Models of Attention in Human-Computer Interaction: From Principles to Applications

Eric Horvitz Carl Kadie Tim Paek David Hovel

Microsoft Research
Redmond, Washington 98025
horvitz@microsoft.com

1 Introduction

One of the main results of Twentieth-century Cognitive Psychology is that, despite the overall impressive abilities of people to sense, remember, and reason about the world, our cognitive abilities are extremely limited in well-characterized ways. In particular, psychologists have found that people grapple with scarce attentional resources and limited working memory. Such limitations become salient when people are challenged with remembering more than a handful of new ideas or items in the short term [20,28], recognizing important targets against a background pattern of items [5,26], or interleaving multiple tasks [6,26].

These results indicate that we cannot help but to inspect the world via a limited spotlight of attention. As such, we often generate clues implicitly and explicitly about what we are selectively attending to and how deeply we are focusing. Given constraints on attentional resources, it is no surprise that communication among people relies deeply on attentional signals. Psychologists and linguists studying communication have recognized that signaling and detecting attentional states lies at the heart of the fast-paced and fluid interactions that people have with one another when collaborating or communicating [2,7]. Attentional cues are central in decisions about when to initiate or to make an effective contribution to a conversation or project. Beyond knowing when to speak or listen in a conversation, attention is critical in detecting that a conversation is progressing. More generally, detecting or inferring attention is an essential component of the overall process of *grounding*—converging in a shared manner on a mutual understanding of a communication [1].

The findings about our limited attentional resources—and about how we rely on attentional signals in collaborating—have significant implications for how we design computational systems and interfaces. Over the last five years, our team at Microsoft Research has explored, within the *Attentional User Interface* (AUI) project, opportunities for enhancing computational systems and applications by treating human attention as a central construct. As an organizing principle, we consider attention as a rare commodity—and critical currency—in reasoning about the information awareness versus disruption of users [12]. We have also pursued the use of attentional cues as an important source of rich signals about goals, intentions, and topics of interest [10,15]. We seek to build systems that sense, and share with users, natural signals about attention to support conversations and other forms of fluid mixed-initiative collaborations with computers [24]. Moving to considerations of computational efficiency, an assessment of a user's current and future attention can be employed to triage computational resources. Investigations in this realm include selective allocation of resources in rendering graphics via relying on models [14,16] or on direct observations [21] of visual attention, and in guiding precomputation and prefetching [11] with forecasts of future attention. Finally, although there is a rich history of prior work on attention from cognitive psychology, we have found that there is much we do not yet understand. Thus, beyond pooling results from prior psychological studies, we need to continue to perform user studies that adapt or extend prior results on attention and memory from cognitive psychology to human-computer interaction [3,4,18,19].

We will describe several principles and methodologies at the heart of research on integrating models of attention into human-computer interaction. Then, we shall review representative efforts that illustrate how we can harness these principles in attention-sensitive messaging and mixed-initiative interaction applications.

2 Models of Attention and Decision Making under Uncertainty

How might we access and use information about a user's attention? To be sure, subtle clues about attention are often available, and a number of these clues can be taken as direct signals about the attentional status of users. For example, sensing patterns of simple gestures such as the touching and lifting of a device in different settings can relay important evidence about attention that can be exploited in a number of exciting ways [8]. Moving to higher-precision sensing, several researchers have pursued the use of gaze-tracking systems, and have used signals about the focus of visual attention in a variety of applications [17,25,27]. As gaze sensors grow in reliability and decrease in cost, we are seeing the evolution of devices that recognize when and how they are interrogated by the spotlight of visual attention.

Nonetheless, we may often be *uncertain* about a user's attentional focus and workload in light of observations, and about the value of alternate actions in different contexts. Thus, we turn to models that can be harnessed to *reason* about a user's attention and about the ideal attention-sensitive decisions to take under uncertainty. Such models and reasoning can unleash new functionalities and user experiences.

We have constructed by hand and learned from data Bayesian models that can be viewed as performing the task of an automated "attentional Sherlock Holmes," working to reveal current or future attention *under uncertainty* from an ongoing stream of clues. Bayesian attentional models take as inputs sensors that provide streams of evidence about attention and provide a means for computing probability distributions over a user's attention and intentions.

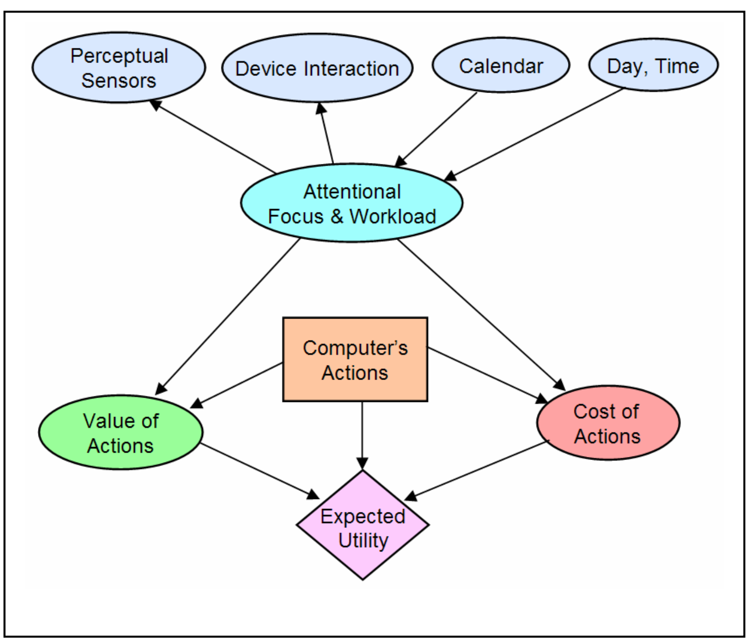

Figure 1. High-level decision model considering a user's attentional focus and workload as a random variable, influenced by the observed state of several sensors.

Perceptual sensors include microphones listening for ambient acoustical information or utterances, cameras supporting visual analysis of a user's gaze or pose, accelerometers that detect patterns of motion of devices, and location sensing via GPS and analysis of wireless signals. However, more traditional sources of events can also offer valuable clues. These sources include a user's online calendar and considerations of the day of week and time of day. Another rich stream of evidence can be harvested by monitoring a user's interactions with software and devices. Finally, background information about the history of a user's interests and prior patterns of activities and attention can provide valuable sources of information about attention.

To build probabilistic attentional models with the ability to fuse evidence from multiple sensors, we leverage the results of accelerated research over the last fifteen years on representations for reasoning and decision making under uncertainty. Such work has led to inferential methods and representations including Bayesian networks and influence diagrams—graphical models that extend probabilistic inference to considerations of actions under uncertainty. Algorithms have been developed which enable us to compute probability distributions over outcomes and expected utilities of actions from these graphical representations.

Figure 1 displays a high-level influence diagram representing sensor fusion and decision making in the context of a user's attention under uncertainty. As portrayed in the figure, a set of variables (oval nodes) representing sensed evidence influence a random variable representing a user's attentional status which, in turn, influences the cost and benefits and overall expected value of alternate actions or configurations. Decisions (rectangular node) about ideal computer actions take into consideration the costs and benefits, given uncertainty about a user's attention. In the end, the expected utility (diamond-shaped node) is influenced by the action and the costs and benefits.

We extend such a high-level, pedagogical view by constructing richer models that contain additional intermediate variables and key interdependencies among the variables. Also, as both devices and people are immersed in time, we move pointwise considerations of the state of variables, to build higher-fidelity temporal attentional models that represent the flow of time. We have employed dynamic Bayesian networks and Hidden Markov Models for representing and reasoning about states of attention and location over time.

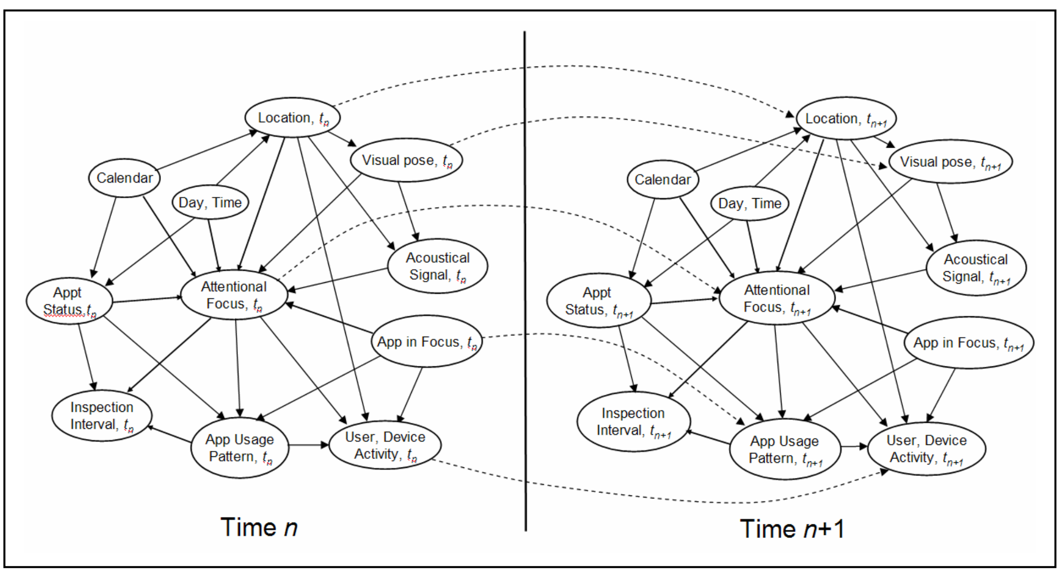

Figure 2. A temporal attentional model, highlighting key dependencies (dashed arcs) between variables in adjacent time slices.

Figure 2 displays two adjacent time slices of a temporal attentional model. Such a model provides a probability distribution over a user's workload and task that was developed for an application that provides selective filtering of messages and communications to users. In this case, the status of attention includes approximately twenty discrete states.

3 Economic Models of Attention and Information

As we can all attest from personal experiences, computers today have little awareness of the value and costs of transmitting messages and alerts to users. Research on the *Notification Platform* project has centered on formulating economic principles of attention-sensitive notification—and on implementing a cross-device alerting system based on these principles. The Notification Platform system modulates the flow of messages from multiple sources to devices by performing ongoing decision analyses. These analyses balance the expected value of information with the attention-sensitive costs of disruption. As highlighted in Figure 3, the system serves as an attention-savvy layer between incoming messages and a user, taking as inputs sensors that provide information about a user's attention, location, and overall situation.

The design of the Notification Platform was informed by several earlier prototypes exploiting context-sensing for identifying a user's workload, including the *Priorities* system [12,13]. Priorities employs classifiers that predict the urgency of incoming email. The classifiers are trained with sample messages, either obtained via explicit training or by automatically *drafting* data sets by observing a user's interaction with an email browser. Priorities also observes a user's patterns of presence at a desktop computer based on time of day, and infers the time until a user will review unread messages. The system computes an *expected cost of delayed review* for each incoming message. This cost is considered, along with a cost of interruption based on activity sensing and calendar information, in automated decisions about if and how to alert and relay information to a user about email, tasks, and appointment reminders in mobile and desktop settings.

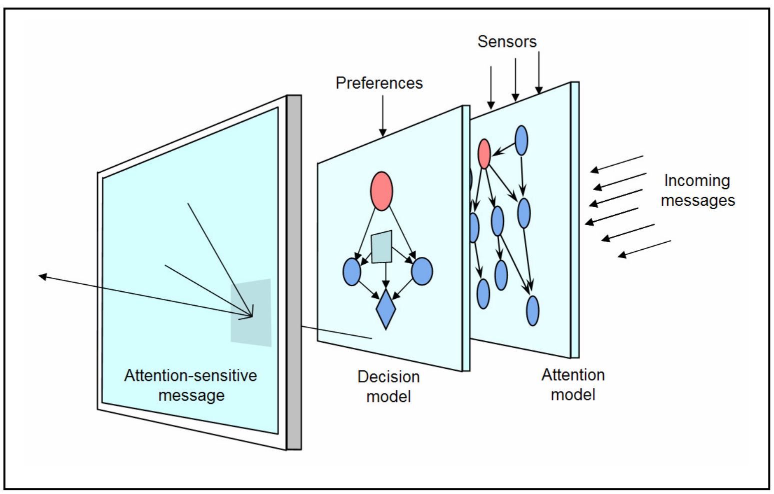

Figure 3. The Notification Platform is a cross-device messaging system that continues to balance the costs of disruption with the value of information from multiple message sources, using a probabilistic model of attention and ongoing decision analyses about ideal message handling.

.

The Notification Platform uses a decision-analytic model for cross-device alerting about information from multiple message sources. The analyses consider a user's attention and location under uncertainty, as well as the fidelity and relevance of potential communication channels. We developed a distributed architecture that executes over multiple devices. Figure 4 displays a schematic view of the architecture of the Notification Platform. Standard interfaces and metadata schemas allow users to subscribe different sources of information and devices to a *Notification Manager*/ At the heart of the Notification Manager is a Bayesian attention model and decision analysis which accesses clues about attention and location from sensors via a module we refer to as a *Context Server*.

The context server accesses several states and streams of evidence, including a user's appointments from Microsoft Outlook, events about device presence and activity, an analysis of ambient acoustics in the room, and a visual analysis of pose using a Bayesian head-tracking system. Key abstractions from the evidence, such as *voice trace detected, task completion occurred within 5 seconds, single application focus, head-tracked—looking away from display*, and *meeting away from office—ending in 10 minutes*, are posted to a volatile store called the *Context Whiteboard* which is continually updated by incoming evidence. The Context Whiteboard is contacted for updated information every couple of seconds by the Bayesian attentional model in the Notification Manager.

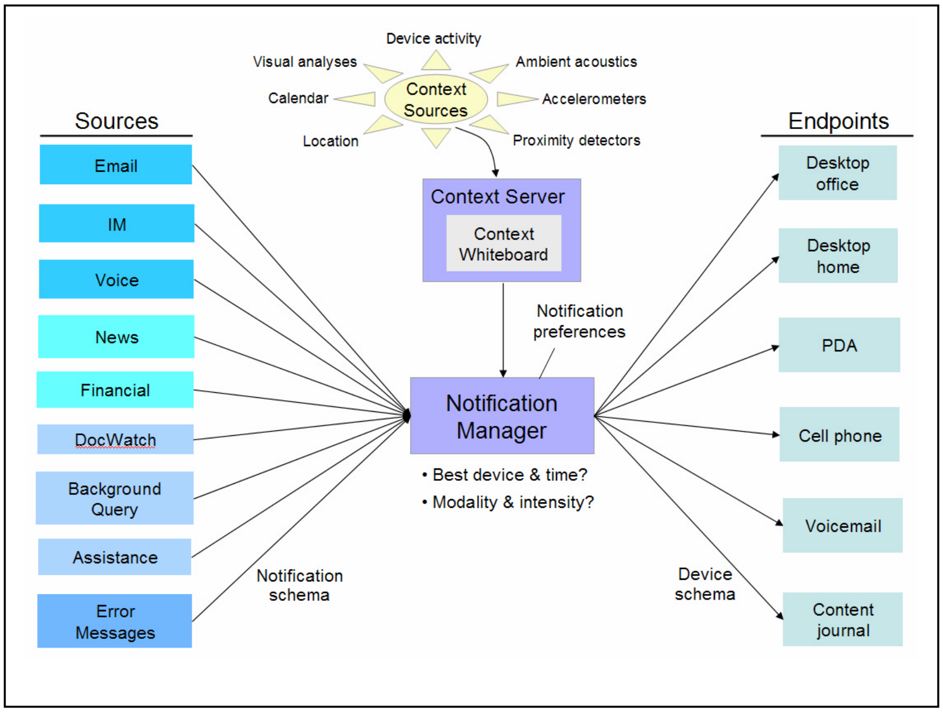

Figure 4. Constellation of components of the Notification Platform. Sources and devices are subscribed to the Notification Manager via a set of standard interfaces. The Notification Manager accesses sensor findings from multiple devices to deliberate about information value, attention, and the best channel and alerting modality.

The Notification Manager's decision analysis weighs the expected costs and benefits of alerting a user about messages coming into the system's *Universal Inbox*. In computing the costs of disruption, the decision model considers the probability distribution over a user's attentional state and location in several places in its analysis, including the cost of disruption associated with different alerts for each device, the availability of different devices, and the likelihood that the information will reach the user when alerted in a specific manner on a device.

The ongoing expected-utility analysis is performed in accordance with a user's preferences, stored in a profile. These include assertions about the cost of disruption for each alert modality, conditioned on users being in different attentional states. As an example, for the case of a desktop computer, the system makes available a set of display alternatives as the product of different visual displays of the alert (*e.g.*, thumbnail, full-alert display) and several auditory cues (*e.g.*, no auditory clue, soft chime, louder alert). The placement of the alert with regards to the current focus of visual attention or interaction is also considered.

Figure 5 captures the deliberation of the Notification Platform about incoming messages. The system computes the expected value of receiving an alert as the difference between the value of alerting the user now and the value that will be obtained by reviewing the alert later. Given probability distributions over a user's attention and location inferred from its sensors, Notification Platform iterates over all alerting and display modalities for each device with an expected-utility analysis to decide if, when, and how to alert a user. As represented by the metaphor of a narrowing funnel in Figure 5, the system considers, for each device and modality, the loss in fidelity of information transmitted. In addition, the system considers the likelihood that an alert will be received, given inferred probability distributions over the attention and location of the user. This reliability of transmission is represented metaphorically in the figure as the chance that a message will make it through a slot in a spinning disk. In the end, the attention-sensitive costs of disruption are subtracted from estimates of the value of alerting, yielding a net value of alerting a user for each channel and alerting modality. The channel and modality with the highest expected value is selected.

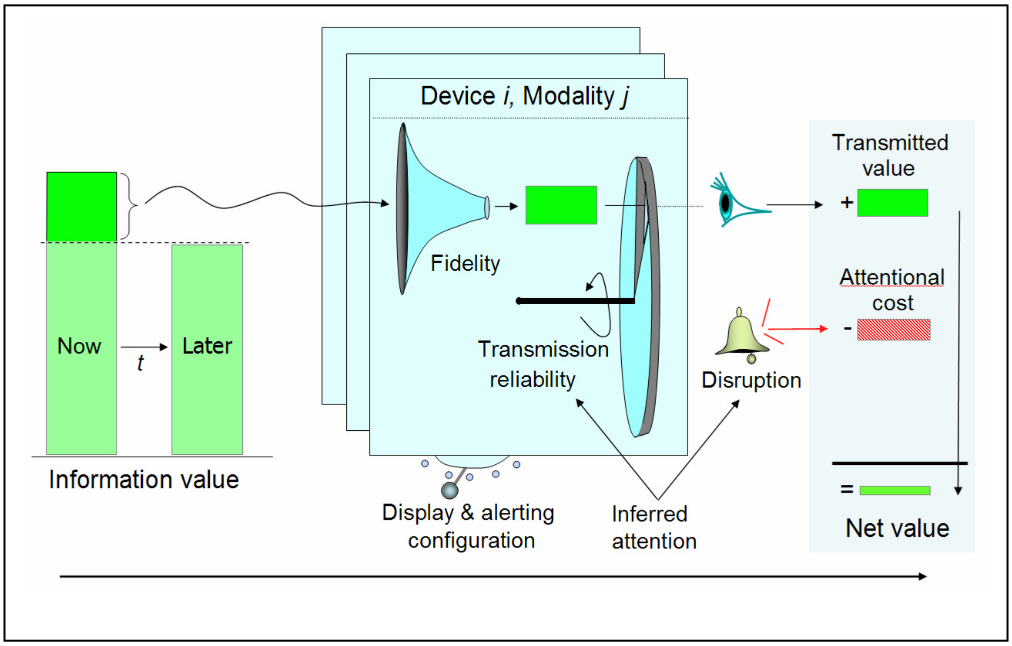

Figure 5. Graphical depiction of cost-benefit analyses in the Notification Platform. The attention-sensitive costs of disruption and value of information are considered, along with the losses due to decreased fidelity (narrowing funnel) and transmission reliability (spinning slotted disk) associated with the use of each alerting modality of all subscribed devices.

Figure 6 displays several aspects of the behind-the-scenes functioning of the Notification Platform. A *context palette* displays current findings drawn from sensor sources. Several views into the decision analysis are displayed, including inference about the time-varying attention of the user. At the current time, the user is inferred to be most likely in a state named *high-focus solo activity*, which has competed recently with *low-focus solo activity*, *conversation in office*, and other less likely states. The Universal Inbox displays messages from several sources, including email, instant messaging, breaking news, and stock prices. Messages have also been received from *DocWatch*, an agent subscribed to by the user that identifies documents of interest for the user. Each message is annotated with the best device and alerting policy, and the associated net expected dollar value of relaying the messages with that channel and mode is indicated. As portrayed in the inbox, it is worthwhile passing on to the user two instant messages. Other alerts are *in the red*, as the cost of disruption dominates the net value of information. In this case, the ideal alerting mode and channel for an instant message is determined to be a visual notification in a large format coupled with an audio herald at the user's desktop system.

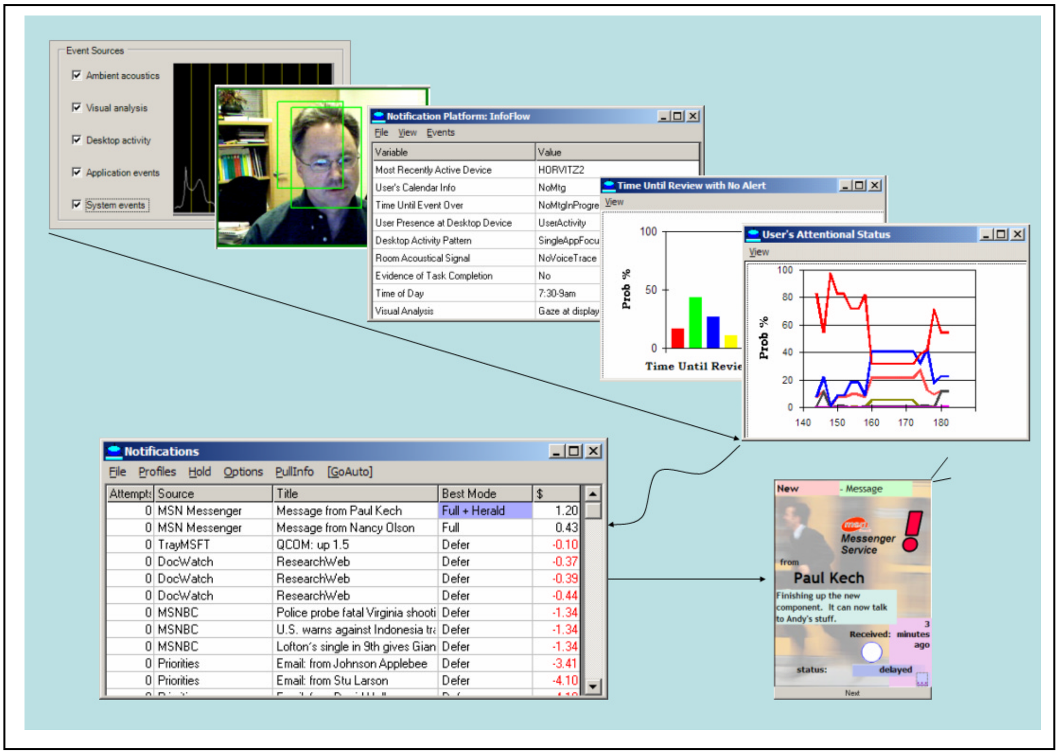

Figure 6. View of portion of the Notification Platform's reasoning. At run time, information from multiple sensors is posted to the Context Whiteboard and fused to infer the user's attentional status and location. Multiple notifications are sorted by net expected value and the channel and alerting modality with the highest expected utility is selected

Ongoing research on the Notification Platform project includes the refinement of preference assessment tools to ease the task of encoding preferences. Currently, users can adjust sliders to change a set of predefined defaults on costs of interruptions. Another key area of work centers on using machine learning for building probabilistic models of attention, location, and cost of disruption from data. Results from machine-learning efforts have been applied to refine the Notification Platform [13,22].

As highlighted in Figure 7, we have also been working to make small devices aware of the attentional status and location of users—and either reporting local sensor information to the central Notification Manager or performing local notification management and related services based on the inferred attention [8]. This research includes the challenges of embedding and leveraging multiple perceptual sensors on small devices, including GPS, 802.11 signal strength, accelerometers, infrared proximity detectors, and touch sensors. Part of this work has explored opportunities for developing devices, such as cell phones that behave with more insight about their disruptiveness by considering coarse models of attention [9].

Additionally, we are continuing to pursue psychological studies of disruption. Formal studies of the costs of disruption began with the early work of Ovsiankina [23] and Zeigarnik [29] nearly seventy-five years ago. The rich body of work in this realm includes studies on memory, problem solving, and overall task efficiency in the face of disruptions. More recent work includes efforts by our team [3,4] and other groups [*e.g.*, 18,19] to probe the influence of notifications of various types and saliencies on the efficiency and satisfaction with performing a variety of computer tasks. The psychological studies and results complement the mathematical models; the economic models provide a principled, flexible foundation which can integrate costs uncovered by user studies as parameters considered in expected-utility decision making.

Figure 7. Sensing PDA, outfitted with multiple perceptual sensors, including proximity, motion, and touch sensors. In the background, accelerometer signals are displayed showing the motion fingerprint of a user walking while looking at the device.

4 Attention, Initiative, and Interaction

In another area of research, we have investigated the use models of attention to enhance the robustness and fluidity of human-computer collaboration. Some of this work focuses on the recognition of attentional cues as coordinative signals in *mixed-initiative* interaction with computing devices. In mixed-initiative interaction, both users and computers take turns in contributing to a project or an understanding [10]. The turn taking of conversational dialog is a prototypical example of mixed-initiative interaction. Psychologists have found that people having conversations with one another rely on attentional cues to signal when a contribution is going to be offered or has been accepted [2]. We have sought to endow computers with an analogous ability to recognize and emit signals to guide the nature and timing of contributions and clarifications in support of mixed-initiative interaction.

DeepListener [15] and Quartet [24] represent efforts in mixed-initiative interaction to incorporate attention in spoken language systems. Both systems tackle what we have referred to as the *speech-target problem*: When a computer with an open microphone and speech recognizer hears an utterance, how is it to recognize that it is being addressed when there are other people or listening devices in a room? DeepListener and Quartet explicitly address this challenge with probabilistic models that infer the likelihood that they are the target of speech.

DeepListener uses a model of attention and intention to guide clarification dialog in a spoken command and control setting. The system considers its uncertainty about whether it is the target of speech, what it has heard, and the likelihood of different intentions. The system continues to make expected-utility decisions about carrying out actions in the world, or about how it should approach users, if necessarily, to clarify their intentions before taking actions. These decisions take into consideration the utilities of alternate dialog actions and the stakes of the world actions.

DeepListener shares its attention and availability by gracefully changing the colors and intensities of an *attentional lens* that glows on its control panel, and via gestures and thoughts of an animated agent that it renders. These affordances provide cues that assist with conversational turn taking.

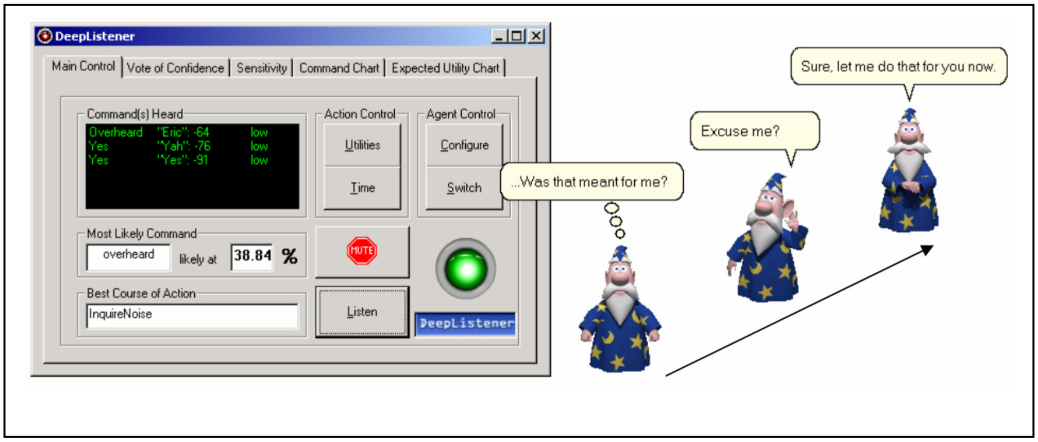

Figure 8. Reasoning about the target of speech and clarification dialog. DeepListener first makes an expected utility decision to quietly share its thoughts about the possibility that it is the target of an utterance. Given additional recognitions, it goes ahead to seek clarification, and finally executes an action for the user.

Figure 8 displays a situation where DeepListener has heard an utterance that was first directed elsewhere in a noisy environment. After analyzing a new utterance a bit later, the system engages the user in a clarification dialog, and then invokes a desired action.

Quartet operates with a continuous speech recognition system, and incorporates a richer model of attention under uncertainty. It examines keyboard events, an analysis of the content and the coherence of natural language parsing, and visual pose analysis to ascertain the attentional status of the user and system with regards to the establishment, maintenance, and disruption of attention between the user and system [24]. Quartet couples speech recognition to a natural language processor. The system continues to attempt to parse the noisy utterances that it has recognized in speech to infer a user's intentions within focused contexts. Figure 9 shows Quartet listening to a user talking *about* the system rather than speaking *to* the system. In this case, Quartet is being used as an assistant to control via voice commands the navigation of slides displayed in a presentation. Utterances directed *to* Quartet about navigation arise intermittently during the more dominant stream of ongoing utterances associated with the presentation. In this example, the user is talking *about* the computer, and, based on a fusion of the user's language and visual pose, Quartet infers that the user is likely speaking to someone else.

Our ongoing research on mixed-initiative and spoken language systems is focusing on several challenges, including the use of sensed or inferred attention to provide clues about a user's intentions, the content and context at hand, and the nature and ideal timing of appropriate contributions. This work includes using sensed or inferred attention to inform speech recognition systems about the specific microcontexts being addressed with utterances. Such narrowing of the spotlight of analysis can be useful for enhancing recognition as it can enable spoken dialog systems to swap in appropriate language models and semantics, and scope of possible actions. Also, robust solutions to the speech-target problem promise to influence significantly the overall sociology of human-computer interaction, by allowing users to interact with multiple devices and people in their proximity with speech and gestures in a manner similar to the way that people interact with one another.

In another realm of innovation, computers with an ability to track and to understand attentional patterns *among people* engaged in conversations can provide new kinds of services and facilities. For example, methods for identifying visual attention among participants in a conversation can be used to automate the control of cinematography, and to capture, organize, and understand a group meeting or videoconference [27]. Thus, beyond enhancing human-computer interaction, sensing and reasoning about attention promises to enhance the way we communicate and collaborate with one another.

5 Conclusion

We described efforts to endow computing systems with the ability to sense and reason about human attention. After reviewing some background on the nature and importance of attention in cognition and discourse, we discussed methods for inferring attention from multiple streams of information, and for leveraging these inferences in decision making under uncertainty. Then, we presented illustrative applications of the use of attentional models in messaging systems and in mixed-initiative interaction. Research on the use of models of attention in computing systems is still in its youth. We expect that continuing refinement of methods for recognizing, reasoning, and communicating about attention will change in a qualitative manner the way we perceive and work with computers.

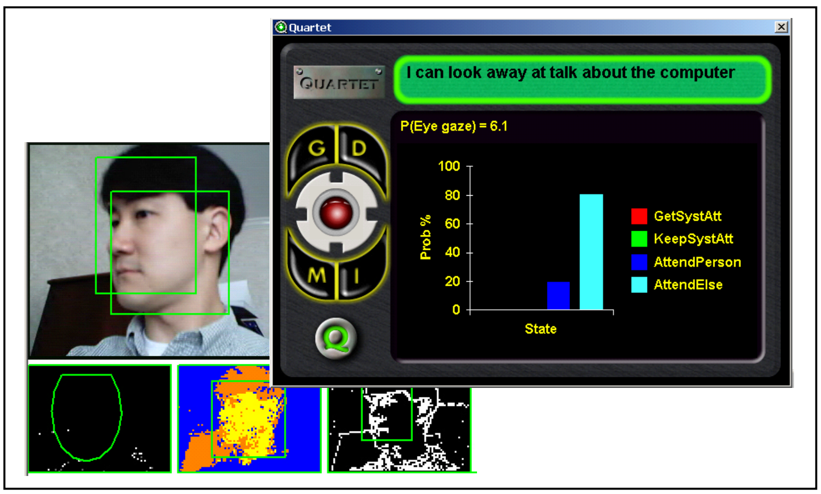

Figure 9. Quartet in action. Quartet's partial recognition is displayed at the top of the display. The system's beliefs about the attentional status of the user, with regards to initiating, maintaining, or breaking out of conversational dialog, is represented as a dynamically changing probability distribution.

Acknowledgments

Contributors on the constellation of efforts on the AUI project at MS Research include Johnson Apacible, Ed Cutrell, Mary Czerwinski, Susan Dumais, Ken Hinckley, David Hovel, Andy Jacobs, Carl Kadie, John Krumm, Paul Koch, Nuria Oliver, Tim Paek, John Platt, Daniel Robbins, Chaitanya Sareen, Joe Tullio, Maarten van Dantzich, and Andy Wilson.

Bibliography

1. H.H. Clark and S.A. Brennan, S.A. *Grounding in communication*. In Perspectives on Socially Shared Cognition, APA Books, pp.127-149., 1991.

2. H.H. Clark and E.F. Schaefer. *Collaborating on contributions to conversations*. Language and Cognitive Processes, 2/1, pp.19-41, 1987.

3. E. Cutrell, M. Czerwinski, and E. Horvitz. Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance, *Proceedings of Interact 2001: IFIP Conference on Human-Computer Interaction*, Tokyo, Japan, July 2001.

4. M. Czerwinski, E. Cutrell, and E. Horvitz. Instant messaging: Effects of relevance and time, In S. Turner, P. Turner (Eds), *People and Computers XIV: Proceedings of HCI 2000, Vol. 2*, British Computer Society, p. 71-76, 2000.

5. C.W. Eriksen and Y.Yeh. (1985), Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 11:583-597, 1985.

6. T. Gillie and D. Broadbent. What makes interruptions disruptive? A study of length, similarity and complexity *Psychological Research*, 50, 243-250, 1989.

7. B.J. Grosz and C.L. Sidner. *Plans for discourse*. In Intentions in Communication, 417-444. MIT Press, 1990.

8. K. Hinckley, J. Pierce, M. Sinclair, and E. Horvitz, Sensing Techniques for Mobile Interaction, In *Proceedings of the ACM UIST 2000 Symposium on User Interface Software and Technology*, November 2000, San Diego, California, pp. 91-100.

9. K. Hinckley and E. Horvitz. Toward More Sensitive Mobile Phones**,** *Proceedings of the* ACM *UIST 2001 Symposium on User Interface Software and Technology*, pp. 191-192.

10. E. Horvitz. Principles of Mixed-Initiative User Interfaces. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, Pittsburgh, PA, May 1999, pp. 159-166. ACM Press.

11. E. Horvitz. Principles and Applications of Continual Computation, In: *Artificial Intelligence Journal,* 126:159-196, Elsevier Science, February 2001.

12. E. Horvitz, A. Jacobs, and D. Hovel. Attention-Sensitive Alerting. In: *Proceedings of the Fifteenth Conference on Uncertainty and Artificial Intelligence*, Stockholm, Sweden, July 1999, pp. 305-313. Morgan Kaufmann: San Francisco.

13. E. Horvitz, P. Koch, C.M. Kadie, and A. Jacobs. Coordinate: Probabilistic Forecasting of Presence and Availability. *Proceedings of UAI '02, Proceedings of the Eighteenth Conference on Uncertainty and Artificial Intelligence*, Edmonton, Canada, July 2002, pp. 224-233.

14. E. Horvitz and J. Lengyel. Perception, Attention, and Resources: A Decision-Theoretic Approach to Graphics Rendering. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, Providence, RI, August 1997. Morgan Kaufmann: San Francisco, pp. 238-249.

15. E. Horvitz and T. Paek. Harnessing Models of Users' Goals to Mediate Clarification Dialog in Spoken Language Systems. In *Proceedings of the Eighth International Conference on User Modeling*, July 2001.

16. D. Luebke, M. Reddy, J. Cohen, A. Varshney, B. Watson, and R. Huebner. *Level of Detail for 3D Graphics*. Morgan Kaufmann, 2002.

17. P. P. Maglio, T. Matlock, C.S. Campbell, S. Zhai, and B.A. Smith, Gaze and Speech in Attentive User Interfaces, In: *Proceedings of the International Conference on Multimodal Interfaces*, Springer Verlag, 2000.

18. D. Scott McCrickard, Richard Catrambone, C. M. Chewar, and John T. Stasko. Establishing Tradeoffs that Leverage Attention for Utility: Empirically Evaluating Information Display in Notification Systems. *International Journal of Human-Computer Studies*, 2003.

19. D. McFarlane. Coordinating the interruption of people in human-computer interaction. Human-Computer Interaction - INTERACT'99, IOS Press, Inc., The Netherlands, pp. 295-303, 1999.

20. G. Miller. The magical number seven plus or minus two; some limits on our capacity for processing information. *Psychological Review*, 63:81-97, 1956.

21. H. Murphy and A.T. Duchowski. Gaze-Contingent Level of Detail. Proceedings of Eurographics, Manchester, September 2002.

22. N. Oliver, E. Horvitz, A. Garg. Layered Representations for Recognizing Office Activity, In *Proceedings of the Fourth IEEE International Conference on Multimodal Interaction*, Pittsburgh, PA, October 2002, pp. 3-8.

23. M. Ovsiankina. Die wiederaufnahme unterbrochener handlungen. *Psychologische Forschung*, 11:302-379, 1928.

24. T. Paek and E. Horvitz. Conversation as action under uncertainty. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pp. 445-464. AUAI, Morgan Kaufmann, August 2000.

25. T. Selker, A. Lockerd, J. Martinez, and W. Burleson. Eye-aRe: a Glasses-Mounted Eye Motion Detection Interface, In Proceedings of the CHI2002 Conference on Human Factors in Computing Systems. Minneapolis, MN.

26. G. Sperling and M. J. Melchner. The attention operating characteristic: Examples from visual search. *Science*, 202:315--318, 1978.

27. R.Vertegaal. The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration. In *Proceedings of ACM CHI'99 Conference on Human Factors in Computing Systems*. Pittsburgh: ACM Press, 1999.

28. N. Waugh and D. Norman. Primary memory. *Psychological Review*, 72:89-104, 1965.

29. B. Zeigarnik. Das behalten erledigter und unerledigter handlungen. *Psychologische Forschung*, 9:1-85, 1927.