An Interaction Fit Analysis Technique and its Applications

Christopher A. Miller, Jeffrey Rye, Peggy Wu, Eric Engstrom
Smart Information Flow Technologies
211 First St. N., Ste. 300 Minneapolis, MN 55401
{cmiller, jrye, pwu, eengstrom}@sift.net

Abstract
We report on the development and applications of a formal representation and reasoning approach for computationally assessing the degree of “fit” between the information needs of a set of tasks and the information conveyed by a set of displays. The core representation, based on information theoretic assessment of information behavior, is independent of implementation and therefore, is inherently multimodal. This computational approach has been validated and used as the core of an adaptive information management system and, more recently, as the core of a tool to aid designer’s in evaluating candidate display sets in both NASA aerospace and U.S. Navy submarine display domains. The tool, known as the Multi-Media Aid for Interface Design or MAID, is described in depth, along with several instances of its use and application.

Introduction
We will report on past work developing, tuning and applying a core representation for computationally comparing the information (input or output) an operator needs to perform a task and the information provided by a user interface. This representation is based on information theoretic properties, thus it can be, and has been, applied to a wide variety of work domains and information and display types. We developed an initial version of this representation and used it to achieve dynamic (1Hz) adaptation of the cockpit of an attack/scout rotorcraft in response to assessed or reported user intent, framed as desired or needed tasks (Miller and Hannen, 1999; Miller, 2000). More recently, we adapted this approach to the task of evaluating and critiquing display format designs for NASA’s space operations (Miller, Rye, & Wu, 2010) and for consistency checking in naval submarine control and sensor system UI design. The representation and reasoning approach generalizes well to describing information types in procedural domains and the tool can analyze sets of display formats for sets of procedures, propose format improvements against a procedure set, project how changes to procedures will affect the suitability of existing formats, and project how changes to formats will improve or reduce their suitability for given procedures. The mathematical nature of the fit computation and scoring approach makes it amenable to a variety of formal analysis and verification techniques (e.g., singular value decomposition analyses—Miller and Morton, 1994), and its grounding in a task partitioned framework suggests that graph and network analytic techniques offer potential benefits as well (Bryce, 2013).

Interaction Fit—Philosophy and Formalism
Our approach to reasoning about information match has been described extensively elsewhere (Miller, 1999) along with its use in adaptive information management systems such as the Pilot’s Associate (Miller, Shalin, Geddes & Hoshstrasser, 1992) and Rotorcraft Pilot’s Associate (Miller & Hannen, 1999). Thus, we will only briefly summarize it here. It is based on information theoretic principles (e.g., Cover & Thomas, 1991) and its range and sensitivity, as well as displays produced by it, have been validated by expert judges (Miller, Levi and Shalin, 1992).

Our basic approach can be summarized as follows: tasks give rise to information requirements (IRs), thus task knowledge can yield knowledge about the information needed to perform tasks. In fact, IRs can be regarded as “arguments” which are required (to a greater or lesser degree) to the process represented by the task or procedure. IRs can be met, again to a greater or lesser degree, by Presentation Elements (PEs). Insofar as the representation for describing IRs and PEs is the same, reasoning about the degree of match or satisfaction between the information needs of a context or task and the information provided by candidate displays is possible.

The core of our approach is thus a quantitative representation information (and its behavior) as required by task(s) and as provided by candidate presentation elements. We associate IRs with tasks, which helps focus and constrain information behavior to a specific context, but a computational representation for reasoning about infor-
information or interaction ‘fit’ mandates developing a tractable representation for IRs and PEs. The representation we developed is a simple data structure that lists the IRs for each task and the PEs conveyed by each candidate display element, along with some parameters describing how the information is needed or conveyed. The fact that the same, quantitative scales are used for both IRs and PEs makes it feasible to computationally compare sets of IRs against sets of PEs and evaluate the degree of coverage and “fit”.

Thus, an example task like Vectoring for Landing Approach might require IRs like Heading, Bearing, Altitude, Gear_Status, etc. IRs represent abstract information needs, independent of any specific display thereby making them inherently agnostic with regards to the methods or modalities used to convey them.

Simply listing the IRs for a task is, we have found, not sufficient for selecting a good presentation method; we must describe how the information is needed for the performance of the task. We accomplish this tractably using a set of descriptive parameters created by Geddes and Hammer (1991) and refined and formalized by us (Miller, 1999). Each IR in a task is described in terms of five parameter/value pairs referred to as SRBIC parameters—after the first letter in their names. Values for each parameter range from 0-10 and represent the ‘proportion’ of that parameter which is needed for the task. The parameters we have used most frequently (though others are possible) are:

- **Scope**—the extent to which simultaneous access to the total range of possible values for the info element is needed by the task (IR) or provided by a display (PE).
- **Resolution**—the need (IR) or ability (PE) to make fine distinctions in the values of the information.
- **Bandwidth**—the need (IR) or ability (PE) to maintain timely awareness of the information value by frequent sampling and/or rapid uptake. High values imply the need to maintain high currency.

- **Importance**—the relative necessity of this information for successful task performance—as distinct from the relative importance of the task itself.
- **Control**—the need (IR) or ability (PE) to affect the information’s value in addition to monitoring it. If the user simply needs to know an IR value but not to control it, the control value will be 0.

The same parameters are used to characterize the information conveyed by a candidate display. The IR(s) for which a PE satisfies are listed in the PE knowledge structure with SRBIC values representing how that PE conveys that information. The definition and scaling of the parameters is similar and a specific scope value for an IR means that that proportion of possible values for the information need to be presented for likely successful task performance, while a scope value for a PE means that that proportion of the values are presented. Figure 1 illustrates the data structures for Heading for the Vectoring task, along with several candidate PEs showing how the representation can describe both information need and presentation.

Note that the scalar values assigned to SRBICs for both IRs and PEs are not simply opinions or even expert judgments, but are based on the information theoretic properties of the information type (e.g., heading) and how it behaves in the task or display of interest (cf. Miller, 1999). We have worked extensively to refine this process and have achieved both computational approaches to deducing SRBIC values from task descriptions and a detailed training manual for assigning them.

The boxed scores at the bottom of each PE in Figure 2 illustrate a mismatch computation (lower scores represent less mismatch) between the task’s information need and information presented by each PE. The Importance value is used to prioritize or weight the IRs within a task so that the information need (IR) can receive better PE matches. Alternate formulae for calculating the match are possible, but a typical one we have used simply takes the absolute value of the difference of each parameter and sums the results. This penalizes each candidate PE for over or under providing the information in the way it is needed. The scores in Figure 1 indicate that the manipulable dial is the best single PE to meet the heading need for this task—which provides some of the needed Control capability.

This ‘raw’ computation, taken over sets of IRs and candidate PE sets, can be augmented with other logic and knowledge to evaluate goodness of UI composition (Miller, 1999). But even simple set comparisons can provide an information theory-based formal analysis of fit and to answer questions such as:

- Adequacy of candidate display(s) for given procedure
- Need for display modifications given procedure mods
- Tradeoffs between candidate displays for a procedure
- Identification of low- or no-value PEs
- Identification of un- or poorly-covered IRs
- Recommendations of alternate display configurations
This core match computation forms the basis of the MAID (Multi-modal Aid for Interface Design) tool we have implemented for NASA. Though comparatively simple, it allows us to draw some powerful conclusions when taken over even small analytic sets, as will be seen in the next section.

Using the Fit Computation: A NASA Analysis

The MAID tool took the computational fit analysis approach outlined above, which had been created initially to support automated display management for military aircraft, and adapted it to provide analysis support for UI designers in NASA contexts—specifically in designing display configurations to support procedure execution.

The main MAID prototype screen is illustrated in Figure 2. MAID provides access to a Library of elements (procedures, tasks, IRs, PEs, displays and devices, etc.) the analyst can manipulate. MAID also provides a number of analysis tools to configure a comparison between IR and PE sets and then compute and visualize the comparison including: a Mismatch table of information need × presentation matches and scores, a Sequence Graph showing both match score by procedure step and a Gantt chart view illustrate context switching between alternate display formats, and Summary Statistics to characterize the match overall.

MAID analyses begin with a procedure of interest. The procedure provides the task model or workflow though other, traditional sources of task modeling are also usable. The procedure must be characterized in terms of its sequence of IRs, each with its individual SRBIC scores. MAID supports inputting or editing these individually or, more conveniently for larger datasets, creating them in a spreadsheet tool such as Excel and importing them as a .csv file. A similar approach must be taken to characterizing the information provided by the presentation elements contained in the candidate displays. In addition, MAID supports associating presentation elements with the displays or display pages/formats they can be presented on, which is a way of addressing issues associated with “legal” screen layouts and configurations.

Once all of the component knowledge is in place, a library can be opened and analyses constructed. To do so, the user selects a defined set of IRs (corresponding to the tasks of interest) and a set of PEs (corresponding to one or more candidate displays). Of course, additional sets of PEs, IRs and PE may be defined and then used in alternate analyses. Once an analysis is configured, MAID can compute the mismatch table for the comparison between the two sets. The result is as depicted in Figure 2 above. The mismatch table shows, on separate lines, each IR appearing in the set, along with the best matching PE for that IR from the designer specified. For each IR × PE match, a computed mismatch score is shown, along with a version of that score weighted by the importance of the IR. This information is also shown in graphical form in the upper part of the sequence graph at the bottom of Figure 2—where numbers along the x-axis correspond to the sequence of IRs used in the procedure steps (first one needed, then second, etc.)
The Gantt chart in Figure 2 (derived from the work flow for the procedure being analyzed) shows the number of times the operator, using this set of PEs (each of which can be conveyed via one or more defined display page or surface), would be required to shift focus of attention from one display surface to another. Summary statistics at the top of the display show:

- The total mismatch score—useful as an overall comparison metric between this procedure-display set match vs. other, possible alternatives.
- The number and percentage of unmet IRs and unused PEs—good indicators of whether the IR set is well covered and the PE set is used efficiently.
- The average weighted mismatch percent—the average degree to which each PE “misses” completely satisfying its corresponding IR.
- The number of context switches across display pages.

One analysis (illustrated in Figure 2) we demonstrated using MAID came from the design of NASA’s Cockpit Avionics Upgrade (CAU) program—a redesign of the Shuttle’s display suite (McCandless, et al, 2005). That effort identified a case in which dual failures in the helium delivery system for the main engines during shuttle ascent could yield conditions where rote adherence to a diagnostic procedure would result in unintended engine shutdown, mission abort, and emergency landing—all very costly and risky. Specifically, the failed status of a helium isolation valve had to be inferred from indirect indicators in the baseline shuttle displays, but was made explicit in the CAU displays. In simulation contexts, 7 of 8 astronaut crews using the baseline displays failed to realize that following the procedure exactly led to inadvertent shut down and mission abort, while 0 of 8 crews made this mistake when using the more explicit CAU displays. In the MAID analysis illustrated in Figure 2, the need for status information about the He valve corresponds to IR #4 which is highlighted in red in both the table and the sequence graph, is shown as having no matching PE, and receives the highest mismatch score (320) in this analysis.

Expanding and Applying Fit Analyses

Even early on, it was apparent that the computational fit analysis approach provided an opportunity to use powerful mathematical tools to evaluate large scale UI designs. In early work (Miller and Morton, 1994), we illustrated methods for using Singular Value Decomposition techniques across large sets of IRs and PEs to identify “regions” of poor information coverage (i.e., un- or poorly-satisfied IRs) and of under-utilized displays (i.e., PEs which provided little IR-satisfaction value).

In subsequent work, we explored techniques for using the core fit analysis approach to perform consistency evaluations for UIs in naval domains. Consistency management is required when UIs are developed or revised by different teams over time and/or geographic distances. Requirements and style guides are the traditional approach to achieving consistency, but they are time consuming, error prone and costly. We designed a tool for aiding this evaluation process by automatically identifying and scoring consistency across UIs and tasks.

We began by reviewing multiple UI guidelines used in naval systems, such as MIL-STD-1472F and shadowed engineers during the design of a revision to an existing submarine UI—including participation in focus groups with operators. This background enabled us to identify six alternate types of consistency and configuration management of use in UI redesigns and comparisons. These were assessing: (1) Guideline Adherence, (2) Required Coverage, (3) Consistency Within a System, (4) Consistency Across Systems, (5) Consistency Across System Redesigns over Time, and (6) Consistency Across Platforms (e.g., Naval Vessel types). We established that MAID alone could address only one of those types (#2), but with comparatively minor enhancements, an extended version (called IMAGE, for “Interface Management through Automated Generation and Evaluation”) could address all six.

We implemented and analyzed test cases from the submarine UI redesign effort using baseline MAID. This illustrated MAID’s ability to provide analyses in another domain, as well as demonstrating type 2 consistency management (ensuring required coverage) and partial type 4 (consistency across systems). This exercise produced results that both confirmed and extended UI designers’ expectations.

In our designed extensions for IMAGE, most consistency checking takes the form of assessing (and scoring) the degree to which similar tasks are performed in similar ways via alternate UIs (within or across systems, over time, or cross platforms). We designed representational extensions to MAID’s data elements, as well as a scoring algorithm to perform this type of assessment. IMAGE will be able to identify task similarity along four major dimensions: to what degree (1) do tasks involve similar cognitive and/or perceptual operations?, (2) are tasks performed using the same type of information?, (3) are tasks performed with the same equipment or systems?, and (4) are tasks part of the same job class or expected to be performed by the same person? MAID already provides one method of assessing information similarity via its SRBIC parameters, but IMAGE will need to perform appearance and behavior similarity assessment between PEs as well. To handle this, we have also defined a suite of PE characteristics akin to those already captured for graphical widgets (type, color, text length size and font, shape, etc.) and an algorithmic approach to scoring appearance similarity using them.

Basic Consistency Assessment is the core intended function of IMAGE and will be performed by algorithmically scoring similarity between tasks and the UIs designed for them. We propose an algorithm which behaves such that the more similar two tasks are, the more similar their
displays need to be for a low inconsistency score. This is because inconsistent displays used for dissimilar tasks cause minimal impact, but similar displays used for dissimilar tasks produce negative impact—and worse still is when inconsistent displays are used for similar tasks. Applications of IMAGE could include Platform Migration Assistance—adapting a display suite to a new ship or other platform/domain. This builds on the basic consistency assessment approach by identifying task and display similarities across platforms, and also opportunities where displays can or need to change. Regulatory Compliance Checking is another potential IMAGE application and can be evaluated whenever we can write a database-style query against the data elements described for tasks or displays. For example, do all information requirements of type “caution/warning” have an associated visual display element which includes status, nature and condition information (as required by MIL-STD-1472F)?

Acknowledgements

Initial work on the fit analysis representation and reasoning approach was funded under USAF, Wright Laboratories contract #F33615-88-C-1739 and U.S. Army AATD contract #DAAJ02-93-C0008. The NASA MAID tool was developed under a Small Business Innovation Research grant, contract # NNJ06JD65C. The IMAGE design was based upon work supported by the Naval Sea Systems Command under Contract # N65538-09-M-0094. We would like to thank contributors Jeff Groat, Duane Dawson, Cheryl Peskar and Barbara Johnson from Lockheed Martin Maritime Systems and Sensors for assistance in Naval Submarine UI design, and Debra Schreckenghost from TRACLabs for her assistance in designing the NASA CAU examples.

References

Bryce, D. Submitted. Explaining verifier traces with EBL. Submitted for inclusion in Working Notes of the 2014 AAAI-SS on Formal Verification in HMI.


