DEFINING SHIP AUTONOMY BY CHARACTERISTIC FACTORS

Ørnulf Jan RØDSETH[1*]

[1] SINTEF Ocean AS, Trondheim, Norway
*E-mail: OrnulfJan.Rodseth@sintef.no

Abstract

Several papers have proposed ways to define levels of autonomy (LOA), i.e. how responsibility is shared between an automation system and a human when the automation system to some degree can operate independently of the human. The different LOAs are developed for different purposes and therefore often have different priorities for what parameters to use in the classification. This makes them difficult to compare. The main purpose of this paper is to propose a more general characterization scheme that can be used to clarify the description of the different LOAs and their practical meaning. It is suggested that the characterization should be done in terms of three main factors: Operational complexity; degree of automation; and operator presence. It is also proposed to subdivide operator presence into two parameters: responsibility onboard and responsibility in remote control center. The other objective of the paper is to propose the concept of "constrained autonomy" as one specific degree of automation. This proposal also includes an argument for how constrained autonomy may be able to improve human-automation interfaces and simplify testing of autonomous control functions.

Keywords: *Autonomous ship; Unmanned ship; Autonomy levels, MASS, Constrained autonomy*

1. Introduction

The International Maritime Organization (IMO) use the term MASS (Maritime Autonomous Surface Ship) for ships that fall under provisions of IMO instruments and which exhibit a level of automation that is not recognized under existing instruments. In the following, the term "autonomous ship" is used to mean a merchant ship that has some ability to operate independently of a human operator. This covers the whole specter from automated sensor integration, via decision support to computer-controlled decision making. An "unmanned ship" is a ship without crew that needs a certain degree of autonomy, e.g. when communication with a remote control center is lost. The term "automation" will be used for the abilities of a control system to implement functions that commonly have been done by humans. The term "autonomy" will be used to characterize a ship system (the ship and its support system onboard and on shore) that to some degree can operate independently of human operators. Thus, automation is necessary to implement autonomy, but automation will not in itself lead to autonomy. This distinction is not always used in the literature and when in the following levels of autonomy (LOA) is discussed, this refers to autonomy as defined above. To avoid confusion, the term "degree of automation" (DA) will be used to refer to automation levels.

Definitions of LOAs have received much attention over the years. One survey of the most common taxonomies investigated 14 different classification schemes [1]. This is to be expected as different LOA are developed for different purposes and with varying emphasis on different properties of autonomy or automation. However, this also introduces significant ambiguity in definitions of LOA, which can be exemplified by the preliminary classification levels used in the IMO regulatory scoping exercise for MASS [2].

This paper proposes another and complementary way to characterize ship autonomy. The main purpose is to provide better terminology to discuss and describe ship autonomy. This also includes the means to provide a more unambiguous definition of what different LOAs mean in terms of general ship autonomy [2]. The proposed taxonomy is still evolving. It is based on the original NFAS classification of ship autonomy [3], as updated in [4] and later adjusted in [2].

1.1 Levels of autonomy and human factors

Kaber [5] discusses LOA in the context of human-automation interaction (HAI). The paper particularly looks at the use of LOA as taxonomies to structure and improve analysis of human performance, workload, and situation awareness as well as some of the problems that this may cause.

Supporting systematic analysis of HAI is one important application of LOA. Many LOA are variants of the "pipeline" model of human information processing [6] as illustrated in Figure 1.

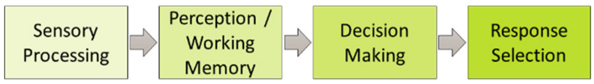

Figure 1. Human information processing pipeline [6]

In this context, higher degrees of automation imply that the human responsibility shifts to the right in the pipeline. This is not necessarily supporting autonomy as it always requires a human to be active or stand-by in the pipeline. Shifts to the right can also be argued to negatively impact situational awareness and cause "out of the loop" problems since it may mentally distance the operator from the physical reality. This can be a problem when the situation rapidly changes. The pipeline principle is also apparent in some LOAs proposed for the maritime sector, e.g. [7] and [8].

One of the points that Kaber makes is that LOA may not always be an accurate tool to predict human behavior or system performance. The introduction of increasing automation changes the way human and machine interact in many ways that may not always be captured by a given LOA classification [5], e.g.:

- *Complacency*: The system operator is satisfied with performance but may lack awareness of other safer or more efficient methods of operation.
- *Satisficing*: This represents an aversion to effort, by accepting a solution that meets minimum requirements, rather than looking for better solutions that are known or suspected to exist.
- *Lack of situational awareness, i.e. out of the loop problems*: Operator does not fully understand the situation and cannot determine the correct actions when human attention is required.

One can again argue that these problems may be related to the pipeline type HAI where humans are supposed to be continuously supervising the system, while most of the work is done by the automation. This paper will not try to fully describe or solve these problems, but section 6 will suggest that constrained autonomy may have the potential to reduce some of these problems.

1.2 One or several definitions of levels of autonomy

LOAs can be used to analyze human and system performance and, thus, support the design of new autonomous systems. There are also other applications of LOA, e.g. for use in safety and risk analysis [13] or as standard terminology in industrial developments [11]. It is not likely that it is possible to define one common LOA, even for one specific type of vehicle, system or application: Different LOAs will be needed for different purposes. However, this may cause confusion among practitioners related to what is meant by the different LOA classifications and how to compare them.

Thus, there is a need to find a more consistent terminology to discuss and compare the different implementation approaches to ship autonomy and scenarios, including definitions of LOA [2]. This may require a reduced focus on the details of HAI and an increased focus on higher-level design options in autonomous ship implementations. This does not mean that the HAI is not essential to safe and efficient operation of autonomous ships and that the classic LOA schemes are less important. Contrarily, it can be argued that a two-level approach to taxonomies can make it easier to apply more detailed and differently targeted LOAs more accurately on the different parts of the autonomous control system, including the human-machine interfaces.

2. Why a special taxonomy for ships?

Regarding autonomy, merchant ships have some special properties that tend to distinguish them from many other autonomous systems. These properties are common to other types of automated vehicles and systems that have been called "*industrial autonomous systems"* [10]. These are systems with high value, high damage potential and absolute requirements to cost-effectiveness. Specifically for autonomous ships, this means:

- Ships are high value assets with a potential for creating dangerous situations for itself or for other ships. *One needs to be conservative in how autonomy is applied.*
- Ship voyages can last for weeks, with long stretches passing by without creating any disturbing events for a remote human operator. However, when situations change, rapid responses may be required. *One needs to be careful in how autonomy is applied*
- Ship operations are very cost sensitive which requires strict cost controls both in capital investments and in operational costs. *One needs to be cost-effective when applying autonomy*.

In addition to these issues, and partly because of them, ships also lend themselves to significant flexibility in how they are operated. At deep sea and in calm weather the ship may be virtually fully autonomous, while in constricted waters and heavy traffic it may have to be fully under direct and remote human control. It is expected that remote control centers will be extensively used to supervise the autonomous ships, which adds flexibility in task assignments between ship and control center. One also has flexibility in how the voyage is planned, e.g. transiting constricted waters may be planned to avoid meeting larger ships. One may also make use of land based infrastructure as it is often better to complement ship sensors with more accurate situation information from shore. Some coastal state authorities are also investigating how to better support autonomous ships. This is in relation to pilotage, vessel traffic services (VTS) and other more general monitoring and control services. This can provide further flexibility in autonomous ship operations.

It is of particular interest to make autonomous ships fully unmanned. This removes living quarters, life support systems and much safety equipment from the ship, saving money, increasing cargo capacity and reducing environmental footprint. Unmanned operation is also of interest if one wants to create a fleet of more frequent and flexible transport systems, where crew costs on multiple ships would otherwise be prohibitively expensive [9].

These features mean that ship autonomy is qualitatively different from autonomy as proposed for, e.g. cars [11]. In cars, it is commonly assumed that one always has a person in the car that can either take part of the control or act as backup in case of failures. Trips are significantly shorter and there is normally not a remote control center.

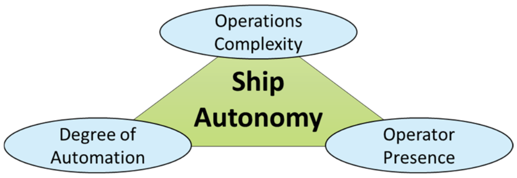

Figure 2. Autonomy as a function of three main factors

To cater for these issues, this paper proposes to define ship autonomy as a function of three main factors as illustrated in Figure 2. This is similar to the ALFUS framework [12], but differs in that "Mission Complexity" and "Environmental Difficulty" are merged into "Operations Complexity" and that "Human Independence" is split into "Degree of Automation" and "Operator Presence". These differences are directly resulting from the special requirements in ship autonomy as discussed above.

"Operations Complexity" and "Degree of Automation" will be defined in terms of "Operational Design Domain" (ODD) and "Dynamic Ship Tasks" (DST) as discussed in sections 3 and 4. "Operator Presence" covers both crews on board and in remote control center as well as sharing of responsibility between the crews. This is discussed in section 5.

3. Operations Complexity

The degree of automation is in principle independent of task complexity. A thermostat is arguably fully autonomous, in the sense that it keeps room temperature constant without needing any human interaction at all. However, its automation is not very complicated. Thus, there is a need to specify operations complexity in addition to the degree of automation.

To capture the complexity of the operations that needs to be performed by a ship, we propose to use the concept of the "Operational Design Domain" (ODD) from the SAE J3016 standard for cars [11]. The operational domain can be seen as multi-dimensional state-space $\boldsymbol{O}$ containing all expected system states s. Each s is normally a vector, but for simplicity, the vector sign is omitted in the following. Note that voyage complexity, level of autonomy and other factors will vary over a ships voyage, i.e. its time t and position p, so the ODD should be defined over the time and positions that are *relevant* for the ship's voyages, i.e. as $\boldsymbol{O}(t, p)$. This was discussed in [3] and [4] and will be returned to later in this section.

In addition to the ODD, one also needs to define a fallback space $\boldsymbol{F}$ that is entered when the ODD is exceeded, as illustrated in Figure 3.

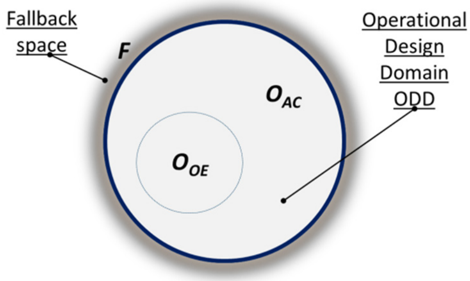

Figure 3. Operations Complexity

The ODD is used to define the operational envelope for the autonomous ship and its support systems. This includes all *anticipated* failures that should be handled by the ship and its systems. The ODD must consider the different ship functions and their constraints, e.g.:

- *G:* Geographic constraints.
- *T:* Other ship and vessel traffic constraints.
- *W:* MetOcean conditions and visibility.
- *V:* Vessel characteristics and capabilities.
- *N:* Navigational infrastructure, aids to navigation, etc.
- *P:* Port facilities and support.
- *C:* Communication system facilities.
- *M:* Mission characteristics.
- *R:* Minimum safety and performance requirements.
- *O:* Other constraints.

The ODD will consist of two sub-spaces, $\boldsymbol{O}_{AC}$, normally controlled by the automation systems, and $\boldsymbol{O}_{OE}$, controlled by the operator exclusively, see Figure 4. Any state that the ship can enter that is not defined in these two spaces will implicitly be in the "fallback space" $\boldsymbol{F}$, which will be discussed later.

$$\boldsymbol{O}(t,p) = \boldsymbol{O}_{AC}(t,p) \cup \boldsymbol{O}_{OE}(t,p) = \{O_G, O_T, O_W, O_V, O_N, O_P, O_C, O_M, O_R, O_O\} \quad (1)$$

In DNVGL class guidelines for autonomous and remotely operated ships [13], the ODD will form part of the CONOPS ("concept of operations"). The other part of the CONOPS will be the "Dynamic Ship Tasks" (DST): The set of tasks that the operator or the automation system must be able to execute to satisfy the ODD. A similar approach was used in the MUNIN project, where Unified Modelling Language (UML) "use cases" described the ODD and DST [14].

Also, the DST concept has been adapted from the SAE J3016 standard [11]. However, DST has been renamed from "Dynamic Driving Tasks" (DDT) to reflect the wider range of ship functions beyond only "driving", including, e.g. energy production, propulsion systems, safety functions and cargo supervision. The word "Dynamic" is used to highlight that these tasks are associated with the execution of a voyage and not the strategic planning or re-planning that takes place before and possibly during the voyage.

The DST is divided into Operator Exclusive tasks (OE-DST), that only a human operator is expected to be able to perform, and Automatic Control tasks (AC-DST) that the automation system is designed to handle, but where an operator, if available, can intervene. This is illustrated in Figure 4.

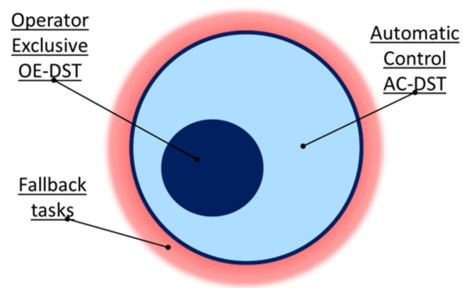

Figure 4. Division of task responsibility

There should be a defined function f in either of the DST function spaces $\boldsymbol{T}_{AC}$ and $\boldsymbol{T}_{OE}$ for each state s that can be entered into in $\boldsymbol{O}$, as shown in equation (2). The function may be to do nothing if the state is safe and sustainable.

$$\forall s \in \boldsymbol{O}(t,p): \exists f(s) \in \{\boldsymbol{T}_{AC} \cup \boldsymbol{T}_{OE}\} \quad (2)$$

DST fallback functions are required to handle reachable states outside the ODD and to bring the ship to an acceptable minimum risk condition (MRC). MRC does not include states related to anticipated failures or problems that are defined to be handled within the ODD. These shall be included in the ODD itself. One may also need more than one fallback strategy and/or MRC to handle different types of problems or situations. DNVGL stipulates that there generally will be two levels of fallback [13]:

- *Minimum Risk Condition (MRC):* Possibly recoverable states where some ship systems remain operational. These may be sustainable only for a limited time.
- *Last Resort MRC (LR):* These shall be sustainable until the ship can receive external assistance but may not be automatically recoverable.

The MRC is similar to what is sometimes called "Fail to Safe", but the MRC makes it explicit that one cannot in general define a completely safe state for a ship and for all external forces or threats that can be applied to it.

Fallback functions and MRC require that all relevant or "reasonably foreseeable" fallback cases can be identified, i.e. all transitions from states in $\boldsymbol{O}$ that can end up outside $\boldsymbol{O}$, as shown in (3). The fallback space $\boldsymbol{F}$ can be divided into $\boldsymbol{F_{LR}}$ for last resort fallbacks and $\boldsymbol{F_{MRC}}$ for normal fallbacks. As for $\boldsymbol{O}$, $\boldsymbol{F}$ will normally also be defined over the relevant time and positions for the ship's voyages.

$$\boldsymbol{F}(t,p)=\{\boldsymbol{F}_{MRC}\cup\boldsymbol{F}_{LR}\}=\{\hat{O}_G,\hat{O}_T,\hat{O}_W,\hat{O}_V,\hat{O}_N,\hat{O}_P,\hat{O}_C,\hat{O}_M,\hat{O}_R,\hat{O}_O\}$$
$$\forall s\in\boldsymbol{O}(t,p),\hat{s}\notin\boldsymbol{O}(t,p),s\rightarrow\hat{s}:\ \hat{s}\in\boldsymbol{F}(t,p)$$
$$\forall s\in\boldsymbol{F}(t,p):\ \exists f(s)\in\{\boldsymbol{T}_{MRC}\cup\boldsymbol{T}_{LR}\} \quad (3)$$

As for the DST functions, one should make sure that all states in $\boldsymbol{F}$ that can be entered into, can be mapped to a (possibly empty) fallback function.

The last resort MRC could be defined by creating states *s* that have the property that all events will still keep the system state in the same state *s*, see equation (4). The last resort states will normally be general and defined for a wide range of MRC states and events.

$$\forall\ s\in\boldsymbol{F}_{LR}(t,p):\ s\rightarrow s \quad (4)$$

Note that the definition in eq. (4) does not allow the ship to recover from a last resort state automatically. It will need some form of operator intervention to return to the ODD. This may be reasonable, as last resort actions most likely will be very basic, e.g. dropping anchor or shutting down propulsion. However, other last resort definitions may be necessary in some cases.

Figure 5 shows the general transitions between states in the DST and MRC, including the dashed arrow showing return to DST. As this transition often is operator controlled, it can be expected to go to OE-DST before going to AC-DST.

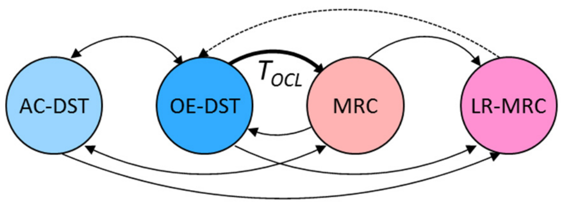

Figure 5. Transition between states in DST and MRC

In general, there will be a maximum time T_{OCL}, the operator command latency (*latency* [2], or *command latency* [13]), between entering the operator exclusive task state and before a fallback action is activated and the system automatically moves to an MRC. T_{OCL} will therefore be the operator's maximum allowed response time and will be discussed in sections 4 and 6.

Thus, the complexity of operation can be qualitatively described by defining the ODD, DST, DST Fallback and MRCs. It is obvious that the complexity rapidly can increase, as there are a high number of interacting components both in the $\boldsymbol{O}$ and $\boldsymbol{F}$ spaces. This has impacts on the testing and approval costs and on establishing acceptance criteria for the safe use of the autonomous ship. Cost-effective development and deployment of autonomous ships requires that the complexity is kept under control. This will be discussed further in section 6.

4. Degree of Automation

Section 1.1 discussed various forms of Human-Automation Interface (HAI) based levels of autonomy. It is clear that these LOA have important applications in various types of human factors research, but for a more general characterization of ship autonomy, they are often too detailed and also to focused on the human interaction issue. While the LOAs could be translated to correspondingly detailed degrees of automation, this paper proposes to limit the degree of automation to five basic cases, where the degrees are defined by the need for a human to be present at the control station, not what type of task the human has, when at the station. This classification was originally suggested in [3], further developed in [4] and updated in [2]. Although the word "autonomous" is used in the names of some of these degrees, this only means that the degree can be used to support system autonomy, not that the system becomes autonomous by the automation degree alone.

- *DA0 – Operator controlled*: Limited automation and decision support is available, as on today's merchant ship. The human is always in charge of operations and need to be present at controls and aware of the situation at all times.
- *DA1 – Automatic*: More advanced automation, e.g. dynamic positioning, automatic crossing or auto-berthing is used. Crew attention is required to handle problems defined in ODD as OE-DST, such as object classification and collision avoidance. The human may use *own judgement* as to how long he or she may be away from the control position. For automated fjord crossing in good weather, little traffic and in sheltered water, the operator may, e.g. be away from the controls for several minutes.
- *DA2 – Partial autonomy*: The degree of automation is higher than for DA1, but there are still limits to the automation system's capabilities. These limits are not defined or constrained (see DA3), so the human operator must still use *his or her judgement* as to the required attention level. However, it is assumed that the need for attention is lower than for DA1.
- *DA3 – Constrained autonomous*: The degree of automation is similar to DA2, but system capabilities are now constrained by programmed or otherwise defined limits. The limits are set to enable the system to detect that limits are exceeded and to alert the operator in time before operator intervention is required. After an alert, the operator has a maximum time of T_{OCL} before he or she needs to be back at controls and take remedial actions.

- *DA4 – Fully autonomous:* The ship automation can handle the full ODD without any intervention from crew. Fallback and MRC can be activated without crew assistance. Crew is not required at any time at control stations.

With respect to the system's ability to operate without human support, DA2 is very similar to DA1, although DA2 is meant to represent a significantly higher degree of automation. As the automation system's limitations in neither degree are known to the operator or the system itself, it cannot be known when operator intervention is necessary, and in both cases the operator may need to intervene on short notice. Section 6 will elaborate on this issue and argue why this degree of automation in most cases should be avoided.

It can also be argued that DA0 and DA1 could be merged into one. With the example of an autopilot in calm weather, little traffic and open sea, one can easily argue that this should fall under category DA1. However, to be able to distinguish between current ship systems and new more advanced systems that still require continuous human presence at controls, it is considered useful to maintain this division.

Thus, the focus of the proposed automation degrees is to indicate to what degree an operator needs to be present at a control position for safe operation of the ship. This is in contrast to some other classifications that put more emphasis on human factor issues and where in the HAI pipeline the automation system and operator meet.

Table I. Sheridan's ten levels versus five degrees (DA0 – 4)

Sheridan's ten LOA / DAn	**0**	**1**	**2**	**3**	**4**
human does the whole job up to the point of turning it over to the computer to implement	■				
computer helps by determining the options	■				
computer helps to determine options and suggests one, which human need not follow	■				
computer selects action and human may or may not do it	■				
computer selects action and implements it if human approves		■	■		
computer selects action, informs human in plenty of time to stop it				■	
computer does whole job and necessarily tells human what it did					■
computer does whole job and tells human what it did only if human explicitly ask					■
computer does whole job and decides what the human should be told					■
computer does the whole job if it decides it should be done, and if so , tells human, if it decides that the human should be told					■

If the proposed five DA are compared with a more traditional LOA scale, e.g. the one proposed by Sheridan [15], a comparison can be set up as in Table I. Dark cells show the correspondence between the two definitions.

DA0 covers the first four levels in the scale and this is caused by different human positions in the decision pipeline, while the human still needs to be continuously available to make the final decision. A similar phenomenon can be seen in the DA4 classification of levels 7 to 10, where the human is not actually needed in any of the cases and can be taken completely out of the control loop if desired. Note also that DA1 and DA2 maps to the same level 5.

5. Operator Presence

Making a ship completely unmanned removes the need for a hotel section, much of the safety equipment and enables completely new ship designs. This has a significantly higher potential for changing and improving maritime transport than just increasing automation onboard [9]. However, unmanned ships will in most cases require some form of remote monitoring and control. Having operator backup on shore is also becoming more common for manned ships, either for maintenance purposes [16] or for general supervision of ship operations [17]. Thus, different combinations of ship and remote control, is a very relevant direction for conventional as well as autonomous ships.

Thus, the third factor that is characterizing ship autonomy is the location and availability of the operators. In the following, crew availability will be designated as *LnRn,* where *L* and *R* means local on ship and remote off ship respectively and *n* specifies the degree of crew availability in each location:

- *0 – None*: There is nobody available to man the control position. The control position may not exist.
- *1 – Backup*: Person(s) are available to operate the control position but are not present. They need to be called and there will be a latency, that generally should be lower than T_{OCL} (see Figure 5), before they can resume full control.
- *2 – Available*: Person(s) are available at the control position but is not actively controlling the ship. In a remote control center, they may control or monitor other ships [18]. The operator can regain full control of the ship at short notice (usually shorter than T_{OCL}).
- *3 – In control*: Person(s) are at the control position, are in charge of and actively controlling the ship.

When two control positions are in use, it is also necessary to define what position is in charge and has the main responsibility for acting when something requires operator attention. The other position will be responsible for fallback responses in case the primary fails to act or acts in a way that put the ship at danger. It will normally be the control position with the highest crew availability that is in charge. In the nomenclature used in this paper, the position in charge will be marked with a star: *L0R2** means that the remote crew is in charge of an unmanned ship; or *L3*R2* means that the ship crew is in charge and directly in control, but have a shore crew actively monitoring operations, without taking control except in exceptional situations. The exception to this is in fully autonomous mode, when no one is in charge and the notation would be just *L0R0.*

For fully unmanned ships, it may also be relevant to have two remote control centers to provide fallback solutions in case the primary center is disabled for some reason. On many of today's ship one will also routinely have a situation where the manned bridge monitors the periodically unmanned engine room (L1R3*). The same classification and considerations can be used also in these cases. Note also that each different shipboard function, e.g. bridge watch, engine control or cargo monitoring may use different control level and operator presence.

Table II. Examples of operator presence and matching DA

Presence	DA	Explanation
L3*R0	DA0	Today's ship bridge
L1*R2	DA0	Periodically unmanned engine
L3*R2	DA0	Supervised operation from shore
L2*R0	DA1	Auto-crossing / berthing
L0R3*	DA0/1	Unmanned, full remote control
L1*R0	DA3	Periodically unmanned bridge
L1R2*	DA3	Supervised, unmanned bridge
L0R2*	DA3	Constrained autonomous
L0R1*	DA4	Monitored, fully autonomous
L0R0	DA4	Fully autonomous

Table II shows some examples of relevant combinations of operator control with degrees of automation (DA). Many other combinations can be envisaged for more specialized operations than those listed here.

The total number of crew, locally or remote, will often be an important cost factor and it is expected that this will be minimized where possible, i.e. one will normally prefer a high degree of automation to reduce crew. DA3 will allow a lower crew number in remote control than DA2 as each crew may monitor more ships.

6. Benefits of constrained autonomy

As pointed out in section 4, the maximum time allowed before a human operator is able to regain control needs to be considered before selecting degree of automation and human presence. From this it follows that partial autonomy (DA2) will not generally give more benefits with regards to manning levels than automatic (DA1). One should in most cases use constrained autonomy (DA3) instead. This section will extend this argument into two specific areas: A potential for avoiding some human factor problems and improved testability when constrained autonomy is used instead of partial.

6.1 Improved human-automation interface

Section 1.1 discusses some problems associated with human-automation interaction (HAI) that causes problems with operators' reduced vigilance and loss of situational awareness. It could be argued that the pipeline model of HAI (Figure 1) may contribute to this in that it increasingly separates the final human decision making from sensory input when degree of automation increases. An alternative model is the hierarchical control model, e.g. as presented by Brooks [19].

Figure 6. Layered control system according to Brooks [19]

This model increases degree of automation by adding new abstraction levels onto the closed loop control system. This relieves the human of tedious and detailed sensory processing and actions and could enable the operator to concentrate more on higher-level monitoring and control functions, without removing him or her from the sensory input. For a ship, this may, e.g. look like Figure 7.

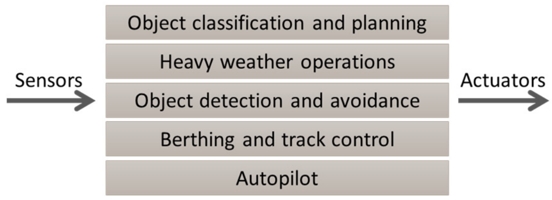

Figure 7. Possible layered control for a ship

For this model, it could be argued that by having humans supervising or acting as backup for higher levels of control, it may offer more deterministic latencies and a better environment for assessing the situation than in the pipeline model. However, checking and verifying this argument is outside the scope of this paper.

As constrained autonomy is defined in this paper, it includes two important principles:

1. The system is able to detect that the ability of the automatic functions will be exceeded; and
2. There is a minimum latency T_{OCL} associated with the automatic transition to MRC, if the operator fails to react to a necessary intervention.

The latter requirements can be expressed in (5), where all possible new states $\hat{s}$ in the Operator Exclusive ODD that can be reached from the current state s in the Automatic Control ODD, need to have a lifetime L greater than T_{OCL}. The lifetime function L estimates how long the new state $\hat{s}$ can safely be maintained without any action from the operator.

$$s \in \boldsymbol{O}_{AC}(t, p), \forall \hat{s} \in \boldsymbol{O}_{OE}, s \rightarrow \hat{s}: \; L(\hat{s}) \geq T_{OCL} \qquad (5)$$

This means that, if the system is correctly designed, it is possible to alert the operator when situations escalates beyond the limits of automation and still give the operator enough time to gain a good understanding of the situation and to plan and execute remedial actions.

It is reasonable to believe that this approach should remove or reduce some of the human factor issues discussed in section 1.1. However, it should be noted that neither the author nor his team has examined this effect in any relevant studies.

6.2. Improved testability of functions

Traditionally, ship safety has been approved as shown on the left hand of Figure 8. Equipment and systems have been approved as "safe to operate" by being tested according to performance requirements, e.g. from class societies or the International Maritime Organization (IMO), and corresponding test standards. The crew has been certified to "operate safely", based on training and qualification requirements from IMO and national authorities. Included in the latter part is the assumption that by giving the crew proper training, the operators will be able to understand what actions to take when new situations or variants of trained situations are encountered.

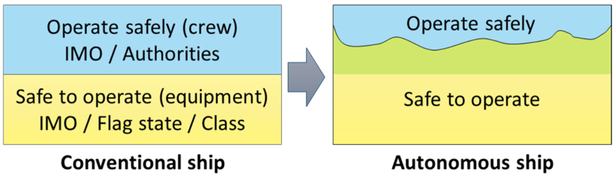

Figure 8. Approving safety of ships

With introduction of automation that takes over all or some of the crew functions, the "safe to operate" approach must now be applied to much more complex problems, where previously one relied on human flexibility and competence to handle variants in dangerous or demanding situations. In addition, one may also find that the limit between operators and automation's responsibilities becomes more blurred.

Traditionally, technical equipment has been tested according to test standards with prescriptive and quantitative test criteria. This will not generally be possible with the new control functions and a more risk based approach will be necessary. This has also been recognized by, e.g. DNV GL [13] and Class NK [21]. However, the agreement is that specific tests and test criteria will still be needed, also for complex functions.

Testing advanced autonomous ship functions poses several problems. The first is to determine the acceptance criteria for the overall functionality of the autonomous ship [20]. DNV GL argues for "equivalent safety" (to manned ships), but this is also problematic, e.g. in that it is not known how many incidents or accidents are *averted* by the crew on today's manned ships as indicated on the right hand side of Figure 9.

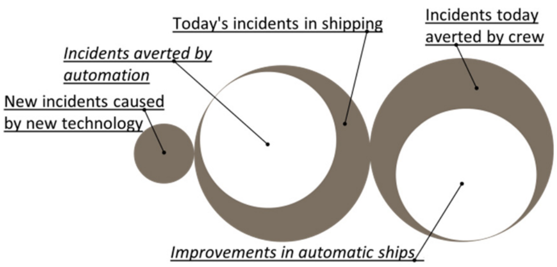

Figure 9. Assumed risk picture for autonomous ships [20]

The next problem is to define exactly what automatic functions are required, i.e. define the AC-DST. Following this, one will need to define performance criteria for all the identified functions.

Finally, one will also need to determine how to test the new functions. All autonomous ship functions are expected to be software intensive and proper procedures for specification and development of the software will obviously be required. However, it is unlikely that critical functions like object detection and classification ("lookout") and collision avoidance can be approved by other than test-oriented procedures that are able to verify the functions' capabilities in simulated or real scenarios.

Testing automated functions with complex behavior will obviously be difficult. As discussed in section 3 and shown in equations (1) and (2), the combinatorial complexity of the operational spaces $\boldsymbol{O}$ and $\boldsymbol{F}$ are normally very high, and this will make it very difficult to design test cases with sufficient functional coverage. In addition, one can assume that there are many rare combinations, particularly in anti-collision cases. To address this, one may use stochastic simulations, but it will still be a challenge to get statistically significant test results for rare cases that may never have been encountered before. Combining constrained autonomy with human assistance for the rare cases may help to solve this problem. Constraints may be defined to move the rarest and possibly unforeseeable cases out from the test space of the automated functions and leave them to the human operator.

Another problem is the use of, e.g. deep learning or similar techniques to train object classification or anti-collision algorithms. Deep learning implicitly makes it impossible to define a priori what the limits of the functions' capabilities are. This makes it difficult or even impossible to device good test-procedures when the test space is virtually unlimited. Constrained autonomy may again be used to define operational limits and by that enable test procedures with sufficient coverage for a more limited operational space.

However, it is not yet clear how easy it is to identify and implement the operational limits required by constrained autonomy. Object detection and classification limits may be relatively easy to quantify, e.g. by defining specific weather and environmental limits for reliable identification of objects of given sizes and types. Anti-collision capabilities may similarly be constrained by overall situational parameters, e.g. number of and distance to other ships as well as navigational complexity due to geographic restrictions.

Given the testability problems discussed above, it may be necessary to introduce constrained autonomy to get sufficiently well tested automated systems for partly or periodically unmanned ships. It will not solve the testing problem for fully autonomous functions, but it can certainly contribute to increased experience and knowledge about testing autonomous functionalities.

7. Summary and conclusions

Classification schemes for autonomous vehicles and ships in particular, are an area in constant development. As the autonomous ship is a relatively new concept, mostly dating back to the MUNIN project from 2012 [22], it is also natural that classifications evolve.

The usefulness of the different classifications of autonomy depends on their intended application. Developing classification schemes for human factors in teleoperations is different from classifications related to testing and approval of industrial autonomous systems like ships. The characterization scheme proposed in this paper is not a replacement for existing or other emerging classifications but is intended as a means to give a better and more formalized description of what the different types of classification actually means in terms of automation of a ship. The three factors used in the characterization, complexity, degree of automation and operator presence is thought to be sufficient and necessary to describe what is meant with an autonomous ship on a high level.

Limiting the degrees of automation to five is also thought to be sufficient for high-level descriptions of autonomous ships. The focus on the operator's need to be present is also suggested as a better metric than, e.g. metrics involving the location in the human information processing pipeline.

Furthermore, the concept of constrained autonomy may be necessary to develop fully or periodically unmanned ships. It can be argued that both human factor issues and testability problems can be significantly reduced by restricting the automation to functions that allows the system to alert operators sufficiently long before they need to take control of the ship. This may be an important step towards reliable and safe semi-autonomous ships and a contribution also to development of fully autonomous ships.

The ideas presented here are still being developed and it is hoped that this paper can be a small contribution to the ongoing discussions and future work.

Acknowledgements

The work presented here is in part funded by the Norwegian Research Council through project number 267860 (SAREPTA). It has also been supported by the Korea-Norway cooperation project on "Development of testbed guideline for autonomous ship".

References

[1] Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied ergonomics*, *53*, 190-202.

[2] ISO, Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS), Proposal for a classification scheme for degrees of autonomy, IMO Document MSC 100/5/1, August 2018.

[3] Rødseth Ø.J. and Nordahl H., *Definitions for Autonomous Merchant Ships*, Norwegian Forum for Unmanned Ships, Version 1.0, 2017-10-10. Available from http://nfas.autonomous-ship.org/resources.html (March 2018).

[4] Rødseth, Ø.J.; Nordahl, H.; Hoem, Å.S. Characterization of Autonomy in Merchant Ships. In: Proceedings of MTS/IEEE Oceans'18, Techno-Ocean 2018 - OTO'18. Kobe, Japan: IEEE 2018 ISBN 978-1-5386-1653-6.

[5] Kaber, D.B. (2018). Issues in human–automation interaction modeling: Presumptive aspects of frameworks of types and levels of automation. *Journal of Cognitive Engineering and Decision Making*, *12*(1).

[6] Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3).

[7] Lloyds Register (2016). *Cyber-enabled ships, ShipRight procedure – autonomous ships*. First edition, July 2016, A Lloyd's Register guidance document.

[8] Australia, Denmark, Finland, France and Turkey (2018), *Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS), Comments on document MSC 100/5*, IMO Document MSC 100/5/6, 12 October 2018.

[9] Rødseth, Ø.J (2018). Assessing Business Cases for Autonomous and Unmanned Ships. In: *Technology and Science for the Ships of the Future. Proceedings of NAV 2018: 19th International Conference on Ship & Maritime Research*. IOS Press 2018 ISBN 978-1-61499-870-9.

[10] Grøtli E.I., Reinen T. A., Grythe K., Transeth A. A., Vagia M., Bjerkeng M., Rundtop P, Svendsen E, Rødseth Ø.J., and Eidnes G. (2015), SEATONOMY: Design, development and validation of marine autonomous systems and operations", In: *Proc. of the MTS/IEEE OCEANS'15.*

[11] SAE J3016 (2016): *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems*, Revision September 2016, SAE International.

[12] Huang, H. M., Pavek, K., Novak, B., Albus, J., & Messin, E. (2005). A framework for autonomy levels for unmanned systems (ALFUS). *Proceedings of the AUVSI's Unmanned Systems North America*, 849-863.

[13] DNV GL (2018), Class Guideline - Autonomous and remotely operated ships, DNVGL-CG-0264, September 2018.

[14] Rødseth, Ø.J., & Tjora, Å. (2014). A risk based approach to the design of unmanned ship control systems. *Maritime-Port Technology and Development 2014*.

[15] Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Report from Massachusetts Institute of Technology, Cambridge Man-Machine Systems Lab.

[16] Latarche M. (2017). The pros and cons of remote monitoring on ship, *ShipInsight*, May 2017.

[17] Erdmann F. (2016). Hamburg-based Carnival Maritime takes over monitoring and support of Carnival UK fleet, *Seatrade Cruise News*, October 2016.

[18] MacKinnon S.N., Man Y., Baldauf M. (2015). *Final report Shore Control Centre*, MUNIN Report D8.8. September 2015.

[19] Brooks, R. A. (1991). New approaches to robotics. *Science*, *253* (5025).

[20] Porathe, T., Hoem, Å. S., Rødseth, Ø. J., Fjørtoft, K. E., & Johnsen, S. O. (2018). At least as safe as manned shipping? Autonomous shipping, safety and "human error". *Safety and Reliability–Safe Societies in a Changing World. Proceedings of ESREL 2018, June 17-21, 2018, Trondheim, Norway*.

[21] ClassNK (2018). *Guidelines for Concept Design of Automated Operation / Autonomous Operation of Ships* (provisional Version), May 2018.

[22] Rødseth, Ø.J., & Burmeister, H.C. (2012). Developments toward the unmanned ship. In *Proceedings of International Symposium Information on Ships–ISIS* (Vol. 201, pp. 30-31).