was rating the test against each other based on the distractors. The ratings based on distractors indicate how much the tests distracted the participants. The ratings given by the participants were very similar by the DelApp-ICU, DAViS and DelSky. Four out of 10 participants gave DelApp-ICU and DelSky same ratings in terms of distractors whereas for DelApp-ICU and DAViS six out of 10 participants gave the tests same ratings. In four out of 10 instances, all the three tests got the same rating from the participants which again highlights the similarity between the three tests.

When looked into how easy the tests were, the DelApp-ICU, DAViS and DelSky tests were again rated similar to each other. All the participants rated DelApp-ICU either 4 or 5 whereas nine of the participants did the same for each DelSky and DAViS. This indicates that the participants found the DelSky and DAViS tests to be slightly difficult than DelApp-ICU, but still found the three tests to be similar to each other.

Thus, taking into account the score of the participants in the attention tests and ratings from the participants

on difficulty, distractors and similarity of the tests, the DelSky and DAViS display similarity with the DelApp-ICU test. However, measures can be taken to consider the difficulty of the DelSky and DAViS tests as they were found to be slightly more difficult than DelApp-ICU by the participants. By the results obtained from the pilot study, it can be concluded that the DAViS and DelSky tests were feasible when performed as tests of attention on the VitalSky and the performance of the participants in these tests were relatable with the performance of the participants in the DelApp-ICU.

Color scheme of the attention tests

The participants liked the high resolution and sharp images in the DelApp-ICU compared to the other attention tests. But, also one of the participant raised the concern that this high contrast can be straining for the eyes of the patients for longer durations.

For DAViS, the participants found the visual content to be unpleasant and too intensive. These remarks from the participants confirms the findings during the initial design phase of the attention tests where the black and white content on the VitalSky was found to be too bright and intensive and also confirms the hypothesis stating that the visual content of the DAViS will be interpreted as unpleasant or uneasy by the participants.

Nine out of 10 participants found the color scheme of the DelSky and VSDT to be pleasant and appealing. This justifies the choice of the white background instead of the black background in case of DAViS. Hence, the participants favoured the visual content of the DelSky and VSDT much more than the visual content of the DAViS.

However, one participant found the white background to be a bit uncomfortable and it was same participant in both DelSky and VSDT who had trouble with the white background. The participant making this remark belonged to one of the youngest age groups of the study and this effect could have been because of the age as well because the brightness perceived by the people can also be attributed to their age and the younger the person, more light is captured by the pupil of the person. However, appropriate measures can be taken to adapt the brightness of the VitalSky according to the age group of the patients.

Another hypothesis made before the pilot study was that participants will prefer the color scheme used in the DelSky over the color scheme of DAViS. To verify this hypothesis, the participants were asked to rate the tests against each other based on the color scheme of the test. The participants had to give a rating between 1 to 4 where 1 indicated most preferred test and 4 indicated least preferred test. The participant had the option of giving same rating to two tests if they felt they were exactly similar.

The most preferred test was based on color scheme was the DelSky test followed by VSDT. This is understandable as both the DelSky and VSDT had similar color scheme. The least preferred color scheme was DAViS and DelApp-ICU was the third preferred test based on the color scheme. The ratings varied for some of the participants but for majority of the participants the preference order was observed to be as stated. This confirms the hypothesis that people will prefer the color scheme of the DelSky over the color scheme of the DAViS. This also confirms the hypothesis that the visual content of the DelSky will be more preferred by the participant than the visual content of the DAViS.

User preference of the response devices

The preference of the users were the same in terms of ease of use, user interaction and usability. The Leap Motion was the least preferred device among all the devices that were used in the pilot study. Respeaker was second least preferred device by the participants. This was in contradiction to the hypothesis that was made before the pilot study. It was believed that responding verbally with respeaker will supposedly be the easiest way to respond and will be the most preferred response method by the participants but this was not the case. This could have happened because of the keyword "Hey Snips", which the participants have to say before saying their response. The participants also stated that they forgot the count of the circle while trying to remember the keyword.

However, the participants had a more positive experience with the respeaker than leap motion and it is also evident by seeing that majority of the participants rated the user interaction and ease of use of the device between three to five (figure 6.12 (Chapter 6)) whereas in leap motion the ratings were spread over the entire scale.

Flic button was the third preferred device and there were quite some variations in the preference of the participants for the flic button. This variation can be explained by the confusion of the participants with the visual feedback provided on the VitalSky and hence rating the device low.

The tablet was the second most preferred device and Bluetooth keypad was the most preferred device by the participants. The participants really liked the ease of use and intuitive user interaction of the device. The device was found to be simple, fast and very easy to use by the participants. However, one of the participants raised a concern that infection control check for the Bluetooth keypad can be difficult. However, appropriate measures can be taken for infection control like covering the top of the keypad with a glass layer which is easy to clean and disinfect.

As seen above, the participants had a clear preference for the Bluetooth keypad. In terms of user interaction, the participants preferred the touch-based devices over the devices in the other domain like verbal and gesture. However, as discussed before the preference of the participants with the respeaker (verbal domain) could have been affected due to the use of keyword in the respeaker before saying the response. But, for the current scenario there seems to be a preference in the participants for the devices in the tactile domain.

7.2. Conclusion

Delirium is a serious neuropsychiatric problem and greatly increases the risk of mortality and morbidity among the ICU patients. Inattention is one of the core symptoms of delirium and is seen in most of the delirium patients. As a result, inattention can be used to detect delirium in the ICU patients. CAM-ICU is the most popular method to detect delirium in ICU patients. However, the results of the CAM-ICU have been found to be dependent on the experience of the tester. Moreover, the patients get used to the questions of the CAM-ICU and can learn how to answer them over the course of time.

The VitalSky is an innovative luminous ceiling used in the ICU to provide light therapy to the ICU patients. Next to light therapy, the VitalSky can also be used to render visual content. Since, the VitalSky is already installed and available in ICU, it was be used to implement objectives tests of sustained attention using visual stimulus.

The primary goal of the study was to develop attention test for the VitalSky and benchmark its performance with the DelApp-ICU. Apart from the primary goal, this thesis also explored the use of different response devices through which the test can be automated requiring minimal intervention from the nurse and assessed the technical feasibility and usability of the automated setup.

The subjective feedback collected from the participants showed similarity between the DelApp-ICU, DAViS, DelSky. The participants had same scores in all the three tests (DelApp-ICU, DAViS and DelSky) and found the tests to be of nearly same difficulty. However, there is still some scope of improvement in the tests because there were varied opinions about the similarity of the tests with the DelApp-ICU. Some people found the tests to be exactly similar whereas some people found the test to be not similar at all but on average the DelSky and DAViS tests were rated to have some similarity with the DelApp-ICU.

In terms of response devices, there was a clear preference among the participants for the devices in the tactile domain, i.e, keypad followed by tablet and flic button in terms of user interaction, ease of use and also usability.

To conclude, the performance of the participants in the DelSky and DAViS tests was found to be comparable with the performance of the participants in DelApp-ICU. Moreover, the tests were found to have some similarity by the participants. A proof of concept of an automated version of the attention test with the VitalSky was also proposed in this thesis and various methods to interact with the VitalSky were explored. Moreover, the automated test was shown to be technically feasible and can be a viable option in the ICU.

7.3. Limitations

1. The study was not tested with delirium patients in a real clinical setting. The results of the pilot study on healthy volunteers indicate that the performance of the participants was comparable among the DelApp-ICU, DAViS and DelSky. But, the results could vary for delirium patients in ICU. Thus, the

DelSky and DAViS tests should be tested with delirium patients in the ICU.

2. The pilot study was limited by the maximum number of volunteers that can be recruited. The maximum number of participants that can be recruited for the pilot study was 10. A higher population would have allowed to have stronger insights about the performance of the participants in the different attention tests and the differences between the tests would have been more substantial. A higher population would have also allowed for a better statistical analysis.

3. Some of the questions of the subjective feedback were later found to be not useful and ignored from the analysis and some questions did not receive the intended answer. This could have been avoided by performing a pilot of the subjective feedback questionnaire before the actual pilot study.

7.4. Future Work

1. Since the participants were confused by the visual feedback provided on the VitalSky for the flic buttons, the visual feedback should be changed to have some other shape or color than the target and distractors to avoid the confusion.

2. Audio feedback should be incorporated along with the visual feedback because the participants seemed to be confused in absence of an audio feedback instructing them when to respond.

3. For the test to be completely autonomous, it needs to be investigated that how to determine the moment to start the attention test and how it can be communicated to the patient that the attention test is about to start. Similarly, it also needs to be sought how to instruct the patients about the task of the test and indicate the ending of the test.

4. Implementation of visual task and behavioural assessment of the DelApp-ICU. In this test, the focus was entirely on the visual sustained attention task of the DelApp-ICU. However, to assess if the patients are fit for the test along with the visual sustained attention task, the behavioural assessment and visual task should be implemented in the VitalSky.

5. For the clinical trials, only a few options of the attention tests will be tested on the patients due to time constraints, the preference should be given to DelSky and DelSky-Smart over DAViS. Additionally, the options of response devices can be reduced by removing the Leap Motion from the list of response devices.

6. Sensors can be used to determine the position of the bed of the patient and adapt the position of the visual content accordingly. This will be helpful to ensure that the visual content still remains in the field of view of the patients even though the position of the bed is changed.

7. Using a neural network and training it for the application of processing audio responses would provide greater accuracy and more flexibility. This will allow to train the model to detect different languages, be more interactive and also tailor according to the healthcare industry standards.

8. Storing the test score of the patients in a database and using data analysis to analyze the fluctuation in the score and therefore the mental status of the patients. This could help in providing insights like if the patient experiences delirium at a particular time of the day or week or if the surgery or treatment that a patient went through induced delirium or increased the severity of delirium.

9. The VitalSky can be used to provide simple exercises for the patients like breathing exercises, simple mobility tasks or even games. In this way, the patients feel more engaged and connected to their surroundings.

Bibliography

[1] Intensive care delirium severity checklist worksheet, https://uploads-ssl.webflow.com/5b0849daec50243a0a1e5e0c/5bad3e04cc14601f7c2dc05f_2013-Tufts-ICU-Delirium-Screening-Checklist.pdf.

[2] Trial making test application, https://www.neura.edu.au/apps/trail-making-test/.

[3] R.-Y. Wang, J.-H. Zhou, Y.-C. Huang, and Y.-R. Yang. Reliability of the chinese version of the trail making test and stroop color and word test among older adults. *International Journal of Gerontology*, 12(4):336–339, 2018.

[4] Computerized version of the spatial span test, https://www.cambridgebrainsciences.com/science/tasks/spatial-span.

[5] A. K. Agarwal, S. I. Shehnaz, R. Khanam, and M. Arifulla. Coffee in class: an alternative to animal experiments in pharmacology? *Alternatives to Laboratory Animals*, 43(2):P17–P21, 2015.

[6] L. J. Brown, C. Fordyce, H. Zaghdani, J. M. Starr, and A. M. MacLullich. Detecting deficits of sustained visual attention in delirium. *J Neurol Neurosurg Psychiatry*, 82(12):1334–1340, 2011.

[7] Z. Tieges, A. Stíobhairt, K. Scott, K. Suchorab, A. Weir, S. Parks, S. Shenkin, and A. MacLullich. Development of a smartphone application for the objective detection of attentional deficits in delirium. *International psychogeriatrics*, 27(8):1251–1262, 2015.

[8] S. Ouimet, B. P. Kavanagh, S. B. Gottfried, and Y. Skrobik. Incidence, risk factors and consequences of icu delirium. *Intensive care medicine*, 33(1):66–73, 2007.

[9] E. W. Ely, A. Shintani, B. Truman, T. Speroff, S. M. Gordon, F. E. Harrell Jr, S. K. Inouye, G. R. Bernard, and R. S. Dittus. Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. *Jama*, 291(14):1753–1762, 2004.

[10] Z. Tieges, J. J. Evans, K. J. Neufeld, and A. M. MacLullich. The neuropsychology of delirium: advancing the science of delirium assessment. *International journal of geriatric psychiatry*, 33(11):1501–1511, 2018.

[11] S. K. Inouye, M. J. Schlesinger, and T. J. Lydon. Delirium: a symptom of how hospital care is failing older persons and a window to improve quality of hospital care. *The American journal of medicine*, 106(5):565–573, 1999.

[12] E. W. Ely, S. K. Inouye, G. R. Bernard, S. Gordon, J. Francis, L. May, B. Truman, T. Speroff, S. Gautam, R. Margolin, et al. Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (cam-icu). *Jama*, 286(21):2703–2710, 2001.

[13] S. K. Inouye. The dilemma of delirium: clinical and research controversies regarding diagnosis and evaluation of delirium in hospitalized elderly medical patients. *The American journal of medicine*, 97(3):278–288, 1994.

[14] Z. Lipowski. Delirium (acute confusional states). *Jama*, 258(13):1789–1792, 1987.

[15] S. K. Inouye, M. D. Foreman, L. C. Mion, K. H. Katz, and L. M. Cooney Jr. Nurses' recognition of delirium and its symptoms: comparison of nurse and researcher ratings. *Archives of internal medicine*, 161(20):2467–2473, 2001.

[16] M. Élie, F. Rousseau, M. Cole, F. Primeau, J. McCusker, and F. Bellavance. Prevalence and detection of delirium in elderly emergency department patients. *Cmaj*, 163(8):977–981, 2000.

[17] J. Kean and K. Ryan. Delirium detection in clinical practice and research: critique of current tools and suggestions for future development. *Journal of psychosomatic research*, 65(3):255–259, 2008.

[18] M. G. Cole. Delirium in elderly patients. *Focus*, 12(2):7–332, 2005.

[19] F. Marquis, S. Ouimet, R. Riker, M. Cossette, and Y. Skrobik. Individual delirium symptoms: do they matter? *Critical care medicine*, 35(11):2533–2537, 2007.

[20] N. Gupta, J. de Jonghe, J. Schieveld, M. Leonard, and D. Meagher. Delirium phenomenology: what can we learn from the symptoms of delirium? *Journal of psychosomatic research*, 65(3):215–222, 2008.

[21] N. A. O'Regan, D. J. Ryan, E. Boland, W. Connolly, C. McGlade, M. Leonard, J. Clare, J. A. Eustace, D. Meagher, and S. Timmons. Attention! a good bedside test for delirium? In *Journal of neurology, neurosurgery, and psychiatry*, 2014.

[22] J. W. Devlin, Y. Skrobik, C. Gélinas, D. M. Needham, A. J. Slooter, P. P. Pandharipande, P. L. Watson, G. L. Weinhouse, M. E. Nunnally, B. Rochwerg, et al. Clinical practice guidelines for the prevention and management of pain, agitation/sedation, delirium, immobility, and sleep disruption in adult patients in the icu. *Critical care medicine*, 46(9):e825–e873, 2018.

[23] E. W. Ely, R. Margolin, J. Francis, L. May, B. Truman, R. Dittus, T. Speroff, S. Gautam, G. R. Bernard, and S. K. Inouye. Evaluation of delirium in critically ill patients: validation of the confusion assessment method for the intensive care unit (cam-icu). *Critical care medicine*, 29(7):1370–1379, 2001.

[24] S. K. Inouye. The confusion assessment method (cam): training manual and coding guide. *New Haven: Yale University School of Medicine*, 2003.

[25] Cam-icu frequently asked questions.

[26] A. Luetz, A. Heymann, F. M. Radtke, C. Chenitir, U. Neuhaus, I. Nachtigall, V. von Dossow, S. Marz, V. Eggers, A. Heinz, et al. Different assessment tools for intensive care unit delirium: which score to use? *Critical care medicine*, 38(2):409–418, 2010.

[27] L. McNicoll, M. A. Pisani, E. W. Ely, D. Gifford, and S. K. Inouye. Detection of delirium in the intensive care unit: comparison of confusion assessment method for the intensive care unit with confusion assessment method ratings. *Journal of the American Geriatrics Society*, 53(3):495–500, 2005.

[28] M. M. van Eijk, M. van den Boogaard, R. J. van Marum, P. Benner, P. Eikelenboom, M. L. Honing, B. van der Hoven, J. Horn, G. J. Izaks, A. Kalf, et al. Routine use of the confusion assessment method for the intensive care unit: a multicenter study. *American journal of respiratory and critical care medicine*, 184(3):340–344, 2011.

[29] N. Bergeron, M.-J. Dubois, M. Dumont, S. Dial, and Y. Skrobik. Intensive care delirium screening checklist: evaluation of a new screening tool. *Intensive care medicine*, 27(5):859–864, 2001.

[30] J. P. L. M. Carvalho, A. R. P. d. Almeida, and D. Gusmao-Flores. Delirium rating scales in critically ill patients: a systematic literature review. *Revista Brasileira de terapia intensiva*, 25(2):148–154, 2013.

[31] C. D. Tomasi, C. Grandi, J. Salluh, M. Soares, V. R. Giombelli, S. Cascaes, R. C. Macedo, L. de Souza Constantino, D. Biff, C. Ritter, et al. Comparison of cam-icu and icdsc for the detection of delirium in critically ill patients focusing on relevant clinical outcomes. *Journal of critical care*, 27(2):212–217, 2012.

[32] K. Plaschke, R. Von Haken, M. Scholz, R. Engelhardt, A. Brobeil, E. Martin, and M. A. Weigand. Comparison of the confusion assessment method for the intensive care unit (cam-icu) with the intensive care delirium screening checklist (icdsc) for delirium in critical care patients gives high agreement rate (s). *Intensive care medicine*, 34(3):431–436, 2008.

[33] S. Grover and N. Kate. Assessment scales for delirium: a review. *World journal of psychiatry*, 2(4):58, 2012.

[34] M. M. van Eijk, R. J. van Marum, I. A. Klijn, N. de Wit, J. Kesecioglu, and A. J. Slooter. Comparison of delirium assessment tools in a mixed intensive care unit. *Critical care medicine*, 37(6):1881–1885, 2009.

[35] T. N. Tombaugh. Trail making test a and b: normative data stratified by age and education. *Archives of clinical neuropsychology*, 19(2):203–214, 2004.

[36] P. T. Trzepacz, R. W. Baker, and J. Greenhouse. A symptom rating scale for delirium. *Psychiatry research*, 23(1):89–97, 1988.

[37] Z. Tieges, L. J. Brown, and A. M. MacLullich. Objective assessment of attention in delirium: a narrative review. *International journal of geriatric psychiatry*, 29(12):1185–1197, 2014.

[38] L. J. Beglinger, J. A. Mills, S. M. Vik, K. Duff, N. L. Denburg, M. T. Weckmann, J. S. Paulsen, and R. Gingrich. The neuropsychological course of acute delirium in adult hematopoietic stem cell transplantation patients. *Archives of clinical neuropsychology*, 26(2):98–109, 2010.

[39] P. Pompei, M. Foreman, C. K. Cassel, C. Alessi, and D. Cox. Detecting delirium among hospitalized older patients. *Archives of Internal Medicine*, 155(3):301–307, 1995.

[40] K. J. Christensen, K. M. Bettin, K. M. Jilk, D. T. Weldon, and J. R. Mach Jr. Neuropsychological tests for monitoring delirium severity in elderly patients. *The American Journal of Geriatric Psychiatry*, 4(1):69–76, 1996.

[41] S. T. O'Keeffe and M. A. Gosney. Assessing attentiveness in older hospital patients: global assessment versus tests of attention. *Journal of the American Geriatrics Society*, 45(4):470–473, 1997.

[42] K. M. Bettin, G. J. Maletta, M. W. Dysken, K. M. Jilk, D. T. Weldon, M. Kuskowski, and J. R. Mach Jr. Measuring delirium severity in older general hospital inpatients without dementia: the delirium severity scale. *The American Journal of Geriatric Psychiatry*, 6(4):296–307, 1998.

[43] L. J. Brown, H. S. Ferner, J. Robertson, N. L. Mills, R. Pessotto, I. J. Deary, and A. M. MacLullich. Differential effects of delirium on fluid and crystallized cognitive abilities. *Archives of gerontology and geriatrics*, 52(2):153–158, 2011.

[44] W. D. W. A. I. Scale and T. Edition. San antonio. *TX: The Psychological Corporation*, 1997.

[45] D. J. Meagher, M. Leonard, S. Donnelly, M. Conroy, J. Saunders, and P. T. Trzepacz. A comparison of neuropsychiatric and cognitive profiles in delirium, dementia, comorbid delirium-dementia and cognitively intact controls. *Journal of Neurology, Neurosurgery & Psychiatry*, 81(8):876–881, 2010.

[46] F. Jabbar, M. Leonard, K. Meehan, M. O'Connor, C. Cronin, P. Reynolds, A. M. Meaney, and D. Meagher. Neuropsychiatric and cognitive profile of patients with dsm-iv delirium referred to an old age psychiatry consultation-liaison service. *International psychogeriatrics*, 23(7):1167–1174, 2011.

[47] D. J. Meagher, M. Moran, B. Raju, D. Gibbons, S. Donnelly, J. Saunders, and P. T. Trzepacz. Phenomenology of delirium: assessment of 100 adult cases using standardised measures. *The British Journal of Psychiatry*, 190(2):135–141, 2007.

[48] A. K. Rajlakshmi, S. K. Mattoo, and S. Grover. Relationship between cognitive and non-cognitive symptoms of delirium. *Asian journal of psychiatry*, 6(2):106–112, 2013.

[49] F. W. Black and R. Strub. *The mental status examination in neurology*. FA Davis Company, 1977.

[50] D. Adamis, D. Meagher, O. Murray, D. O'neill, E. O'mahony, O. Mulligan, and G. McCarthy. Evaluating attention in delirium: A comparison of bedside tests of attention. *Geriatrics & gerontology international*, 16(9):1028–1035, 2016.

[51] R. P. Hart, J. L. Levenson, C. N. Sessler, A. M. Best, S. M. Schwartz, and L. E. Rutherford. Validation of a cognitive test for delirium in medical icu patients. *Psychosomatics*, 37(6):533–546, 1996.

[52] B. Agrell and O. Dehlin. The clock-drawing test. *Age and ageing*, 27(3):399–403, 1998.

[53] D. Adamis, D. Meagher, D. O'Neill, and G. McCarthy. The utility of the clock drawing test in detection of delirium in elderly hospitalised patients. *Aging & mental health*, 20(9):981–986, 2016.

[54] G. L. Bryson, A. Wyand, D. Wozny, L. Rees, M. Taljaard, and H. Nathan. The clock drawing test is a poor screening tool for postoperative delirium and cognitive dysfunction after aortic repair. *Canadian Journal of Anesthesia/Journal canadien d'anesthésie*, 58(3):267–274, 2011.

[55] P. Tognoni, A. Simonato, N. Robutti, M. Pisani, A. Cataldi, F. Monacelli, G. Carmignani, and P. Odetti. Preoperative risk factors for postoperative delirium (pod) after urological surgery in the elderly. *Archives of gerontology and geriatrics*, 52(3):e166–e169, 2011.

[56] A. Smith. The serial sevens subtraction test. *Archives of Neurology*, 17(1):78–80, 1967.

[57] R. Katzman, T. Brown, P. Fuld, A. Peck, R. Schechter, and H. Schimmel. Validation of a short orientation-memory-concentration test of cognitive impairment. *The American journal of psychiatry*, 1983.

[58] G. Blessed, B. E. Tomlinson, and M. Roth. The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. *The British Journal of Psychiatry*, 114(512):797–811, 1968.

[59] J. L. Rudolph, R. N. Jones, L. J. Grande, W. P. Milberg, E. G. King, L. A. Lipsitz, S. E. Levkoff, and E. R. Marcantonio. Impaired executive function is associated with delirium after coronary artery bypass graft surgery. *Journal of the American Geriatrics Society*, 54(6):937–941, 2006.

[60] J. Meagher, M. Leonard, L. Donoghue, N. O'Regan, S. Timmons, C. Exton, W. Cullen, C. Dunne, D. Adamis, A. J. Maclullich, et al. Months backward test: a review of its use in clinical studies. *World journal of psychiatry*, 5(3):305, 2015.

[61] L. J. Ball, G. B. Bisher, and S. J. Birge. A simple test of central processing speed: an extension of the short blessed test. *Journal of the American Geriatrics Society*, 47(11):1359–1363, 1999.

[62] N. A. O'Regan, D. J. Ryan, E. Boland, W. Connolly, C. McGlade, M. Leonard, J. Clare, J. A. Eustace, D. Meagher, and S. Timmons. Attention! a good bedside test for delirium? *J Neurol Neurosurg Psychiatry*, 85(10):1122–1131, 2014.

[63] G. Bellelli, A. Morandi, D. H. Davis, P. Mazzola, R. Turco, S. Gentile, T. Ryan, H. Cash, F. Guerini, T. Torpilliesi, et al. Validation of the 4at, a new instrument for rapid delirium screening: a study in 234 hospitalised older people. *Age and ageing*, 43(4):496–502, 2014.

[64] C. Green, K. Hendry, E. S. Wilson, T. Walsh, M. Allerhand, A. M. MacLullich, and Z. Tieges. A novel computerised test for detecting and monitoring visual attentional deficits and delirium in the intensive care unit. *Critical care medicine*, 45(7):1224, 2017.

[65] E. Tang, M. Laverty, A. Weir, E. S. Wilson, T. S. Walsh, M. Allerhand, S. Parks, A. M. MacLullich, and Z. Tieges. Development and feasibility of a smartphone-based test for the objective detection and monitoring of attention impairments in delirium in the icu. *Journal of critical care*, 48:104–111, 2018.

[66] W. S. Helton, R. P. Kern, and D. R. Walker. Conscious thought and the sustained attention to response task. *Consciousness and cognition*, 18(3):600–607, 2009.

[67] A. P. Zanesco, E. Denkova, S. L. Rogers, W. K. MacNulty, and A. P. Jha. Mindfulness training as cognitive training in high-demand cohorts: An initial study in elite military servicemembers. *Progress in brain research*, 244:323–354, 2019.