

A Model for Notification Systems Evaluation—Assessing User Goals for Multitasking Activity

D. SCOTT McCRIKARD, C. M. CHEWAR, JACOB P. SOMERVELL,
and ALI NDIWALANA

Virginia Polytechnic Institute and State University

Addressing the need to tailor usability evaluation methods (UEMs) and promote effective reuse of HCI knowledge for computing activities undertaken in divided-attention situations, we present the foundations of a unifying model that can guide evaluation efforts for notification systems. Often implemented as ubiquitous systems or within a small portion of the traditional desktop, notification systems typically deliver information of interest in a parallel, multitasking approach, extraneous or supplemental to a user's attention priority. Such systems represent a difficult challenge to evaluate meaningfully. We introduce a design model of user goals based on blends of three critical parameters—interruption, reaction, and comprehension. Categorization possibilities form a logical, descriptive design space for notification systems, rooted in human information processing theory. This model allows conceptualization of distinct action models for at least eight classes of notification systems, which we describe and analyze with a human information processing model. System classification regions immediately suggest useful empirical and analytical evaluation metrics from related literature. We present a case study that demonstrates how these techniques can assist an evaluator in adapting traditional UEMs for notification and other multitasking systems. We explain why using the design model categorization scheme enabled us to generate evaluation results that are more relevant for the system redesign than the results of the original exploration done by the system's designers.

Categories and Subject Descriptors: H.1.2 [Models and Principles]: User/Machine Systems—*Human factors*; H.1.1 [Models and Principles]: Systems and Information Theory—*General systems theory*

General Terms: Design, Human Factors, Theory, Evaluation

Additional Key Words and Phrases: Peripheral systems, design model, claims reuse, usability

1. INTRODUCTION

As people everywhere become increasingly more insistent on integrating additional computing tasks with routine and critical daily activities—a behavior

Authors' address: Department of Computer Science, 660 McBryde Hall, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0106; email: {mccricks, cchewar, jsomerve, andiwala}@cs.vt.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 1515 Broadway, New York, NY 10036 USA, fax: +1 (212) 869-0481, or permissions@acm.org.

© 2003 ACM 1073-0616/03/1200-0312 \$5.00

fueled by demand for pervasive and ubiquitous information—there is a growing gap within HCI research. Certainly, much progress has been made toward understanding and refining typical desktop interfaces used during extended periods of concentrated attention with orderly, predictable task action flow. However, different usage situations, expectations, and error consequences govern the growing breed of applications and devices being introduced to support multitasking information demands. Referred to as *notification systems*, these interfaces are generally desired as a means to access valued information in an efficient and effective manner without introducing unwanted interruption to a primary task [McCrickard and Chewar 2003]. They can be found in many implementation forms and on a variety of platforms. Perhaps classic desktop systems are the most readily identifiable—instant messengers, status programs, and news and stock tickers. Other familiar examples such as Weiser’s dangling string representation of network traffic [Weiser and Brown 1996], in-vehicle information systems, ambient media, and multi-monitor displays hint at the potential range of systems.

While use of these systems and the range of solutions has skyrocketed, our ability to scientifically recognize, pattern, and improve success within the HCI community has not kept pace. There are surprisingly few efforts in the literature that effectively evaluate usability of the information and interaction design for notification systems. For example, while some notification systems support collaborative activities and are studied from a CSCW perspective, disparate agendas lead to inconsistent definitions of successful design, inhibiting cross-initiative influence.

An umbrella approach is needed, tying together knowledge and addressing challenges in notification systems design throughout the HCI community. We build on the vision by several recent dialogues within the HCI research community. First, we heartily endorse research approaches such as the systematic establishment of critical parameters (based on Newman [1997]) and reference tasks argued by Whittaker et al. [2000]. Second, we recognize the enormous potential of psychological models applied to create macrotheories that describe interactions within a mental architecture [Barnard and May 1999, 2000], especially as a basis for early-phase, predictive usability evaluations. Finally, we receive inspiration from Sutcliffe’s [2000] notion of “claim families,” advocated as a mechanism for incremental improvement of design guidelines within a scenario-based approach.

With this impetus, we describe a novel approach for modeling and classifying the core concepts within the notification systems spectrum, which allows an improved usability evaluation process to emerge. This article provides a first look at an extensible philosophy for studying other instances of multitasking or collaborative performance. We argue that the models and framework presented here will improve the HCI community’s ability to classify and evaluate existing and emerging notification systems, as well as to catalog information and interaction design guidelines and lessons learned in a cohesive, collective manner. In the next section, we present a more thorough overview of notification systems appearing in recent literature and itemize general user goals, providing motivation and background material for the model we present in Section 3.

2. EMERGENCE OF NOTIFICATION SYSTEMS

In recent years, the research community's pursuit of facilitating the use of multiple, simultaneous information sources is demonstrated by many innovative interface design approaches.

Several efforts can be characterized by their attempt to deliver information of interest with small desktop or task tray icons implemented as sidebar or corner applications, specifically designed to provide glanceable awareness without disturbing other tasks or becoming annoying. The Scope [van Dantzich et al. 2002], Sideshow [Cadiz et al. 2001], and Irwin [McCrickard 1999; McCrickard and Chewar 2003] applications adopt this strategy, although differing in finer design objectives. As an alternative to dedicating constrained screen space to tickering displays and other notification tools, Harrison et al. [1995] argue that transparent user interface elements, as a layered, space multiplexing technique, can provide awareness of other information and enhanced context while minimally disrupting focused attention on standard interface objects. Other desktop applications that are intended to be used with other tasks do not seem to be concerned with preventing distraction, since they proactively provide prompts that are intended to guide or enhance activities. Certainly, Microsoft's Office Assistant (Clippit) and Rhodes and Maes' Remembrance Agent [Rhodes and Maes 2000] are examples of these types of applications.

Other innovative work has demonstrated the feasibility and utility of presenting information within a user's environment, although there are many different approaches here as well. Large screen displays are used in both MacIntyre's Kimera augmented office environment [MacIntyre et al. 2001] and efforts like Informative Art [Redström et al. 2000], but there are fundamental differences in the objective amount of user attention necessary to extract information and gain meaning. Kimera's wall displays seek to provide quickly understood background awareness cues that complement the flow and context of work, while Informative Art provides a hidden representation data that is enjoyed during moments of deeper reflection. Techniques for subtly altering elements of the user's environment to convey information for background processing were demonstrated in the ambientROOM and elsewhere with projections of water ripples, natural soundscapes, spinning pinwheels, patterns of light patches, and the Information Percolator's air bubbles [Ishii et al. 1998; Dahley et al. 1998; Heiner et al. 1999]. Other work has described how physical widgets (called phidgets) were produced to display information states with curious, physical objects, such as an artificial flower arrangement or Phidget eyes [Greenberg and Fitchett 2001].

Although many of these examples are designed to enhance user efforts on desktop platforms and in office environments, similar research interest (and HCI expertise) often extends to cover more ubiquitous displays, such as vehicle and wearable navigation/information systems, heads-up displays (HUDs), and augmented reality applications. Collaboration tracking and groupware systems also tend to have multitasking design components, where information of interest is presented in a divided-attention situation. While most of the systems mentioned so far are not described by their contributors as notification systems,

they all share a few general goals, allowing a more cohesive view. As we seek to understand how to better model multitasking situations for usability studies, address known usability problems, and adopt research approaches that promote knowledge application and extension, we find it essential to synthesize as much previous work as possible.

2.1 A Unifying Theme

Since these example systems share several usage goals they can be more broadly classified as *notification systems* [McCrickard et al. 2003]. Notification systems are defined as interfaces that are typically used in a divided-attention, multi-tasking situation, attempting to deliver current, valued information through a variety of platforms and modes in an efficient and effective manner [McCrickard and Chewar 2003]. The benefits of notification systems can be numerous, including rapid availability of important information, access to nearly instantaneous communication, and heightened awareness of the availability of personal contacts. Apparent usage goals (fully detailed in McCrickard and Chewar [2003]) present an important distinction between notification systems and traditional HCI research:

The success of a notification system hinges on accurately supporting attention allocation between tasks, while simultaneously enabling utility through access to additional information.

This design paradigm provides a unifying theme for notification systems research that is quite different and more specific than typical interface study.

Computer users have long used notification systems like clocks, email alert tools, and system load monitors, suggesting that people may be willing to tolerate or even welcome an interruption if the information presented proves to add utility—often providing a competitive advantage, enhanced knowledge, better communication, or increased happiness—through appropriate, timely reaction, long-term comprehension, or possibly by simply facilitating information access. While demand for these types of displays appears to be increasing, questions remain regarding the effects of notifications on ongoing tasks. They are often perceived as distracting, but the degree to which they distract a user is not well understood. On the other hand, a compelling recent work showed cases of intrinsic utility in interruptions for managers [Hudson et al. 2002]. If trade-offs can be determined for information design options across platforms and information types, then various usage scenarios can be reliably supported with optimal presentation features. However, before any enduring progress can be made toward this end, we must be able to recognize and gauge deficiencies and successes in notification systems interface designs.

2.2 Evaluation Challenges

As one of the two important research challenges asserted by Abowd and Mynatt [2000] for the ubiquitous computing field, they motivate the imperative for

assessing progress toward real human needs with quantitative and qualitative evaluation methods that capture authentic context of system use. While some early studies of notification systems have captured some guidelines and design tradeoffs and serve as initial models [Cutrell et al. 2001; Mamykina et al. 2001; McCrickard et al. 2001], few efforts have been conducted and reported to explicitly afford knowledge application and reuse, or even facilitate study replication and extendibility—clearly objectives of empirical and analytical evaluation. While much of the dual-task experimentation (especially cockpit design) done within the human factors and engineering psychology fields seems highly relevant to this area of research, for example the numerous references provided. Wickens and Hollands [2000], it does not seem to be readily applied to notification systems or many other HCI multitasking requirements.

The lack of a unified perception of the notification systems field has resulted in numerous fragmented efforts addressing similar problems, resulting in few general guidelines for evaluating the effectiveness of systems and enabling little reuse of empirical conclusions that do emerge. As is common with emerging fields, many researchers seem to feel that summative evaluations are too challenging, and instead tend to demonstrate intrinsic value of design paradigms only through the generation of unique implementations, with usability claims supported by perhaps a few user comments or an isolated user study.

A primary outcome of research should be the incremental advancement of understanding how to support interactions that computer users desire. While this requires a strong theoretical base, recognizing successful models and implementations consistently comes from the ability to apply models to measure and compare analytical and empirical evidence collected over time. While articulating a theme for notification systems provides a common way of considering the disparate systems we reviewed earlier, we need to be able to model systems in a manner that allows comparison. If practitioners are to use and value our research, we must find methods that increase cohesion, extendibility, and replicability of individual results within the frame of a larger model [Gray and Saltzman 1998]. Part of the challenge in developing this tool is capturing accurate design model descriptions that represent user goals, as well as the resulting cognitive complexity from multiple system interactions within these divided-attention situations.

In the next section, we develop an argument that three critical parameters—*interruption*, *reaction*, and *comprehension*—can describe user notification goals. This argument forms the basis of our modeling technique. We describe how we considered many parameters of the usage experience, reducing a much broader set to this useful abstraction. This leads to the presentation of an initial framework for classifying notification systems, a cognitive process model that suggests evaluation and redesign imperatives, and a claim-centered mechanism for conducting reusable, comparable usability studies. These contributions provide a firm base, allowing incremental, useful advancements in the research field of notification systems with progress guided and measurable by well-defined critical parameters.

3. MODELING USER NOTIFICATION GOALS

Thus far, we have articulated a theme that expresses general goals and characteristics for this group of multitasking systems. We have provided some insight into the challenge and need for better evaluations of notification systems, demonstrated by the slow convergence of usable and extendible studies. This section lays the groundwork for our approach to discerning usability of these systems.

First, we look at a method of simultaneously describing a design model with the critical parameters. This allows us to consider and label general combinations, forming a descriptive and prescriptive design space. Using a simple model of human information processing allows deeper understanding of the regions within the design space through identification of action models. We demonstrate the utility of this novel approach for notification systems classification at the end of this section by integrating several examples of existing applications within the framework and illustrating how reusable design guidelines are then possible through a claims-centered approach to usability evaluations.

3.1 Critical Model Parameters

In order to conduct meaningful usability evaluations that will allow systems to become progressively better, Newman [1997] argues that we first must define or adopt *critical parameters*, or figures of merit that transcend specific applications and focus on the broader purpose of the technology. He implies that well selected critical parameters can function as benchmarks—“providing a direct and manageable measure of the design’s ability to serve its purpose”—and indicate the units of measure for analytic methods that predict the success of an early design. Newman provides examples and makes several recommendations for identifying critical parameters that support core user tasks and goals.

3.1.1 *Evaluating and Selecting Options.* Our first step in selecting critical parameters for a model of notification systems was to identify key user tasks and usage constraints. We developed a long list for both. Users notification goals include typical tasks such as receiving information that is more important than current activities (perhaps prompting task transition), regularly monitoring a secondary information source over an extended period of time, becoming informed about timely instructions or information states to advise critical primary task actions. Constraints to notification system use include information complexity and granularity, situational context, available cognitive resources, associated familiarity and enjoyment, and delivery mode and method (continuity and encoding). To reduce the complete collection of tasks and constraints to a manageable set, we employed two processes: 1) separating design model and user model attributes, and 2) identifying dependencies in order to focus on root causes. Each process is described in turn.

First, we considered the distinction between two types of information about users that designers should have available. Following Norman’s [1986] terminology [1986], the *design model* describes the designer’s conceptual model of the user’s background, goals and tasks, and processing limitations. Likewise, the

user model refers to the conceptual model that the user forms according to their expectation and experience of the actual system. Our thought is that modeling each according to similar criteria would be ideal (allowing easier comparison), forcing consideration to be on *anticipated and actual effects* of an interface artifact on a user—which ought to correspond to user goals. This implies that implementation details (e.g. information or notification delivery characteristics that may impact sense of privacy, aesthetics, and subjective satisfaction) should not be a first-order variable within the model, but should be thought about as a system characteristic, modifiable at some level to accommodate less flexible design requirements.

Second, we looked at the dependencies in our list of key user tasks to determine primary factors and generalize the tasks as much as possible. The biggest challenge in doing this is identifying critical parameters that are “measurable” and “manageable,” yet ensure that those parameters characterize essential facets of user interaction. Much guidance comes from Whittaker, Terveen and Nardi’s argument for reference tasks [Whittaker et al. 2000]. Since the purpose of our model is to aid comparison of designs that are created to support similar user goals and facilitate recognition of design progress, we do not want to select critical parameters that cannot be modified by an interface design. While situational context is certainly an important facet in the success of a notification goal, and it is tempting to include it as a critical parameter, designers are often unable to anticipate or address context variables. Therefore, we reserve aspects of context as an essential element of artifact descriptions and claims, but do not include it as a primary critical parameter for our model. Likewise, while user satisfaction and enjoyment with a notification system may be an independent goal, we believe that satisfaction is typically derived from efficient and effective delivery of the notification according to a positive balance of the attention-utility theme (as described earlier) [McCrickard and Chewar 2003]. Our current determination not to include satisfaction as a critical parameter may be reassessed with further research.

However, as we inspected the general tasks that contribute to notification utility through “access to additional information,” we recognize that user interruption, near-term reaction, and long-term comprehension are the immediate results of such access. More importantly, these three parameters are manageable through design choices, measurable in empirical user testing, and capable of being modeled in terms of cognitive processes. Certainly, each has received much attention in the multitasking and notification research communities (as we proceed to describe). Each of the three can also be thought about as a guiding force of a design model and the desired or undesired consequence of the information presentation of the user model. Therefore, they are the root causes of a design’s success—the main factors that ultimately cause a shift in the balance of the attention-utility theme. Based on this argument, we recognize three critical parameters for modeling of notification system user goals and system designs: user interruption, reaction, and comprehension.

3.1.2 Interruption. User goals and usage scenarios for notification systems often have some requirement regarding the interruption of primary tasks. In

the context of notification systems study, we define *interruption* as an event prompting transition and reallocation of attention focus from a task to the notification. Some situations, such as driving a car equipped with an in-vehicle information system (IVIS), require that a notification system not intrusively disrupt user attention devoted to a main task. Guidelines established in the area of IVISs suggest defining limited numbers and types of interactions with the displays, restricting the amount that displays change, and limiting the time that a display is present [Ballas et al. 1992; Green 1999; Tufano et al. 1996; Sheridan 1991]. However, other situations, such as monitoring a nuclear reactor, explicitly call for notification-prompted task-switching. Horvitz's models and inference procedures present some hope for this design objective, an imperative driven by his belief that human attention is the most valuable commodity in HCI [Horvitz et al. 1999; Horvitz 1999]. These models are designed to improve notification utility by considering cost of user interruption and introducing notification presentation appropriately. McFarlane describes a taxonomy and empirical study describing the major dimensions and design tradeoffs related to interruption [McFarlane 1998, 2002]. The tentative guidelines he established exhibit design goal tradeoffs among the coordination methods, although negotiation-based interruption coordination appears to be best for many cases. Selection of information design for a notification system that is driven by inferred suitability of interruption will likely have impacts on the two other design objectives (reaction and comprehension) and affect overall system utility.

3.1.3 Reaction. The second critical parameter we propose is the rapid and accurate response to the stimuli provided by notification systems, an effect which we refer to as *reaction*. Often, notification systems present cues intended to inform the user of information of interest, often requiring them to differentiate between values. As such, several studies have investigated how to improve reaction to notifications using preattentive processing, which considers how information can be assimilated and understood rapidly by using colors, shapes, and motion [Enns and Rensink 1991; Healey et al. 1996; Healey and Enns 1999; Bartram 1998; Bartram et al. 2001; Bartram 2001]. Other work has examined moving and changing text as a method for presenting information in hands-off displays, observing the perceptibility and readability of rapid serial visual presentations (RSVPs) of letters, strings, and words [Foster 1970; Duchnicky and Kolers 1983]. These types of studies investigated rapid reaction to information, yet they did not consider more in-depth and memorable understanding of it, our third measure of notification systems.

3.1.4 Comprehension. While rapid and accurate reaction to an informational cue is important in many situations, often it is also (or only) vital to use notification systems with the goal of remembering and making sense of the information they convey at a later time. We refer to this as *comprehension*. Again, we consider research relating to textual motion as an initial example for studying relative comprehension of secondary display information. Juola found that comprehension of information was comparable when presented as RSVPs and

in multi-line paragraph format [Juola et al. 1982]. A study led by Granaas found that in scrolled displays, larger jumps (four to ten characters) led to better comprehension than smaller jumps (one to two characters) [Granaas et al. 1984]. Kang and Muter [1989], in comparing a tickering effect to a non-animated RSVP effect, found no difference in comprehension for a reading task. Other efforts have focused on evaluation of various attributes (position, area, and color) in secondary displays for supporting information extraction and comprehension as part of tasks requiring detection, estimation-ratioing or estimation-compare [Chewar et al. 2002]. We found that the three attributes are significantly different in enabling comprehension at various levels of primary task degradation.

Notification systems research should focus on exploring balances between the interruption, reaction, and comprehension design objectives. However, most of these studies seem to focus on one, or perhaps two, of these critical parameters, seeking to identify forms of information representation that provide the best support for accepted design tradeoffs. In order for critical parameters to add value to research, all three should be acknowledged in an evaluation process and have standard representational methods. We go on to propose such a model.

3.2 The IRC Characterization Framework

As we conveyed in the discussion of challenges to evaluation of multitasking systems, one of the most difficult and important aspects is to adequately consider multiple critical parameters that gauge different outcomes of a single resource. In the case of notification systems, various levels of interruption, reaction, and comprehension result from and cause changes in attention allocation. Since notification systems are typically used in a divided-attention situation where they are not the main focus of attention, assessing these critical parameters often requires consideration of both a primary task and the notification task. As if conceptualizing concurrent and perhaps conflicting design model objectives and modeling them as a user study is not difficult enough, understanding what the evaluation results indicate about the user model and using this insight to guide iterative prototype refinement can be quite complicated. The various approaches to these problems taken by different design teams make extending knowledge to new applications difficult as well.

To improve this impasse, we propose characterizing all notification systems according to their blend of the three critical parameters. In doing this, we strive for a mechanism that captures the design model—the objective system based on anticipated user goals. Keeping this as simple as possible, we are initially only considering combinations of *high* (1) or *low* (0) levels of each parameter. For example, a user goal can require a notification system that provides immediate reaction to new information without introducing interruption to a primary task or gaining a deep understanding of information over time. This design model can be described as low interruption, high reaction, and low comprehension, or IRC 010. When we consider this specific parameter combination, it seems to describe an *indicator*—a passive device used for conveying information status and allowing quick recognition of, and reaction to meaningful data.

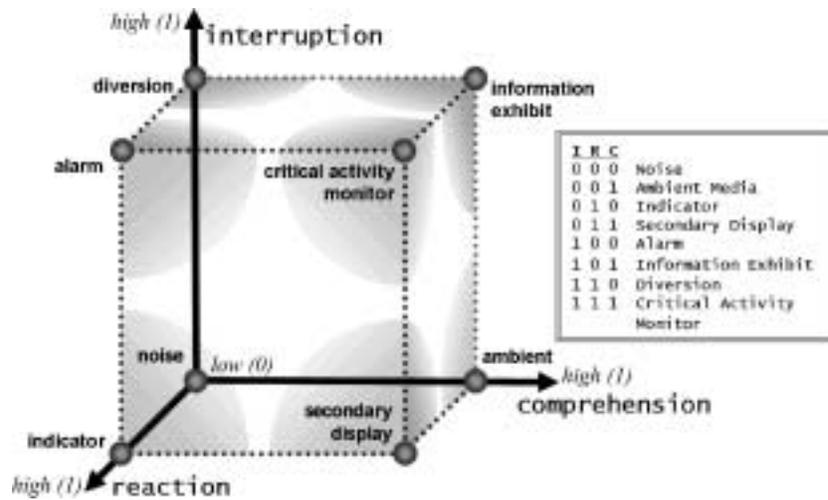


Fig. 1. Notification systems categorizations according to blend of design model objectives (representing user goals) of interruption (I), reaction (R), and comprehension (C), simplified as low (0) or high (1).

Extending the same approach to the other seven combinations of parameter levels, we are able to conceptualize user need scenarios and identify a descriptive name for each design goal. Figure 1 provides a list of all eight combinations with names, as well as a useful visualization of these regional relationships. We represent each critical parameter as an independent, orthogonal axis, increasing from a low to high objective level. While the IRC categorization only precisely describes the corners of our notional cube, we believe it is useful to initially consider regions as extending from objective to near-mid range levels. We fully expect that as this framework is tested and used to describe other user need scenarios, additional logical regions and associated IRC levels will be identified, serving as refined categories of notification systems.

Several ideas may be initially non intuitive as these design model blends are considered. First, high interruption may appear to be an unlikely user goal for a notification system. However, users often multitask in anticipation or vigilance of the introduction of a certain information state or receipt of a message. For example, in a collaborative document writing activity, a user may be editing a section of the document while waiting for certain actions to be completed by colleagues, maintaining awareness of collective progress with a notification system. When various states of progress are achieved, the user may desire an interruption from the current task that prompts transition to a more important task. Likewise, stock brokers or other decision makers may perform less important activities while they monitor news and stock prices, needing and valuing interruption when important information states are presented—not only to prompt task transitions for immediate reaction but also to enable deeper, immediate inspection of the information. In other cases where interruptions could be valued, users may rely on notification systems to provide advice or guidance

for primary task execution—software agents and surgery support systems are compelling examples.

A second potentially perplexing notion implied by this framework is that reaction or comprehension (or both) can occur without interruption to other tasks. At this point, it is important to recall that the IRC levels represent our understanding of *user goals*—the design model—or, objective performance to be facilitated by the system. If designers can leverage skilled memory, task automaticity, and preattentive processing capabilities of users, possibly through use of efficient encoding, rich affordances and metaphors, and cross-modal information conveyance, such design models may be realized. These concepts are discussed in greater detail in the next section.

An important consideration for this conceptualization is the validity in our assumption that these three critical parameters can be considered as orthogonal. Since each IRC blend seems to correspond to potentially realistic usage scenarios and system classifications, this seems like a plausible initial framework. The action models presented later in this section also reinforce orthogonal representation of these critical parameters. However, to convince the skeptical reader and fully clarify ideas encapsulated by the region labels, we present a brief description of a likely usage scenario, motivating and articulating each corresponding design model and IRC blend.

- *Ambient Media*(001)—an office worker without a window effortlessly maintains awareness of the weather throughout the day with dynamically changing desktop wallpaper. Although knowledge about the weather may be applied in a later conversation or decision, reacting to sudden changes or specific instances is not important.
- *Indicator*(010)—a traveler in an unfamiliar city uses a vehicle navigation system to prompt required turns along the route. He has no interest in learning his way around the city, and is only concerned with negotiating traffic and arriving at the destination quickly and safely.
- *Secondary Display*(011)—while an editor works on part of a document that is distributed among co-workers, she monitors a groupware tool on the office's large screen display that shows various progress meters for the different parts. Information presented is important for pacing or technique adjustment, as well as an overall understanding of team contributions.
- *Noise*(000)—a student working on a slide presentation may not need network access, but perceiving a functional information channel (perhaps providing Internet radio) may be reassuring.
- *Diversion*(100)—a home computer user enjoys using his computer more with lower stress if a friendly agent occasionally pops-up with a joke.
- *Alarm*(110)—as a businessman attends to various tasks throughout the day, he relies on calendar and email alerts to keep appointments and quickly view important emails. Redirecting activity to the right place, at the right time is the only important consideration.
- *Information Exhibit*(101)—a factory supervisor performs routine administrative tasks while maintaining awareness of overall operations. While she

expects operational details to be handled by lower level managers, frequent updates are critical for seeing how statuses change over time, allowing assessment of long-term strategy, subordinate decision-making, and operational trends. Understanding this important information often requires close examination due to complexity.

— *Critical Activity Monitor*(111)—while performing many routine activities, a system administrator uses a network monitor on a small portion of the desktop. Many users critically depend on his quick and insightful response to network problems, but he is even more valued for understanding specifics or patterns relating to problem prediction and enabling fault-free preventative network maintenance.

Having illustrated a possible usage scenario for all eight blends within the IRC characterization framework, several differences are readily apparent in design model information interaction approaches. For instance, perception of information changes can be expected to be performed with quick but frequent, non-interruptive glances, careful study during self-defined task breaks, or through peripheral or background perception. Some scenarios called for information presentation that could be fully interpreted and acted on without other information, while others suggest that new information would only be meaningful when associated with previous knowledge or if additional details were accessed. This range of expected interaction approaches implies that different usage situations can be modeled in different ways. The next section presents one possible modeling approach.

3.3 Notification Action Models

Norman's [1986] theory of action provides the HCI community with a common representation of activity stages required to complete a task. Having this theoretical tool aids the task analysis process, since inspection of interface performance (information or interaction design evaluation) can focus on specific stages or transitions, particularly during a scenario-based design approach [Rosson and Carroll 2002]. However, when considering a multitasking situation typical of notification systems use with critical parameters like interruption, reaction, and comprehension, the tool remains an important influence but seems overly abstract in its ties to cognitive processes. For a theoretical model to be useful for understanding notification systems, it needs to demonstrate parallel processing limitations within and between activity stages, allowing designers to discern conflicts between primary and secondary activities.

Better representations of task flow should be more closely tied to cognitive architectures, providing both the stage-based focus of a theory of action and the rich link to cognitive science research. Evaluators can target specific areas of the action model for empirical investigation, and seek problem explanations and associated iterative redesign strategies (which Barnard and May [2000] refer to as *microtheories*) from a well established field. Computational cognitive modeling, as demonstrated by SOAR [Newell 1990] and the more recent EPIC and ACT-R/PM models [Kieras and Meyer 1997; Anderson and Liebiere 1998], simulate and predict user performance with interfaces, and may be a long term

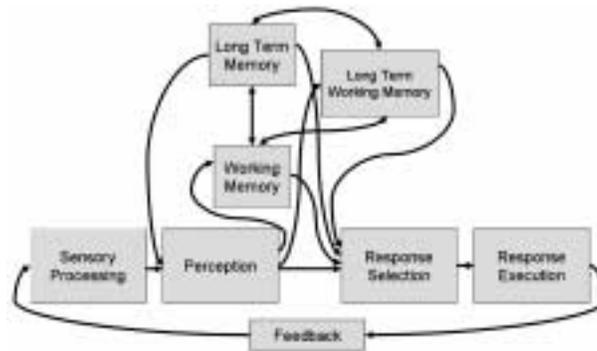


Fig. 2. A human information processing stage model, from Wickens and Hollands 2000, pp. 11 and 295.

solution. However, if research, evaluation, and interface design approaches are incompatible with modeling methods, dividends will be slow coming regardless of the model robustness.

Barnard and May [2000] argue that we should consider a system's behavior as "a trajectory governed by systematically structured sets of constraints." When several systems simultaneously support user goals and resulting interaction, the larger system should be modeled according to a *macrotheory*, with the interaction trajectory providing the center of interest. To form the psychological component of a macrotheory, the authors present a cognitive architecture (Interacting Cognitive Subsystems, or ICS) describing interactors and organization between subsystems that handle sensory input, action coordination, and high-order abstraction of information. This model is quite useful for realizing the processing stages required for and potentially constraining task performance.

With similar motivation to understand possible interaction trajectories characteristic of notification system design models, we surveyed theories of human information processing stages and found models presented in Wickens and Hollands [2000] to be most useful for our purposes (see Figure 2). This representation and the related material provided in this reference is particularly handy, since it allows mapping of various trajectories, provides tight integration with our critical parameters, and aids understanding of parallel processing opportunities and bottlenecks.

Using this abstracted model of human information processing, we mapped notification task trajectories for each of the eight broad scenarios for a user's receipt and processing of a notification (discussed earlier). Since arrows depict the possibilities for attention flow, we considered the available flows from each cognitive process that could be used for attention allocation to the notification. As we reasoned about the likely information processing paths for each scenario, we used the associated IRC classification to recall the generalization of user goals. For instance, we can think about interruption as the disruption and resetting of working memory—an inevitable effect of context switch and attending to unfamiliar or complex information for anything but a few seconds. Comprehension requires flow of attention to the long term memory in order

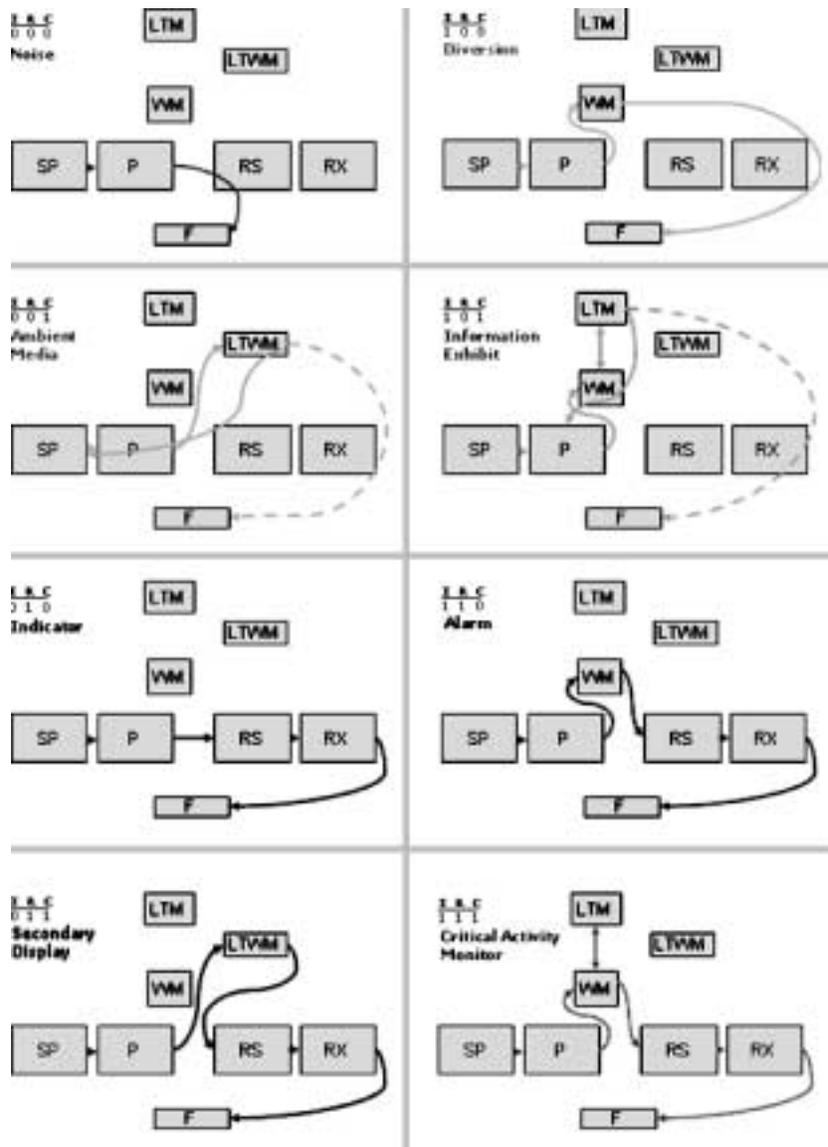


Fig. 3. Design model flows through the human information processing stage model (see Figure 2) for each of the eight main notification systems' IRC categorizations (see Figure 1). Note unique design path trajectories for each categorization.

to link new information to existing knowledge. Reaction is the observable outcome of the response selection and execution stages. Realizing the presence of each goal (as well as the approximate order in which each goal would be fulfilled) ensured inclusion of attention flow to appropriate cognitive processes. The notification task trajectories for each of the eight general scenarios (single notification assumed) are depicted in Figure 3.

This provides expected action models for each of the eight design model characterizations. Each of the eight trajectories would need more extensive reasoning before they could be thought of as more than “useful approximations,” however there are still several points of interest in this result. First, each path is unique, further supporting our assumption of orthogonal critical parameters. Long term working memory theory plays an important role in our trajectories, and although many ideas are currently debated in psychology channels, the most compelling evidence for this more efficient and less volatile skilled memory comes from dual-task and task switching experimentation [Ericsson and Kintsch 1995]. Trajectories for the ambient media and information exhibit categorizations contain a top-down processing element, in which an interface is searched for specific information rather than simply reacting to presentation of stimuli.

If designers are able to gauge user expectations for notification interruption, reaction, and comprehension (forming a design model IRC), they can design the information and interaction display in a manner that promotes ideal flow of attention between cognitive processes, as depicted in the appropriate section of Figure 2. For example, it may be argued that in the case of an alarm a user would access long term memory to recall the steps for reaction. However, as depicted with the alarm trajectory, an *ideal* alarm design would attempt to avoid accessing long term memory, perhaps conveying all necessary information in a highly compact manner. As the interface only supports the notification task (not the tasks that would result from an attention transition, which would be performed as primary tasks), this is a realistic design goal. Thinking about the human information processing model in these terms clarifies its usefulness in a design process.

Similarly, when testing a particular design model claim, notification systems evaluators can use these action model trajectories to refer to studies within the cognitive psychology field. Generally accepted testing and reporting methods can be leveraged to capture more precise measures of interruption, reaction, and comprehension. For instance, Rogers and Monsell [1995] studied the cost of task switching. Not only do they provide an excellent review of related work, but they introduce a method of employing alternating task switch and non-switch trials, effectively arguing that the task switching costs captured describe the need to switch tasks better—essential for understanding usability of systems that provide information guiding primary task performance. Similarly, experiments conducted by Baddeley [1996] to validate the conception of a homunculus as a model of working memory provide guidelines for dual-task experimentation isolating working memory performance. See et al. [1995] provided a review of sensitivity decrement studies for vigilance tasks (particularly useful for evaluating critical activity monitors, secondary displays, and indicators) which not only summarizes important design considerations, but provides a meta-analysis and common view of conclusions from 42 similar studies throughout the literature—a feat that seems quite intractable within our field.

Not only can we use these action models to guide our evaluation processes (such as conducting a cognitive walkthrough), but understanding concepts such as bottleneck theory (expressed in single-channel theory of the psychological

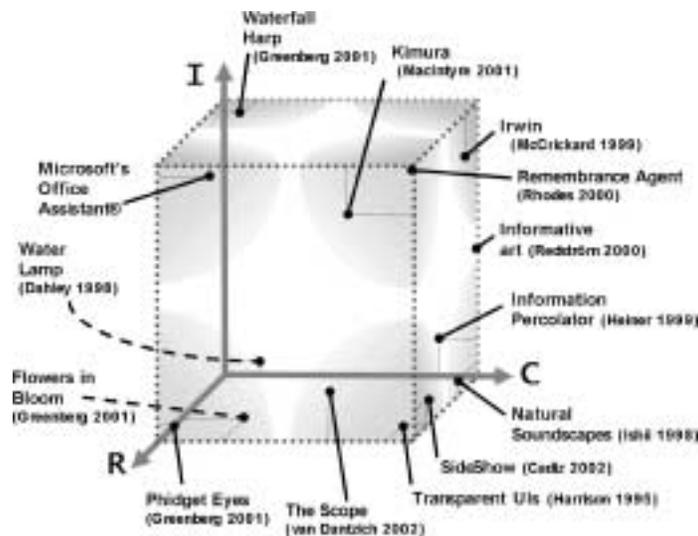


Fig. 4. Inferred IRC categorizations (design model interruption, reaction, and comprehension objectives) of several notification systems.

refractory period), cross-modal sensory perception, automaticity, and preattentive processing provide valuable insight for addressing identified user problems. Further discussion of these topics is beyond the scope of this paper, but an excellent review and additional references is available in Wickens and Hollands [2000].

Having discussed the initial foundations of a classification and modeling system for notification systems, we turn our focus to applying the IRC categorization framework and notification action models.

3.4 Using the Notification Systems Design Space

Since we have set forth a framework for classifying design models of notification systems according to IRC categorizations, we can revisit some of the existing systems discussed early in Section 2. Although it may appear as though most of these systems had very little in common with each other, we identified an attention-utility theme that expressed goals common to all of these systems. Each takes different implementation approaches, but they all seek to provide some utility by presenting additional information while appropriately preserving desired attention distribution. Implementation differences are motivated by the designer's expectation of the differing interruption, reaction, and comprehension levels desired by a user during their interaction with the notification information.

From the claims made by the authors in describing these systems, we can only infer design model details regarding the critical parameters. We use these inferences to provide an initial IRC classification of each system (see Figure 4), but we hope that additional, collaborative analysis with designers of these and other systems will refine the classification and overall understanding of the

framework. To clarify our method of assigning an IRC classification, we describe the process for four systems well dispersed throughout the design space: Informative art [Redström et al. 2000], Water Lamp [Dahley et al. 1998], Remembrance Agent [Rhodes and Maes 2000], and Flowers in Bloom [Greenberg and Fitchett 2001].

- Informative art—In Redström et al.’s description of these computer amplified, dynamic works of art, they present this class of displays as distinctly different from ambient media or information visualizations, specifically mentioning that these are not intended to reduce information overload by enabling peripheral perception of information (*not low* interruption). Instead, the period of time required to view and decipher deep meaning (*high* comprehension) provides a valued moment of rest and reflection for users (*some* interruption), although the displays are intended to be non-obtrusive, aesthetically pleasing objects during times of non-use (*not high* interruption). Furthermore, no user utility gain is anticipated by prompting responses like spontaneous informal communication (*low* reaction). *IRC characteristic: (.5/0/1)*
- Water Lamp—Dahley et al. provide an example usage scenario for their ambient projection of light through water ripples created by computer-controlled solenoids: enabling a sense of connection to a loved one by displaying their actual heart beat. The projected ripples are intended to be casually perceived and processed at a user’s “periphery of attention” (*low* interruption), without invoking moment-to-moment responses (*low* reaction), but providing some awareness of the loved one’s activity levels (*slight* comprehension). The true utility gained by a user of this system is anticipated to be an added feeling of closeness. *IRC characteristic: (.1/0/.25)*
- Remembrance Agent—Rhodes and Maes discuss the goals a user would fulfill with their just-in-time information retrieval agent: as a user types a document he receives an alert (*some* interruption) about related documents with one-line summaries provided at the bottom of the text window. Suggested documents can be old emails, notes, webpages, and so on, leveraging and linking existing knowledge (*high* comprehension) or inspiring new ideas for the editing task (*high* reaction). Clicking on the summaries (*high* reaction) allows an easy and desired task transition—access to the full text of the suggested documents (*high* interruption). *IRC characteristic: (1/1/1)*
- Flowers in Bloom—Representing information in a continuum of states according to the bloom-level of an artificial flower arrangement, this device is intended to be non-intrusive within an environment (*low* interruption), providing a single value in each glance (*slight* comprehension) that would facilitate appropriate action (*some* reaction). *IRC characteristic: (0/.75/.25)*

To better understand our other characterizations in Figure 4, interested readers should look at the cited papers describing the applications. Although we focused this analysis on assigning an overall IRC characterization for any combination of design goals, a per-task method of assessing IRC levels is thought to be more descriptive. Tasks that are accomplished simultaneously can be

thought of as supertasks. The Scope application [van Dantzich et al. 2002], described in the Section 4 case study, provides an example of this analysis.

This application of the IRC framework readily illustrates the expected source of utility, in terms of critical parameters, provided by the notification system. Referencing the applicable notification action model for the design space region's IRC blend also allows a basic understanding of the anticipated cognitive trajectory, which can be overlaid on the primary task action stages for better dual-task usability engineering. Furthermore, while we have demonstrated the plotting of *design model* IRC characterizations, *user model* or actual characterizations (evident from evaluation results) can also be plotted, indicating a design disparity vector that should be closed through iterative reengineering. While this may allow a single application to become progressively better, we are more interested in facilitating contributions that enhance the collective notification systems research effort.

In order to accomplish this, we must understand how to compare systems in a formative or summative evaluation, generalize design guidelines for future applications, and gauge overall process against benchmark critical parameters. Success in these endeavors can only be proven through prolonged, popular use of the models presented here, but we can make several recommendations for continuing progress.

3.5 Adapting Traditional UEMs

Sutcliffe [2000] argues that HCI research should focus on producing designer digestible packets of HCI knowledge in the form of claims, grounded on good theory and allowing general reuse. He defines *claims* as situated advice about design rationale that expresses the upsides and downsides of the usability of an artifact. Claims analysis is accomplished by evaluating artifacts, and claims can be written generically, classified and organized in a catalog that allows association of artifacts with established design tradeoffs. He cites two potential problems to this approach: "creating a generic version of the claims and artifacts and then matching appropriate claims with a new application context."

We believe that the notification systems design space, as described by the IRC characterization framework, is concise enough to facilitate the creation of generic claims, resulting naturally from generic usability evaluation method implementations. That is, each region of the IRC framework can have corresponding method implementations that can be used in any evaluation. For analytic methods, this could mean using associated action models to guide a walkthrough process, or using heuristics that are specifically designed to capture targeted levels of interruption, reaction, and comprehension. Regions can also prescribe experimental metrics and procedures, as well as methods for field studies and items for questionnaires that can be used to capture comparable data. We tested this notion in the case study that follows, by creating a claim-specific list of questionnaire items that could be used to evaluate the same claims of other artifacts.

This implies a solution strategy for convincingly conducting summative evaluations, as well as matching established claims to new applications. Summative

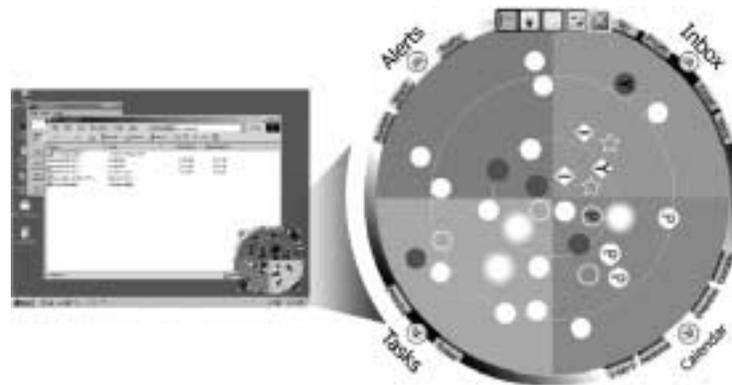


Fig. 5. The original prototype design of Scope, the application used as the focus of our case study. Scope sits in a corner of the desktop, presenting notification items as symbols within categorical quadrants. Urgency ratings correspond to centrality within radar metaphor. Scope is fully described in van Dantzich et al. [2002].

evaluations for systems within a common categorization region become simple with generic UEM implementations. Benchmark levels for critical parameters within each region can be determined in due process, and could be quite useful for judging design potential of new artifacts in early development stages. New design model concepts can be matched with claims that are correspondingly cataloged within common IRC characterizations, allowing reuse and enhancing opportunities for incremental progress within the field. Assessing the potential and procedures for intra-regional comparisons and claim applications will be more difficult, but will also add immense value to our understanding of notification systems usability.

4. CASE STUDY

To test the utility of the IRC characterization framework, the corresponding action model, and our notion of generic IRC-based UEMs, we conducted a case study. Our case study compared two formative usability evaluations where questionnaires were used as the primary evaluation method. The original was conducted by researchers at Microsoft as part of the iterative design process for their notification system, the Scope van Dantzich et al. [2002]. The second was conducted in our lab with a similar study developed using the IRC framework and a simulated version of the prototype. Guidelines derived from the two evaluations were compared to determine which evaluation was more effective.

The interface under consideration in both evaluations is the Scope, a notification system developed by Microsoft researchers to help users stay aware of information using a radar-like circular display with higher urgency items located closer to the center of the Scope (see Figure 5). The application constantly resides in a corner of the desktop, providing information on and an access point to notifications. The initial prototypes of the Scope divided the space into four categories: the email inbox, a calendar, a task list, and general alerts. The

| Questionnaire Item: | Avg. Rating (1=Disagree, 7=Agree) |
|---|---|
| 1. The ability to change priorities for an item by drag and drop was easy. | 6.33 |
| 2. It was easy to move between low and high levels of detail. | 6.167 |
| 3. The use of pulsing to show "new" items allowed good detectability. | 6 |
| 4. The ability to expand a wedge was useful. | 6 |
| 5. I was able to see an overview of my high priority items from the Scope at a glance. | 5.83 |
| 6. The different shapes of the items were easy to learn and useful. | 4.83 |
| 7. The Scope Alerts prototype was easy to use. | 4.5 |
| 8. The different textures/motions of items (pulsing, fuzzy or inverted) were easy to learn and useful. | 4.33 |
| 9. The ability to delete items from the Scope view was easy. | 4.33 |
| 10. It was clear what region of the Scope meant "high priority" and what region meant "lower priority". | 4.167 |
| Overall Average Subjective Rating: | 5.3 |

Fig. 6. Questionnaire and ratings used by the Scope design team, reported in van Dantzich et al. [2002].

appearance of items in the Scope reflects information such as recipient lists for emails and expired deadlines for calendars and task lists.

We selected this interface for our case study since van Dantzich et al. are exceptionally thorough in reporting their design objectives, justifications, usability study, and iterative refinement decisions (in van Dantzich et al. [2002]). Recognizing that such scholarship is vital for incremental advancement of ideas in any research field, we were particularly grateful for a well documented effort of this type. According to its designers, the Scope is intended to “direct a user’s attention to high urgency items” yet in general require “minimal attention to stay aware of incoming notifications” [van Dantzich et al. 2002]. According to our IRC model, this means that the Scope should act both like an alarm, supporting high interruption and reaction but low comprehension (IRC 110), and like an ambient display, supporting high comprehension but low reaction and interruption (IRC 001). That is, the Scope is intended to support the alarm-ambient supertask where it must simultaneously enable detection of urgent notifications while facilitating task transition decisions and provide awareness of all pending notifications without distracting other tasks. Scope’s IRC depiction in Figure 4 represents this supertask characterization.

4.1 Evaluations

After the initial design phase, the Scope developers conducted a pilot usability study intended to identify major usability problems to be addressed in the next design iteration. In the study, six participants performed a series of eleven tasks using the Scope in a standalone setting. Tasks included identifying high urgency items that met certain criteria, and interacting with the Scope at appropriate times in appropriate ways. For the tasks, completion times and verbal protocols were collected. After performing the tasks, participants completed a questionnaire consisting of ten questions that participants rated on a 7-point Likert-type scale (see Figure 6).

While the general style of the Scope study might be reasonable for traditional pilot studies, it failed to account for the unique interactions users have with

notification systems. In the Microsoft study, participants used the Scope just as they would a word processor, spreadsheet, or visualization tool, and many of the questions on the questionnaire probe standard interface issues despite the fact that the designers claim the Scope is intended to be used quite differently than a typical interface. This seemed to make it difficult for the designers of the study to use the results of the questionnaire in establishing future design iterations.

In our study, participants experienced a similar training base through task completion, but with the added benefits of a dual-task situation to provide a truer sense of the effectiveness of the interface. Rather than using the Scope by itself, our participants kept the notification system running in support of a secondary task, with the primary focus on a document editing task. Participants completed two five-minute rounds, with high-urgency items of interest specified before each round and general awareness questions asked after each round. After answering the questions, participants were informed of the correctness of their responses and reactions to provide them with a sense of their performance. In performing the tasks, participants were instructed that their primary goal should be to complete as much of the editing task as possible while still reacting to certain high-urgency items and staying aware of the general state of the information. We feel that a dual-task situation is necessary to encourage users to consider their behavior given two claim categories: alarm and ambient.

To further enhance the participants' alarm experience, in each round participants were asked to click on specific high-urgency items (such as a new email sent just to you) just as they would when using the Scope in a real setting. In terms of the notification action model we discussed earlier, this requires participants to experience stimulus perception, working memory dump, and response selection resulting in task transition. By completing several such alarm-style interactions, participants should be better prepared to judge the Scope's ability to support alarm interaction. To encourage the ambient experience, participants were informed that at the end of the round they would be asked questions about the information that appeared in the interface (such as the total number of items or the category in which the most new items appeared). In terms of the corresponding action model, this requires participants to experience stimulus perception, maintain their working memory, and yet expand their semantic memory with new information. By answering several such ambient-style questions over multiple rounds, participants should be better prepared to judge the Scope's ability to support ambient interaction.

Going into the questionnaire, our participants had experienced a more realistic usage environment and should be better prepared to assess the ability of the Scope to act as a notification system in the ways intended by the designers. Our questionnaire is divided into three parts: an alarm assessment category, an ambient assessment category, and an alarm-ambient supertask category (see Figure 7). We developed the questionnaire to be of comparable length to the questionnaire that the Scope evaluation team used. Question selection was based on our assessment of the Scope's design model (as discussed earlier in the case study description) and is intended to explore the tradeoffs between interruption, reaction, and comprehension experienced by the participants. For

| | Questionnaire Item: | Avg. Rating (1=Disagree, 7=Agree) |
|------------------------------------|---|---|
| Alarm claim | 1. The interface effectively supports rapid reaction to important information. | 3.17 |
| | 2. It is easy to distinguish important from non-important information. | 4.67 |
| | 3. Appropriate reactions to the interface are obvious and intuitive. | 3.00 |
| | 4. The interface instilled confidence that all important items were seen. | 2.33 |
| | 5. It is obvious when a reaction to information in the interface is more important than continuation of the current task. | 2.67 |
| | | 3.27 |
| Ambient claim | 6. The interface provides an overall sense of the information. | 3.83 |
| | 7. The interface provides an ability to select and understand clusters in the information. | 3.67 |
| | 8. The interface supports easy understanding of how information changes over time. | 4.33 |
| | 9. The interface supports easy understanding of links between different types of information. | 4.83 |
| | | 4.03 |
| Supertask claim | 10. The interface provides these overall understandings of information in a non-intrusive manner. | 4.0 |
| | 11. Important new information can be quickly recognized and accessed, and it is possible to stay aware of all information without using your place in the document. | 3.17 |
| Overall Average Subjective Rating: | | 3.81 |

Fig. 7. Questionnaire designed based on IRC claim categorization (alarm, ambient, or supertask), with ratings obtained in our user study. Apparent from the mean claim ratings (3.27, 4.03, and 3.17 respectively), the Scope facilitated ambient goals best and was most lacking in support for simultaneous (supertask) goals.

instance, designers of the Scope anticipate users will welcome brief interruption to properly react to sporadic, high urgency notifications. Support for this alarm goal is assessed with the alarm portion of the questionnaire. As we thought about key reaction questions, tenets of signal detection theory outcomes were influential. However, normal use of the Scope is expected to allow longer-term awareness of notification items with glances that do not interrupt the primary task or invoke immediate reaction. This ambient design model is tested with a different series of questions, which probe user satisfaction for support of typical and general ambient notification tasks. Rather than trying to speculate about the combined effect on users that results from simultaneous and disparate design models, we added a final question to test the supertask.

All questions were intentionally designed to be generic so that they could be readily applied to other interfaces supporting similar design models—thus enabling benchmarks and comparison. While continuing work focuses on validation and factor analysis of testing instruments that are adapted for IRC models, our intent with this case study was to demonstrate the performance of a testing tool that could be mapped back to the IRC model.

To judge the merits of our redesigned evaluation method, we compared the findings from both questionnaires with the actual redesign, which was based not only on the original questionnaire but also on user comments and expert reviews. One concern with the original evaluation was that many of the apparent findings from the questionnaire were not followed in the new design, suggesting that it did not probe the issues properly and it did not provide the participants with a realistic user experience. For example, the third question in the original questionnaire suggested that pulsing of new items for three seconds supports good detectability, which may be true when using the application in a standalone manner but which may not be adequate when simultaneously engaged in another task. In fact, many of the responses to the revised questionnaire suggest that the alarm functions are not adequate, a feeling clearly shared by

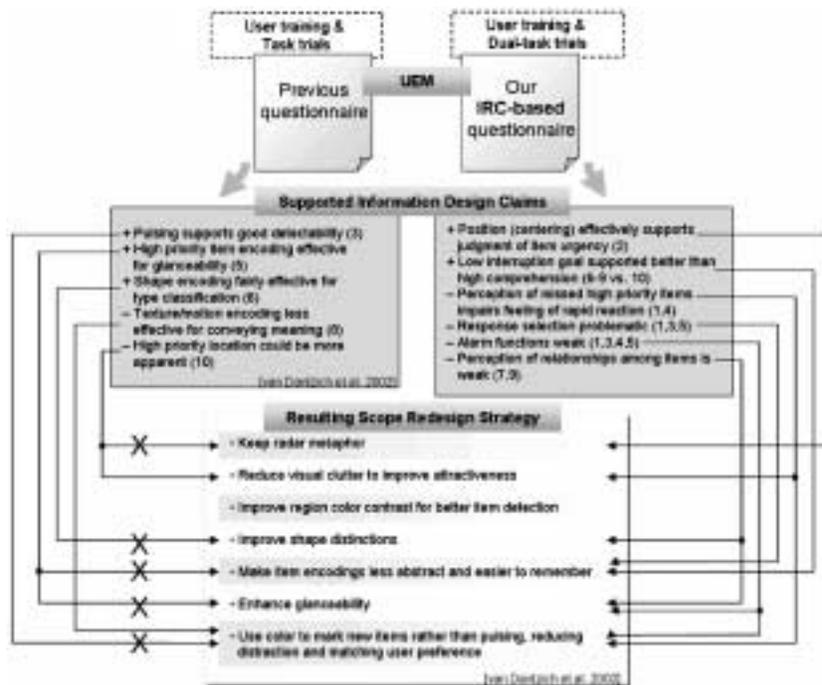


Fig. 8. Case study summary. From the previous questionnaire and our IRC-based version, we extracted information design claims and then mapped them (using arrows) to the redesign strategy actually selected and reported in van Dantzich et al [2002]. Note that an (X) on an arrow denotes inconsistency between identified claim and redesign action. Clearly, the IRC-based questionnaire supported the actual redesign strategy decisions better.

the Scope designers, who chose to revise the way they highlighted new items, but not supported by the original questionnaire results. Numerous other such false design claims emerged from the original but not the revised questionnaire; Figure 8 provides an overview of all of our conclusions.

4.2 Discussion

Our study employs a reusable approach such that other applications can be judged using similar methods. The dual-task usage scenario experienced by participants provides a good model for other studies of notification systems. The questionnaire provides a reusable base that can be applied to other notification systems with design model claims of supporting either alarm, ambient, or alarm-ambient supertask interface functionality for formative and summative evaluation.

In conducting other types of evaluations, the approach we undertook in designing this evaluation can map to other empirical methods or analytic approaches. Our previous work, instrumental in the development of the IRC framework and notification action models, examined the evaluation of notification systems in empirical studies with primary task degradation, timed rapid response tasks, and answer correctness as dependent variables [McCrickard

et al. 2001]. In extending to other evaluation styles, it is necessary to provide realistic experiences and probe the use of the notification system according to tradeoffs among interruption, reaction, and comprehension. For example, the primary task degradation used to study interruption in our empirical studies was examined using questions 5 and 10 in the case study questionnaire (see Figure 7) and could be explored, say, by observing decrease in productivity during high email periods in an ethnographic study in the workplace.

The advantage of this evaluation approach is that knowledge gained can be directly applied to new design processes, isolating design challenges for iterative refinement while retaining the link to critical parameters. As the area advances, there emerges a cataloging of design models and information design claims, providing a richer base for future notification systems researchers to use for comparison and inspiration. The next section examines more closely the utility of the IRC framework and action models, relating to the general multitasking approach and extending the approach to a broader class of computing experiences.

5. CONCLUSION

We have presented a novel approach for classifying and modeling the attention-impacting and utility-producing parameters—interruption, reaction, and comprehension, or *IRC*—that affect the success of notification systems. These contributions can extend far into the HCI community:

- Applying the IRC categorization framework provides a unified view of the notification systems design space and allows an improved usability evaluation process to emerge, as demonstrated by our case study.
- Adopting a common theme, classification system, and evaluation method implementation will increase research cohesion, extendibility, and replicability.
- By weaving critical parameters tightly into the classification process, evaluation design, and claims catalog, we can advance research faster and produce knowledge that is valuable to practitioners.
- Our enhanced ability to articulate the strengths and weaknesses of designs will make systems better suited to user needs and expectations.

While this article specifically addresses notification systems, many of the general concepts discussed here can be more broadly applied for studying other multitasking or collaborative systems. As off-desktop computer usage continues to extend and new applications are introduced at ever-increasing rates, the research community must regear and regroup with approaches that are firmly rooted in science, yet provide Sutcliffe's [2000] notion of "designer digestible" packets of HCI knowledge. Our concept of centering classification systems, cognitive models, UEMs, and claims catalogs on blended critical parameters can be extended to other domains, allowing claims about information and interaction design to be collected in a cohesive, efficient manner.

There is much to do in the way of future work. While we presented a sampling of notification systems that have appeared in recent literature, there are many others. Through collaborative efforts, notification systems researchers

should identify disparate efforts, specify design model objectives, and classify systems accordingly. Years of research and experimentation in this field and others have produced many valuable theories and guidelines, which need to be contextualized to readily apply to our common design space view. These theories should ground claims that correspond to tasks and artifacts associated with IRC characterizations. Standard UEM implementations should be postulated, tested, and adopted for general use. Adaptation of these UEMs can be guided by action models for various IRC combinations. Studies to verify claims should be reported in such a manner that allows establishment of benchmarks and gauging of progress over time, which relies on a common conceptual framework like the IRC design space. The dividends that will result from these efforts provide our motivation—enhancing the computer user experience for a new and exciting generation of notification system applications.

ACKNOWLEDGMENTS

We are particularly grateful for the exceptionally thorough reporting of the Scope notification system by van Dantzich et al., which allowed us to build on their work to accomplish our case study. We thank the anonymous reviewers of this work, who provided very insightful comments that made this work better.

REFERENCES

- ABOWD, G. D. AND MYNATT, E. D. 2000. Charting past, present, and future research in ubiquitous computing. *ACM Trans. Comput.-Hum. Inter.* 7, 1, 29–58.
- ANDERSON, J. R. AND LIEBIERE, C. 1998. *The Atomic Components of Thought*. Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey.
- BADDELLY, A. 1996. Exploring the central executive. *The Quarterly Journal of Experimental Psychology* 49A, 5–28.
- BALLAS, J. A., HEITMEYER, C. L., AND PEREZ, M. A. 1992. Evaluating two aspects of direct manipulation in advanced cockpits. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '92)*. Monterey, CA, 127–134.
- BARNARD, P. AND MAY, J. 1999. Representing cognitive activity in complex tasks. *Human-Computer Interaction* 14, 93–158.
- BARNARD, P. AND MAY, J. 2000. Systems, interactions, and macrotheory. *ACM Trans. Comput.-Hum. Inter.* 7, 2 (June), 222–262.
- BARTRAM, L. 1998. Enhancing visualizations with motion. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis '98)*. Raleigh, NC, 13–16.
- BARTRAM, L., WARE, C., AND CALVERT, T. 2001. Moving icons: Detection and distraction. In *Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT 2001)*. Tokyo, Japan.
- BARTRAM, L. R. 2001. Enhancing information visualization with motion. Ph.D. thesis, Simon Fraser University, Canada.
- CADIZ, J., VENOLIA, G. D., JANCKE, G., AND GUPTA, A. 2001. Sideshow: Providing peripheral awareness of important information. Tech. Rep. MSR-TR-2001-83, Microsoft Research, Collaboration, and Multimedia Group. Sept.
- CHEWAR, C. M., MCCRICKARD, D. S., NDIWALANA, A., NORTH, C., PRYOR, J., AND TESSENDORF, D. 2002. Secondary task display attributes: Optimizing visualizations for cognitive task suitability and interference avoidance. In *Proceedings of the Symposium on Data Visualization (VisSym '02)*. Eurographics Association, Barcelona, Spain, 165–171.
- CUTRELL, E., CZERWINSKI, M., AND HORVITZ, E. 2001. Notification, disruption, and memory: Effects of messaging interruptions on memory and performance. In *Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT 2001)*. Tokyo, Japan, 263–269.

- DAHLEY, A., WISNESKI, C., AND ISHII, H. 1998. Water lamp and pinwheels: Ambient projection of digital information into architectural space. In *Proceedings of the Conference on CHI 98 Summary: Human Factors in Computing Systems*. ACM Press, 269–270.
- DUCHNICKY, R. L. AND KOLERS, P. A. 1983. Readability of text scrolled on visual display terminals as a function of window size. *Human Factors* 25, 6, 683–692.
- ENNS, J. T. AND RENSINK, R. A. 1991. Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review* 98, 335–351.
- ERICSSON, K. A. AND KINTSCH, W. 1995. Long term working memory. *Psychological Review* 102, 2, 211–245.
- FOSTER, K. I. 1970. Visual perception of rapidly presented word sequences of varying complexity. *Perception and Psychophysics* 8, 215–221.
- GRANAAS, M. M., MCKAY, T. D., LAHAM, R. D., HURT, L. D., AND JUOLA, J. F. 1984. Reading moving text on a CRT screen. *Human Factors* 26, 1, 97–104.
- GRAY, W. D. AND SALTZMAN, M. C. 1998. Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction* 13, 3, 203–261.
- GREEN, P. 1999. The 15-second rule for driver information systems. In *Proceedings of the ITS America Ninth Annual Meeting*. Washington, DC, CD-ROM.
- GREENBERG, S. AND FITCHETT, C. 2001. Phidgets: Easy development of physical interfaces through physical widgets. In *Proceedings of the ACM Conference on User Interface Software and Technology (UIST '01)*. Orlando, FL.
- HARRISON, B. L., ISHII, H., VICENTE, K. J., AND BUXTON, W. A. S. 1995. Transparent layered user interfaces: An evaluation of a display design to enhance focused and divided attention. In *Conference Proceedings on Human Factors in Computing Systems (CHI '95)*. ACM Press/Addison-Wesley Publishing Co., 317–324.
- HEALEY, C. G., BOOTH, K. S., AND ENNS, J. T. 1996. High-speed visual estimation using preattentive processing. *ACM Trans. Hum. Comput. Inter.* 3, 2, 107–135.
- HEALEY, C. G. AND ENNS, J. T. 1999. Large datasets at a glance: Combining textures and colors in scientific visualization. *IEEE Transactions on Visualization and Computer Graphics* 5, 2, 145–167.
- HEINER, J. M., HUDSON, S. E., AND TANAKA, K. 1999. The information percolator: Ambient information display in a decorative object. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '99)*. Asheville, NC, 141–148.
- HORVITZ, E. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '99)*. Pittsburgh, PA, 159–166.
- HORVITZ, E., JACOBS, A., AND HOVEL, D. 1999. Attention-sensitive alerting. In *Conference on Uncertainty and Artificial Intelligence (UAI '99)*. Stockholm, Sweden, 305–313.
- HUDSON, J. M., CHRISTENSEN, J., KELLOGG, W. A., AND ERICKSON, T. 2002. “I’d be overwhelmed, but it’s just one more thing to do”: Availability and interruption in research management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. ACM Press, 97–104.
- ISHII, H., WISNESKI, C., BRAVE, S., DAHLEY, A., GORBET, M., ULLMER, B., AND YARIN, P. 1998. ambient-ROOM: Integrating ambient media with architectural space. In *Proceedings of the Conference on CHI 98 Summary: Human Factors in Computing Systems (CHI '98)*. ACM Press, 173–174.
- JUOLA, J. F., WARD, N. J., AND McNAMARA, T. 1982. Visual search and reading of rapid serial presentations of letter strings, words, and text. *J. Exper. Psych. General* 111, 2, 208–227.
- KANG, T. J. AND MUTER, P. 1989. Reading dynamically displayed text. *Behaviour and Information Technology* 8, 1, 33–42.
- KIERAS, D. E. AND MEYER, D. E. 1997. An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction* 12, 4, 391–438.
- MACINTYRE, B., MYNATT, E. D., VOIDA, S., HANSEN, K. M., TULLIO, J., AND CORSO, G. M. 2001. Support for multitasking and background awareness using interactive peripheral displays. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (UIST '01)*. ACM Press, 41–50.

- MAMYKINA, L., MYNATT, E., AND TERRY, M. A. 2001. Time aura: Interfaces for pacing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. ACM Press, 144–151.
- MCCRICKARD, D. S. 1999. Maintaining information awareness with Irwin. In *Proceedings of the World Conference on Educational Multimedia/Hypermedia and Educational Telecommunications (ED-MEDIA '99)*. Seattle, WA.
- MCCRICKARD, D. S., CATRAMBONE, R., AND STASKO, J. T. 2001. Evaluating animation in the periphery as a mechanism for maintaining awareness. In *Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT 2001)*. Tokyo, Japan, 148–156.
- MCCRICKARD, D. S. AND CHEWAR, C. M. 2003. Attuning notification design to user goals and attention costs. *Comm. ACM* 46, 3, 67–72.
- MCCRICKARD, D. S., CZERWINSKI, M., AND BARTRAM, L. 2003. Introduction: Design and evaluation of notification system interfaces. *Inter. J. Hum.-Comput. Studies* 8, 5, 509–514.
- McFARLANE, D. C. 1998. Interruption of people in human-computer interaction. Ph.D. thesis, George Washington University, Washington DC.
- McFARLANE, D. C. 2002. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human Computer Interaction* 17, 3.
- NEWELL, A. 1990. *Unified Theories of Cognition*. Harvard University Press, Cambridge, MA.
- NEWMAN, W. M. 1997. Better or just different? On the benefits of designing interactive systems in terms of critical parameters. In *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '97)*. ACM Press, 239–245.
- NORMAN, D. A. 1986. Cognitive engineering. In *User Centered System Design: New Perspectives on Human Computer Interaction*, D. A. Norman and S. W. Draper, Eds. Lawrence Erlbaum Associates, 31–62.
- REDSTRÖM, J., SKOG, T., AND HALLNÄS, L. 2000. Informative art: Using amplified artworks as information displays. In *Proceedings of DARE 2000 on Designing Augmented Reality Environments*. ACM Press, 103–114.
- RHODES, B. AND MAES, P. 2000. Just-in-time information retrieval agents. *IBM Syst. J.* 39, 3–4, 685–704.
- ROGERS, R. D. AND MONSELL, S. 1995. Cost of a predictable switch between simple cognitive tasks. *J. Exper. Psych. General* 124, 2, 207–231.
- ROSSON, M. B. AND CARROLL, J. M. 2002. *Usability Engineering: Scenario-Based Development of Human-Computer Interaction*. Morgan-Kaufman, New York, NY.
- SEE, J. E., HOWE, S. R., WARM, J. S., AND DEMBER, W. 1995. Meta analysis of the sensitivity decrement in vigilance. *Psych. Bull.* 117, 2, 230–249.
- SHERIDAN, T. 1991. Human factors of driver-vehicle interaction in the IVHS environment. Tech. Rep. DOT HS 807 837, National Technical Information Service, Springfield MA.
- SUTCLIFFE, A. 2000. On the effective use and reuse of HCI knowledge. *ACM Trans. Comput.-Hum. Inter.* 7, 2 (June), 197–221.
- TUFANO, D., KNEE, H., AND SPELT, P. 1996. In-vehicle signing functions and systems concepts. In *Proceedings of the 29th International Symposium on Automotive Technology and Automation (ISATA) Dedicated Conference on Global Deployment of Advanced Transportation Telematics/ITS*. Florence Italy, 97–104.
- VAN DANTZICH, M., ROBBINS, D., HORVITZ, E., AND CZERWINSKI, M. 2002. Scope: Providing awareness of multiple notifications at a glance. In *Proceedings of the 6th International Working Conference on Advanced Visual Interfaces (AVI '02)*. ACM Press.
- WEISER, M. AND BROWN, J. S. 1996. Designing calm technology. *PowerGrid Journal* 1.01.
- WHITTAKER, S., TERVEEN, L., AND NARDI, B. A. 2000. Let's stop pushing the envelope and start addressing it: A reference task agenda for HCI. *Human-Computer Interaction* 15, 75–106.
- WICKENS, C. D. AND HOLLANDS, J. G. 2000. *Engineering Psychology and Human Performance*, Third ed. Prentice Hall, Upper Saddle River, NJ.

Received July 2002; revised April 2003; accepted August 2003