Brain–computer interfaces (BCIs) hold great promise for improving information delivery and preserving user attention, but this promise has not yet translated to practical use. A prototype BCI that optimizes email notifications in noisy, complex environments, CARSON combines multiple measures from the brain to predict both cognitive workload and message relevancy to determine the optimum time to interrupt the user.

People consume information at many different times and in many different ways. In 1971, political scientist and economist Herbert Simon pointed out that “a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.” Since Simon’s time, demands on our attention have greatly increased. At work or home or on the go, we use multiple devices to read news articles, shop online, stay updated with our social networks, sift through instant messages and emails, stream videos, and track our activities. Emerging augmented-reality technologies such as smart glasses and car head-up displays move information from peripheral to immediately noticeable locations. As people engage with more and more data, attention has become a precious resource.

The consequences of misappropriating attention are well documented. Consuming too much or the wrong information, or consuming information at the wrong moment, can not only decrease working performance but also increase stress and anxiety. For example, an analysis of 10,000 programming sessions found that the average programmer takes 10–15 minutes to resume editing code after a single interruption.

While the advent of novel information delivery services and new types of devices has expanded our capacity to access information, our capacity to perceive and understand information has not changed. This tension between greater information availability and biological constraints on attention leaves user interface researchers at a crossroads: How can we prevent increasing amounts of information from overwhelming users?
When is the best time to give users new information so that it won’t be disruptive? Which information is most helpful (or unhelpful) at any given moment? What is the best way to deliver timely information to users?

Physiological signals from the brain could serve as inputs to systems that help manage user attention. While brain–computer interfaces (BCIs) hold great promise for improving information delivery, this promise has not yet translated to practical use. CARSON (Cognitive Automatic Real-Time Selection of Notifications) is a prototype BCI that combines multiple measures from the brain to regulate email delivery in complex environments. We use CARSON to explore the space of attention-aware systems, outlining design principles that enable BCIs to operate in noisy conditions. Although significant challenges lie ahead for BCIs in real-world scenarios, we envision an interface that could someday optimize information delivery to match users’ constantly changing needs.

TOWARD INTELLIGENT INFORMATION DELIVERY

In everyday social interactions, humans draw on a wealth of information acquired from personal experience and observation and act on this information in real time. For example, teachers who detect boredom in their students will use more demonstrative gestures and modify the tone of their voice to regain the students’ attention. Student engagement improves because teachers perceive the problem and quickly adapt to address it.

Computing devices cannot recognize the often subtle social cues that indicate a cognitive state like boredom, happiness, or frustration. Consequently, they cannot adapt to retain our interest or prevent us from becoming overwhelmed. Yet people unconsciously treat computers as social entities, and research suggests that “polite” computers can significantly improve user interaction. If the computer is a collaborator, it is often a clueless one.

This limitation underscores a fundamental problem in human–computer interaction (HCI): the amount or quality of information that a computer has about us pales in comparison to what we naturally intuit about one another. The bandwidth of communication between computers and users is too narrow to construct an attentive system.

In the hope of broadening this bandwidth, researchers have begun monitoring the brain as a powerful source of input to build an intelligent system that always delivers relevant information when and how users expect it. Unfortunately, as Edward Cutrell and Disney Tan pointed out, “reality is a bit more sobering; we are a long way from the Matrix.” The brain is complex, neural data from sensors is noisy, and users’ natural working environments often introduce complex and difficult-to-control variables. Consequently, while BCIs have existed for decades, their translation to everyday life has largely remained a vision for the future.

The sidebar “BCIs and Attention Awareness” outlines the promises and challenges of this approach and provides context for our work.

CARSON: AN INTELLIGENT INTERRUPTION SYSTEM

Playfully named for the butler who adeptly navigates complex social settings in the hit British TV series Downton Abbey, CARSON is a BCI that passively monitors signals from the brain and uses those measures to deeply personalize the computing experience to improve how people engage with information. In particular, it leverages functional near-infrared spectroscopy (fNIRS) signals to predict both the user’s cognitive workload and the relevance of incoming emails to determine the optimum time to interrupt the user and deliver the emails. To address some of the challenges of building functional, attention-aware BCIs in the wild, CARSON applies design strategies that mitigate the impact of misclassifications on the user experience. This is critical as physiological computing systems begin to migrate to more general, less-controlled environments.

Measuring brain activity

Observing many of the same physiological parameters as functional magnetic resonance imaging (fMRI), fNIRS is an optical brain-sensing technology that fundamentally consists of light sources and detectors, as Figure 1 shows. To detect brain activity, fNIRS sends light in the near-infrared range (650–900 nm) into the head, where it is diffusely reflected by the scalp and skull but absorbed by the oxygenated and deoxygenated hemoglobin in the brain. Following a banana-shaped path, a small portion of the light sent into the head returns to the sensor, allowing researchers to calculate the concentration and oxygenation of blood at the brain’s surface.

Because active regions of the brain demand more oxygen, these parameters can be used to estimate localized brain activity at depths of 1–3 cm. Despite a somewhat slow physiological response on the order of 5–8
When it comes to incorporating physiological signals from the brain into intelligent systems, there are many open questions about how to process this unique, but potentially noisy, input: What information should be emphasized? When should it be delivered to the system? How should it be presented to the user? The building blocks for brain–computer interfaces (BCIs) continue to appear in research, even if they are not explicitly cast in that light. Taken together, they offer a promising but still incomplete picture of future information delivery.

The fact that attentional resources are finite is foundational to modern psychology and cognitive science.1 The more information sources (distractions) there are, the fewer resources that are available to devote to a given task. Consequently, building computers that automatically preserve user attention has been a focus of human–computer interaction (HCI) research for more than 20 years.2 In presenting an early system designed to improve email notifications, Eric Horvitz, Andy Jacobs, and David Hovel wrote that “human attention is the most valuable and scarcest commodity”3 in HCI. Attention awareness likewise was the basis for the CLUES system, which used personal information in a person’s work environment to automatically prioritize messages.4

For the past decade, researchers have explored using BCIs to manage user attention. Daniel Chen and Roel Vertegaal used physiological sensing to identify interruptible moments,5 Anton Nijholt and Desney Tan proposed using BCIs for intelligent systems,6 a team led by Jan van Erp described a BCI framework for multimodal interaction and multitask environments,7 and Thorsten Zander and his colleagues advocated BCI-based implicit control of interfaces.8

Our own work has applied BCIs to information filtering, a valuable mechanism commonly realized in recommendation systems that helps users deal with the profusion of choices characteristic of online entertainment.9 Using functional near-infrared spectroscopy (fNIRS) as the sole input to a proof-of-concept movie recommender, we found that implicit measures of preference can be used to recommend movies that subjects rate higher than when recommended by a control condition. Unlike systems that require users to explicitly rate or flag their preferences, ours makes recommendations passively so that the user need not expend cognitive resources.

Despite a growing consensus that physiological computing systems in general, and BCIs in particular, offer solutions to information overload, many practical hurdles to their implementation remain. Improvements in signal processing and sensor design are needed to increase the classification accuracy of models. In addition, many studies focus on a single cognitive state, which might not be enough to capture complex real-world conditions.

References
seconds, fNIRS is relatively tolerant to minor movement artifacts, increasing its potential to attain ecologically valid measurements.5

Predicting interruptibility
Physiological sensors such as fNIRS devices can detect periods of low and high cognitive workload. Low-workload periods are often correlated with task breakpoints, moments when users can handle incoming information without severe disruption. While user workload is a valuable heuristic for determining opportune moments of interruption, research indicates that the incoming information’s relevance is also critical. For example, a study of Microsoft developers and managers found that turning off email notifications resulted in some users checking their email account more frequently to avoid missing an important message.6

Given this tension between the user’s desire to focus on a task and to maintain a high level of awareness about incoming information, the context of an interruption is clearly important. However, reliably predicting both message relevance and engagement without direct user input has been challenging. How can we design systems to function in working environments where the cost of mistakenly delaying a critical message is high? What measures can we use to better capture the multidimensional complexity of interruptions?

As Figure 2 shows, CARSON aims to intelligently adapt to complicated real-world scenarios by combining two measures: the potential cost of an interruption based on how busy the user is (when are you available?) and the relevance of incoming notifications (which messages are most

Irrelevant email is deferred when user has high workload
Relevant email is delivered immediately
All emails are delivered immediately when user has low workload

Irrelevant email is delivered when workload drops

High workload
Low workload

Time

FIGURE 1. Using functional near-infrared spectroscopy (fNIRS) to measure oxygenation levels in the blood that correspond with low and high brain activity. (a) A user wears fNIRS sensors while interacting with CARSON. (b) The probes secured on users’ foreheads each have four light sources and one detector. (c) Light entering brain tissue follows a banana-shaped path from source to detector. (Photo credits: Tufts University/Steffan Hacker)

FIGURE 2. Combining predictions about the user’s cognitive workload with an email’s relevance allows CARSON to deliver messages at the most opportune moment.
important to you?). It uses fNIRS signals to predict cognitive workload and message importance and then incorporates these measures into an interruptibility formula that builds on Eric Horvitz and Johnson Apacible’s calculation of the expected cost of interruption (COI): 7

\[
COI = (W_{\text{wl}} \times \text{COI}_{\text{wp}})(W_{\text{util}} \times \text{COI}_{\text{util}}),
\]

where COI\text{wp} represents the cost of an interruption when cognitive workload fluctuates between low (0) and high (1), and COI\text{util} represents the cost of an irrelevant email (1) versus a highly relevant email (0). Each is weighted by CARSON’s confidence in those values. These weights are critical in designing a system that can function in different environments.

**Designing for complex environments**

Automating information delivery is dangerous because people reasonably expect important messages not to be delayed. Given unpredictable environments and noisy sensors, however, inaccurate classification models could be an unavoidable constraint in the near future. As a result, we applied two design principles to CARSON: we created graded, non-critical responses to brain input, and personalized the system to consider the environment’s complexity.

**Building noise into adaptation.** The most obvious way to deal with a costly interruption by an incoming message is to defer notification to a better time for the user. However, this can strip users of their sense of control. A notification should not be deferred indefinitely simply because the system has not determined a good moment for the interruption.

To counter this problem, CARSON slowly decreases the cost of a delayed interruption over time, regardless of cognitive workload, making delivery more and more likely. When CARSON has high confidence in its predictions, it will delay the message longer. When it has low confidence, it sets a much lower maximum wait time. This graded strategy guarantees that lengthy delays occur only when the system is sure that a message is not important.

**Personalizing interruptions.** Selecting the ideal threshold at which a user should be interrupted is also challenging. If the threshold is set too high, all messages will be deferred; if the threshold is set too low, all messages will immediately be delivered. This problem is more acute with physiological data, as the translation of real-time measurements to classifications results in wide variations in model accuracy.

CARSON manages its threshold based on a distribution of classification probabilities that occurs during actual system use—it selects the point that maximizes the distance between the probabilities of truly relevant and irrelevant emails. This approach creates personalized interruption parameters that are sensitive to noise in the environment.

**TRAINING CARSON TO READ THE BRAIN**

Intelligent systems can automatically identify user state in real time. As with many BCIs, CARSON requires a calibration period to “learn” how a user’s physiological signals correspond to his or her cognitive workload.

As an example of this process, CARSON learns to predict the difference between interruptions containing important information and those that contain unimportant information. This capability extends work by Erin Solovey and her colleagues, who demonstrated that fNIRS could detect three multitasking states: branching, in which a user performs secondary tasks while still holding in mind the primary task’s goals; delay task, in which a user ignores secondary tasks; and dual task, when a user frequently shifts between the primary task and secondary tasks. 8 For CARSON, we hypothesized that dealing with irrelevant email notifications would increase brain activity, while relevant notifications (which support the primary task) would ultimately reduce it.

To predict cognitive workload, we repeatedly interrupt users working on a difficult information-monitoring task with either important or unimportant notifications for 15 minutes.
CARSON monitors each interruption’s impact on the corresponding fNIRS signal for 20 seconds. Each light source/detector pairing in the fNIRS probes is considered a different channel of information. As Figure 3 shows, we record each information source’s signal activity over time, extract high-level features that describe the signal, and input that data into a model that predicts message relevancy. Although the best combination of signal features remains an open question, inputting the mean change in and slope of the signal to a support vector machine yields high accuracies.\(^9\)

Once these fNIRS features are trained, CARSON streams them to the model for real-time classification. The output is both a prediction of cognitive workload and a probability (or confidence value) of that classification. We use these probabilities—which are critical to negating harmful misclassifications—to weight our classifications in calculating COI.

We tested our model’s accuracy on 14 participants during real-time use in an interruption task and discovered that 12 exhibited data at above-chance levels. However, the average probability estimate between relevant and irrelevant emails differed by only 10 percent on average (less than zero is below chance). While this provides weak evidence that fNIRS can detect the value of incoming messages, we viewed it as an opportunity to test CARSON’s design principles in real-world working environments that are difficult to control.

**TESTING CARSON**

To put CARSON to the test, we constructed a hypothetical scenario in which users act as information specialists for a news station. Their job involves monitoring a Twitter feed about the day’s events and periodically retweeting messages to keep followers informed. As Figure 4 shows, users are instructed to click on every third tweet of each topic they are assigned to follow. These topics are assigned by the system, and can vary in both content and number. The more topics a user must follow, the more challenging the task becomes. Manipulating the number of topics is a way to explore high-workload scenarios in which interruptions are disruptive versus low-workload moments when users can handle additional information.
While tracking their assigned topics, users are also instructed to respond to emails that are either relevant or irrelevant to the information-monitoring task. Relevant emails specify that users should prioritize messages from a particular Twitter account and stop doing their normal job until they identify that account. Irrelevant emails act as distractors, mentioning tweets about topics the users are not currently monitoring.

**Experimental design**

Given our goal of moving BCIs into real-world environments, we conducted a within-subjects experiment using our hypothetical scenario. We sought to answer two main questions: do misclassifications harm user performance, and how does CARSON impact users’ sense of control of their information? The study involved 14 participants from 18 to 23 years old, with a mean age of 21. To improve experimental control and focus on a physiological signal that is less prevalent in literature (relevance), we simulated the detection of cognitive workload.

For each study participant, we built a model to predict message importance by delivering a series of relevant and irrelevant email notifications and recording the user’s fNIRS signals associated with each interruption.

Existing email systems such as Gmail estimate message priority. We simulated this model by running a second session in which we predicted the relevance of new emails that participants engaged with. The resulting accuracy of these predictions served as the basis of CARSON’s real-time assignment of message priority.

We tested participant performance under three conditions:

- **Adaptive**—CARSON calculates COI based on email relevance and the user’s simulated cognitive workload, trying to delay notifications to an opportune moment for the user.
- **Immediate**—CARSON immediately delivers all email notifications to the user.
- **Maladaptive**—CARSON inverts its calculated COI. This helps determine whether the sensors are providing meaningful information. If they are, there should be a significant difference in performance between the maladaptive and adaptive conditions.

For each condition, participants were interrupted by 20 relevant and 20 irrelevant notifications. Users interacted with seven 45-second trials of interruptions during high cognitive workload (monitoring two topics) and seven during low cognitive workload (monitoring one topic), presented in random order. The independent variable was notification delivery time.

For each condition, we recorded the following dependent measures:

- **Twitter task performance**—this was calculated using the Levenshtein edit distance between strings representing ideal performance and the user’s performance; and
- **the number of errors that we could attribute to participants not responding quickly enough to an important email.**

After testing, we surveyed the study participants to determine whether they found CARSON to be useful, confusing, or efficient, and how much control they felt they had over the system, under each of the three conditions.

**Experimental results**

In our experiment, users did significantly better under the adaptive condition than under the maladaptive condition in both Twitter task performance ($t(13) = –3.1049, p < 0.01$) and in responding to relevant emails ($t(13) = –7.79, p < 0.0001$). As Figure 5 shows, these results suggest that although CARSON’s interruption strategy has a low classification rate, fNIRS signal data can still be meaningfully mapped to message relevancy.

User performance did not significantly differ under the adaptive and immediate conditions. Because CARSON has very low confidence in identifying important emails, it severely capped the maximum time to delay incoming notifications. Irrelevant emails were delayed, on average, only about 5 seconds. Messages were thus rarely delayed until a natural task breakpoint, indicating that under low-accuracy conditions CARSON will only occasionally aid but will not hurt the user. This is encouraging for future work: increased model accuracy would extend CARSON’s willingness to delay a message to a more opportune moment for the user.

In their survey responses, study participants rarely noted differences between the three conditions. One user observed that the adaptive condition was “easier to handle/manage than the last one” (referring to the immediate condition). More generally, users did not express any loss of control or inefficiency between one condition and another. This attitude is best reflected by a participant who stated, “It seemed like it was the
same as before.” In light of known system misclassifications, this suggests that users might not detect CARSON’s adaptiveness.

LESSONS FROM CARSON
Our work on CARSON has yielded several insights about designing BCIs and deploying them in the wild.

Complex problems need more than one measure
Current applications and devices provide various mechanisms to help manage information. Gmail, for example, tries to capture the importance of emails with a priority inbox, and many smartphones have settings to assign higher priority levels to specific contacts. CARSON aims to expand the bandwidth between user and computer—to build an attention-aware system that has access to more information than existing applications and leverages untapped physiological resources to present that information to the user in the best possible way.

The advantage of physiological signals is that they can be recorded without direct user involvement, thus preserving users’ attentional resources. And because physiological responses can be generalized to broader contexts, they could become the basis for comparing and prioritizing text messages, emails, and other types of notifications across domains. However, physiological responses to different user states could overlap or interfere with one another, and using multiple models increases the risk of misclassification. Future BCI research must resolve these challenges as the technology continues to mature and support more complex user problems.

Designing for noise and misclassification
System interventions can disrupt a user’s perception of control, especially when they are the result of misclassifications. This presents a dilemma for designers of interruption systems driven by brain input because accuracy can be difficult to guarantee. However, as our participants showed, users generally are not aware of the timing or content of incoming notifications until they arrive, allowing the system some flexibility in manipulating their delivery.

Our study also highlights the effectiveness of graded adaptation strategies. CARSON’s physiological-based email deferral policy provided subtle adaptations. While users performed significantly better in the adaptive and immediate conditions, the maladaptive condition was not disruptive enough to register on any dimension of our survey.

Brain and body metrics are increasingly accessible with the emergence of wearable computing. At the same time, the amount of information we process has grown rapidly. Our devices are capable of remarkable power and speed, but they also remain insensitive to our mental states and thus frequently cause annoying and disruptive interruptions.

However, CARSON and other comparable systems provide concrete examples of how physiological measures can improve the way we engage with information. While inaccurate classification models and noisy signals have traditionally impeded the translation of BCIs to the consumer space, good design principles could help mitigate the impact of misclassification and improve the viability of this emerging technology.

At the same time, CARSON’s measurement of cognitive workload is not the end game. Capturing the richness
of our environment will likely involve many metrics that weigh numerous contextual factors, and this raises several challenging questions: What is the best combination of sensors to capture user state? What states can be detected simultaneously? How do user movement and changing environments impact measures of attention?

Our computing devices could soon leverage information from brain and body sensors to respond in a manner that is more appropriate or familiar to us. As technology advances, we can begin to imagine an entirely new generation of BCIs that are sensitive to our changing state—optimizing how, when, and which information is delivered to us. Such interfaces have the potential to someday transform our computers from tools to collaborators.

REFERENCES