Abstract – The functional role of a data fusion system is to provide timely and accurate information to the user. At the lowest level, the fusion system filters relevant incoming data and at the highest level, the fusion system presents actionable information to the user. In many cases, the fusion system design is dependent on user capabilities and knowledge. Thus, only presenting the fused data might not enhance operator performance, while presenting all the multimodal information might overload the operator's cognitive capabilities. In this paper, we use the JDL-User model to design an interface which utilizes image fusion and multimodal characteristics to enhance user performance. The design allows for the user to passively monitor and actively interact with the fusion system with controls for image overlays, queries and audio warnings to support unattended areas, and an eye tracker to improve attention. Implemented in JAVA, we present a tracking display design that utilizes a multimodal interface to (1) take advantage of the human's ability to fuse multisensory information, (2) facilitate situational awareness, and (3) enhance the user's cognitive task analysis.

Keywords: Fusion, Multimodal interface design, Situational Awareness, Cognitive task analysis

1 Introduction

The key to a fusion system design is the ability of the system to convey information desired by the user. The integration of senses by the human is natural form of data fusion and extensions to human processing include computer displays to guide attention, facilitate efficient task analysis, and provide cues for situation awareness. Passive computer presentations might accurately present the fused data, however if the mental model of the human is not consistent with the display, the system usability might not improve performance. One way to improve operator performance is to design an interactive interface which allows the human the capability to interact and refine the display. As an example, the human might be able to refine the data collection so as to get more image perspectives for a 3D target presentation or multiple looks on the same target [1]. When the human is allowed to interact with the system, the human is transformed from an passive operator to an active user. Another way to improve operator performance is to design a multimodal interface.

Multimodal interface design (MID) is a key concept that leverages the human's ability to filter incoming data. MIDs augment display information with control inputs, audio warnings, and some proprioceptive information. Such examples are audio warnings for threat assessment and control interactions which afford the human the ability to cue in on featural data [2] or drill down to areas of interest. Additionally, we might think of classifying image fusion selections as an extension to the MID design. For instance, allowing the user to select between electrical-optical(EO), synthetic-aperture radar (SAR), and infrared (IR) for day, high-altitude, and night missions, respectively, would enhance the multimodal visual capabilities (see Waxman[3]). If the user was allowed to select independent channels or a fused presentation would assist in time-constrained targeting for varying weather conditions. Other examples include overlaying vehicle separation plots on detected tracks for target tracking and identification.[4] A key component of the MID is the ability to facilitate the human's cognitive ability to fuse multiple image presentations for task completion.

Task analysis is a method to assess the actions taken by the user. Contrasted with a task analysis is a cognitive task analysis (CTA) [5] which seeks to look at the mental model of the operator. As an example, for a time constrained targeting scenario, the task analysis would look at the key strokes of the human interacting with the display while the CTA would address what tasks and thoughts were used in accurately identifying the target. For fusion design, it is important to design the system so that the user can accomplish the cognitive task. It is a question of “why” and not “how” the user selects which multimodal inputs to use and fuse. To explore the cognitive fusion task, the JDL-user model (Shown in Figure 1) is used to explore the interactions between the human and the fusion system.

A general theme of the fusion community to explore the higher levels of information fusion which includes the human as a key component in the fusion system. Blasch
and Plano [2] have described the interaction between the human and all the levels of the JDL fusion model with the highest level (Level 5) being that of User refinement. The model defends the case that the human is always active in a fusion system design. Such examples are selecting incoming data (level 0), choosing objects of interest (level 1), defining an area of coverage (Level 2), determining the level of threats (level 3) and refining the location of the sensor placements (level 4). Examples of user issues includes object assessment by Waxman [3] with EO, IR and SAR data, and medical image fusion [6] for health monitoring and diagnosis, situation awareness by Endsley [7], Wright [8] and Roy et al [9, 10], and information querying by Anken [11], Masters[12], and Fransson[13].

This paper addresses a multimodal fusion design for the Wright State Aegis Simulator Platform (WASP). The rest of the paper is as follows: Section 2 details human-computer interactive fusion. Section 3 details the problem statement, multimodal display, performance metrics, and a task analysis used to improve operator performance for a targeting task. Section 4 summarizes the need for a focused strategy for fusion designing the human into the fusion process.

## 2 Human-Fusion Interaction

One fusion model that has been utilized by fusion community is the JDL fusion model, which has been adapted to highlight the need for human refinement [1]. In the JDL-User model, the focus is the human’s ability to perform active reasoning capabilities over the other fusion stages.

![Figure 1. JDL-User Model](image)

In the model, level 5 – User refinement is described as an element of Knowledge Management: adaptive determination of who queries information and who has access to information. (e.g. information operations) and adaptive data retrieved and displayed to support cognitive decision making and actions (e.g. altering the sensor display). While many fusion models explore the capabilities of the human, it is not a question of where the boxes go, but what is the focus of fusion system. A computer is good at processing large amounts of object data while a human is good at inferencing over the data. Thus, the key to the human refinement question is the “what framework the human has to inference over the data and what capabilities allow for active reasoning?”

To develop the user refinement capabilities to include in a design, three issues are paramount (1) situational awareness, (2) cognitive task analysis, and (3) multimodal interface.

### 2.1 Situational Awareness

The Human in the Loop (HIL) of a semi-automated system must be given adequate situation awareness (SA). According to a pioneer and continued leader in the SA literature, Endsley stated that "Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.” [7]. This now-classic model translates into 3 levels:

- **Level 1 SA - Perception of environmental elements**
- **Level 2 SA - Comprehension of current situation**
- **Level 3 SA - Projection of future states**

Operators of dynamic systems use their situation awareness in determining their actions. To optimize decision making, the SA provided by a system should be as precise as possible as to the objects in the environment (Level 1 SA). A situation awareness approach should present a fused representation of the data (Level 2 SA) and provide support for the operator's projection needs (Level 3 SA) in order to facilitate operator's goals.

Issues associated with the awareness of the human are (1) workload, (2) attention and (3) trust [2]. In the case of workload, we minimize the amount of information the human must process which also helps to highlight the need for different types of “cause” information for the user. Attention is the concept in which the user is directed to the relevant information at the correct time. Trust includes the ability to convey information on the computer as extended sensory information, about the truth of the world. False data would hinder the direct connection between the user and the world around them.

### 2.2 Cognitive Task Analysis

To reduce the set of information to a dimensionally attractive set of information requires that a hierarchy of needs is determined for the human processed by the fusion system. A hierarchy of needs helps the fusion system match the output to the human’s current work-oriented goals, to perform active reasoning over the data.
in a useful order. In this case, we can postulate the needs by addressing the impact assessment (Level 3) questions through employment of a number of work domain analysis [14] and task analysis methodologies [5, 14, 15]. Some of the needs that can be included in a hierarchy include:

Goals:
- Context
- Inputs expected to accomplish work
- Timing
- Relationships
- Outcomes and products of work

Concepts:
- Objects – Number and types
- Threat – whether harmful, passive, or helpful
- Location – close or far
- Basic primitives – features
- Existence known or unknown – (i.e. new objects)
- Moving or static

Processes:
- Measurement system reliability (uncertainty)
- Ability to collect more data
- Delays in the measurement process

It is the user that must interact with a fusion system to determine the priority. Supervisory control concepts are relevant to this interaction, especially in choosing a level of automation. Management by consent [16] is likely to be tested as a paradigm by which the fusion system proposes information for decision support, upon which the human explicitly consents to actions in order to proceed to the next step. The priority of information presented is algorithmically related to the information sought, but to provide the human with the ontological ability to understand the world, they must have the ability to choose or select both the objects of interest as well as the processes from which the raw data is converted to the fused data. The success with which a human can understand the origin of the information or the pedigree of the data and information, in a timely enough manner, under high workload conditions, with stressors upon maintaining situation awareness, is a critical issue in the processing and presentation of fused information.

2.3 Multimodal Interface

To actively reason over a complex set of data, a usable set of information must be available from which the human can act on. The display for fusion information is based on the ontology chosen which includes (1) cognitive reasoning, and (2) display issues such as semantics, efficacy, and a taxonomy.

2.3.1 Cognitive Reasoning

User cognition follows a variety of mechanisms: monitoring situations actively or passively, planning by reacting to new data or proactively controlling the fusion system. Thus, the human also must choose cognitive roles to receive and convey information, such as Monitor (active versus passive) or Planner (proactive vs. reactive.)

When a human interacts with a system, it is important the ability of reasoning be available for the fusion system. The human has the abilities to quickly reduce the search space of the fusion system and hence, guide the fusion system process. Such an example is when the human cues a fusion system to look for an object in a certain area of the earth. Developing a framework of a user refinement system must have some sort of semantics or interface actions that allow the system to coordinate with the user. Such an example is a query system in which the human seeks questions and the system translates these requests into actionable items. [13]

2.3.2 Display / Interface

The display interface is key to allowing the user to have control over the data collection and fusion processing. [17] Without designing a display that matches the cognitive perception of the information, it is difficult for the user to reason over the fused result. While many papers and books address the interface issues (i.e. see Billings [16], Wickens [18], Ware [19], Preece [20]), it is of concern for the fusion community to address the cognitive user issues to ensure that the fusion system designed is to emulate the functional roles required of the fusion system.

**Semantics** associated with a display design are important to ensure that the information displayed is understood cognitively. The use of the correct terminology/vernacular that is similar as experienced in training is critical to comprehension. In the development of the MID, we developed a query systems so as to allow the user to engage with the system to ask for information as well as determining the display categorical labels.

**Efficacy** is an important concept for the performance questions involving a human and a computer system. If the user has confidence and reliability in the fusion system, the fusion system will be further utilized for its capabilities. Without having trust in automation, the effects of the inferencing over the data will not result.

A **taxonomy** is a classification of something – such as algorithms, processes, and things. In order to execute the prioritization information, there has to be some sort of classification. The difficulty is that each user, unless trained, sees the world slightly differently and thus, has a different classification for objects and processes. In order to develop efficient and effective fusion systems, it is important to develop a taxonomy that relates to the ontological perspectives of what the human expects in the world. Such an example is tracking a target. The expected processes are detection, recognition, classification, and identification. If the object is identified, then it is assumed that the target is detected and not the other way around. Seeking a valid taxonomy of
objects and fusion strategies will help the user in determining what role they should play in the active reasoning over data and fused information.

3 Problem Description and Proposal

The real-world operator of a shipboard target tracking system, simulated here by the Wright-State Aegis Simulator Platform (WASP) [21], must strive for perfect task execution in a dynamic and high-stakes environment. Performance of human operators as well as the system can suffer from discrepancies between optimum and real execution results. The errors in performance (discrepancies), measured against goal objectives, must be reduced. The interfaces to the WASP must be augmented to provide a cleaner, more efficient, robust interaction between the WASP interface and operators’ sensory modalities. The goal of such augmentation is therefore to provide better situation awareness in the simulation environment, hence reducing objective time to goal on specific performance parameters. These include efficiency in visual search, attention allocation, keystrokes and other tactile manipulations.

3.1 Problem Statement

The purpose of WASP Fusion Multimodal Interface Design (MID) project is to model an interface that would improve the human problem-solving performance in real-time, in a multi-task domain, where the human operator must share attention among competing task demands. The problem is to improve the interactions of an operator with an existing system interface. The problem space is a simulation system of an Aegis-class ship-based legacy system, which supports operators in tracking and intercepting targets in the local zone around a deployed naval battle group. The current primary simulation interface between the user and the radar data and tasks is called WASP. As an interface, WASP currently fulfills all required tasks. Figure 2 is a screenshot of WASP.

The goals of the controller/user, are to defend his ship and not to engage friendly aircraft. The performance of the operator is measured explicitly against a set of prescribed actions, called Rules of Engagement (ROE).

The Rules of Engagement are:
1. Engage Track within 20 NM
2. Illuminate Track if within 30 NM
3. Issue Level1 warning if track within 50 NM
4. Issue Level2 warning if track within 40 NM
5. Issue Level3 warning if track within 30 NM
6. Recall CAP if further than 256 NM
7. Dispatch CAP is closer than 20 NM
8. Assign primary Track ID for all unknown Air Tracks
9. Assign track designation

Figure 3 illustrates the Identify Friendly, Foe, Neutral (IFFN) task flow, per ROE.

**Time windows** are the mechanism used to measure performance of the system/user against the (Rules of Engagement) ROE. A ‘time window’ defines a time interval within which the operator is required to complete the task, and accounts for the corresponding operator response (urgency and priority) to the task demand. Things that impact performance within time windows are training, concurrent demands, and facility of the system to aid the operator in completing the task under given conditions within that time window.

Although WASP fulfills all required tasks, the problems with using WASP for attaining performance goals:

---

1 WASP is a branch of a program called GT-ASP, created at Georgia Institute of Technology. This program has been used as a teaching and research design tool at Wright State University, as well as other universities, due to the dissemination of it by original designers, such as Dr. Ling Rothrock, who is now at Penn State University.
• The high number of steps the user takes to get these tasks accomplished successfully
• The low fidelity of graphics and interface controls and general data displays, given advances in modern software and multi-modal capabilities.
• The lack of best-practices cognitive support such as attention-directing features,

The motivation for updating the WASP design exists since performance success is critical to battle group safety and overall success, and because the tasks have narrow time windows for success.

3.2 Human Factors Input to the Design

While WASP fulfills base requirements, the Fusion MID design proposed is predicated upon understanding that human factors inform good engineering design of complex system interfaces. An analysis of layout, functionality, navigability and interactivity of the existing interface comprises the bulk of this project. Some human factors analyzed and proposed in the MID interface are:

• User requirements of the system, particularly in the visual realm but encompassing the work and context
• Limitations and ranges of the human visual, auditory, and tactile capabilities and systems
• Limitations and ranges of current technology to provide advanced modal input to users
• Interaction parameters which can be optimized between the human and the software display
• Current trends in interaction design for efficient mapping of familiarity/problem solving skills
• Conceptual model [20]: e.g., terrain data might suggest different conceptual designs of multiple screens.
• Conditions under which display is used
  o Context, situation, environment
  o Ambient light, ergonomic positioning of the display
  o Duration (time) of sessions with the system
  o Other visual and ergonomically related tasks concurrent with using the system

3.3 WASP Fusion MID

We propose an augmented system that supports the same functions and tasks, while providing better human-computer interaction based on a human factors analysis. This design architecture, Figure 4, introduces these components to the existing system:

• Audible warning system generator
• Haptic pointing device for pointing
• Fused video display for range overlays
• Fused video display for target identification/filtering
• Visual enhancement to radar display
• Eye tracking to synergize system and ID alerts to current operator status

Figure 4. WASP Fusion MID Architecture

• Auxiliary monitor/panel with touch screen and new menu options

Figure 5 shows the image fusion enhancement to the WASP design. In the image fusion component, the various display types (indicating different target alternatives) can be filtered and fused using selected threshold and operator selections. The user has the ability to switch back and forth and determine the change detection capability of the system as to the updated target information. Not pictured is the audible information from which warnings are shown. The audible warnings can be filtered on the display to increase the attention of the user to focus on the targets of importance.

Figure 5. WASP Fusion MID Video Enhancement.

Figure 6 shows the range overlays to distinguish the spatial aspects of the targets, one of the earliest human factors developments in such displays. One of the difficulties in a drill-down display is determining the location of the targets of interest. For the proposed design, the idea of a topological map was included to display the distance of the targets to give the operator a quicker and better understanding of the spatial information.
of the targets. While the range concentric circles would enhance operator understanding, other ideas proposed included putting color-coded information on the targets to determine the last update of information.

Correctly performing the required action within a given time window (or window of opportunity) is recorded. The information for measuring this parameter is given in WASP_Log.txt file. [21]

The variables are:
- action_exists
- close_time
- CNNO
- open_time
- platRule_index
- rule_ID
- window_ID

These parameters permit quantification of the number of mistakes, the total number of actions missed, late command, early, and omitted actions.

The other parameters of performance will be derived from interactions of the operator with WASP in relation to data in the following tables of the database file:

- Sounds
- Mouse Events
- Keyboard Events
- Global Environment
- Local Environment
- Actions
- Sequences
- Static information

The error discrepancy detector is constrained as a theoretical structure using the performance data as its input. For this design, the focus centered upon improving the performance of the user in executing a task scenario within the expected relevant time windows. The loci for intervention of the augmented system include the following performance measures (dependent variables during later evaluation/validation-of-design phase):

**Accuracy**
- Increase number of identifications that fall in the expected timeframe (windows)
- Increase accuracy of ID: decrease number of incorrect identifications
- Increase the number of warnings given before targets enter inside range
- Increase confidence / trust of identifications
- Increase spatial awareness of target types

**Time**
- Decrease the time to visually attend to a new target stimulus
- Decrease workload for user selection of targets
- Increase speed of dispatching commands to F-15
- Increase speed of requesting track updates
- Increase speed of initiating a change in the display

Many concepts were considered, from hardware and technology to efficient modality strategies, to promote dynamic task execution in a high-stakes environment. This is the first design iteration, to illustrate concept variants and principles. Table 1 lists several MID concepts, noting the rationale whether a concept was included in the current design proposal. Concepts were screened and rated
against performance requirements, above, and constraints. [22-38]

### Table 4. Alternative MID Concepts Compared

<table>
<thead>
<tr>
<th>Hardware Sols.</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head-Up Display HUD</td>
<td>• Another way to do an overlay</td>
<td>• Hard to align</td>
</tr>
<tr>
<td>Head Mounted Display HMD</td>
<td>• Moves with operator, avoids adding extra panel</td>
<td>• Hard to align</td>
</tr>
<tr>
<td>Haptic Mouse or Joystick</td>
<td>• Feed back</td>
<td>• Difficult Hand-Eye coordination</td>
</tr>
<tr>
<td></td>
<td>• Familiar technology</td>
<td>• External Wires</td>
</tr>
<tr>
<td></td>
<td>• More degrees of freedom if useful</td>
<td>• Can be easily lost/broken</td>
</tr>
<tr>
<td>Thin Transparent Display</td>
<td>• Portable</td>
<td>• New Technology</td>
</tr>
<tr>
<td></td>
<td>• Thin</td>
<td>• May scratch in rough environment</td>
</tr>
<tr>
<td></td>
<td>• Programmable</td>
<td></td>
</tr>
<tr>
<td>Video Mixer</td>
<td>• No additional “visible” space needed</td>
<td>• Installed behind the cabinet – installation and access may be difficult</td>
</tr>
<tr>
<td></td>
<td>• Familiar technology</td>
<td></td>
</tr>
<tr>
<td>Programmable Buttons</td>
<td>• Expert system drives button displays</td>
<td>• None listed</td>
</tr>
<tr>
<td>Audio Feedback</td>
<td>• Familiar</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Easy to deploy</td>
<td>• May have too much “noise” to recognize audio</td>
</tr>
<tr>
<td>3D Audio Input</td>
<td>• Directionality</td>
<td>• Not required</td>
</tr>
<tr>
<td>Multiple Panels</td>
<td>• Allows spatial memory</td>
<td>• Possible distraction of looking around</td>
</tr>
<tr>
<td></td>
<td>• Segregating info</td>
<td></td>
</tr>
<tr>
<td>Touch Screens</td>
<td>• Faster selection</td>
<td>• Spatial acuity required in touch</td>
</tr>
<tr>
<td></td>
<td>• Can make large enough for fingers/gloves</td>
<td></td>
</tr>
<tr>
<td>Stylus/Pen Input</td>
<td>• Precision pointing and selecting</td>
<td>• Technology dependence on manufacturer</td>
</tr>
<tr>
<td></td>
<td>• Reduces size of selection controls</td>
<td>• Not as easy with non-dominant hand as touch-screen</td>
</tr>
<tr>
<td>Cognitive Decision Aids, Automated</td>
<td>• Offloads workload to algorithm</td>
<td>• Requires immense analysis and data</td>
</tr>
</tbody>
</table>

3.5 Task Analysis

One of the methods to determine the human interaction is to perform a task analysis, which is mission or scenario dependent. A task analysis can give insights into the time of processing of information so as to determine when the fused data is more useful for the human. If the needed results are presented too early or too late, it is difficult for the human to utilize the information conveyed. Additionally, there need to mechanisms for the human to be able to query information from the system so as to get the needed inputs.

There are 2 ways to ID an unknown air or surface target:

1. Check the sensor values (EWS)
   a. Displays a sensor value like ARINC564 ON (Sensor value, which were Friendly and which were Hostile)
   b. Look for the sensor value and decide Friendly or Hostile

2. Bridge Or Visual Identification (VID)
   a. This was useful only when the F15’s were very close to the unknown airs or the surface, so that a visual identification is possible. It displays “HSTL or FRND” otherwise it displays “VID NOT POSSIBLE”

Task 1: ID Unknown Air as Assumed Friend using EWS, VID
   EWS (F9) → Update Hooked Track (F5) → Primary ID (F5) → Assumed Friendly (F5) → Execute ID Changes (F1) (The shape of the aircraft changes from an open rectangle to a dotted rectangle)
   Bridge (F11) → VID(F1) → Update Hooked Track (F5) → Primary ID (F5) → Assumed Friendly (F5) → Execute ID Changes (F1) (The shape of the aircraft changes from an open rectangle to a dotted rectangle)

Task 2: ID Unknown Air as Friend using EWS or VID
   EWS (F9) → Update Hooked Track (F5) → Primary ID (F5) → Friendly (F6) → Execute ID Changes (F1) (The shape of the aircraft changes from a open rectangle to a arc)

The other tasks include:

Task 3: ID Unknown Air as Assumed Hostile using EQS, VID

Task 4: ID Unknown Air as Hostile using EWA or VID

Task 5: Identifying Unknown Surface as Assumed Friend using EWS, VID

Task 6: ID Unknown Surface as Friend using EWS, VID

Task 7: ID Unknown Surface as Assumed Hostile (EWS, VID)

Task 8: ID Unknown Surface as Hostile using EWS or VID

Task 9: Changing the course of F15

Task 10: Changing the speed of F15

Task 11: Changing the altitude of F15

Task 12: Changing the course of F15 to where the mouse points

Task 13: Issue Level Warning 1 to an aircraft

Task 14: Issue Level Warning 2 to an aircraft

Task 15: Issue Level Warning 3 to an aircraft

Task 16: Engage Aircraft (within 20 nm from own ship)
Task 18: Zoom In /Zoom out Display From Own ship
Task 19: Enter Range Circle with ref to the own ship
Task 20: Display Range Rings with ref to own ship
Task 21: CAP Too Far (> 250nm from own ship)
Task 22: CAP Too Close (Within 20nm from own ship)
Task 23: Assign Air Track designations

4 Conclusions

As presented in this article, data fusion involves the integration and application of many multimodal data and controls to satisfy a user need. The development of the multimodal interface enhances the Level 5 “user refinement” to interact with the traditional 4 levels of the JDL model. The paper described various methodologies used to increase user situation awareness, assist in the cognitive task analysis, and leverage the human’s ability for multimodal sensory fusion for target tracking. As an example, the user has the ability to select single channel video displays, image overlays, target identification results, and simple queries to the system for region of coverage. Further work includes incorporating fused image results into the user selections as opposed to single channel inputs, ability to control sensor placement for enhanced target identification, and multimodal tools to enhance operator performance.

References