Designing a Technology Coach

BY WENDY A. ROGERS, IRFAN A. ESSA, & ARTHUR D. FISK

A multidisciplinary team models a system that can alert users of a complex medical device when they make an error.

Technology in the home has the potential to support older adults in a variety of ways. The success of such technology depends on understanding the needs and capabilities of the user and on developing the technology to provide seamless and appropriate support.

Our goal was to develop a technology "coach" that could support older adults in learning to use a medical device – in this case, a blood glucose meter. Our approach was interdisciplinary: It involved human factors/ergonomics (HF/E) researchers with expertise in cognitive psychology and a computer scientist with expertise in computer vision.

Based on our analysis of user capabilities and task demands, we developed a computer vision system that could noninvasively observe, track, recognize, and interpret a person's interaction with the meter. We assessed the relative benefits of different feedback types to correct errors. This research illustrates the potential for the development of in-home personal assistants and the necessity for interdisciplinary approaches to the design of "smart" home technologies.

Potential Benefits of a Technology Coach in the Home

Health care is a concern to adults of all ages, but particularly to older adults, who often have at least one chronic condition such as arthritis, hypertension, or diabetes. Monitoring chronic conditions and learning new medical procedures, usually accomplished with some technology, are often part of their daily routines. Unfortunately, such technology can be difficult to master. Innovations in automated activity recognition may support the use of home medical devices by providing instruction, monitoring use, guiding interpretation of results, troubleshooting errors, and reminding about maintenance tasks.

Consider the following scenario:

> Mrs. Q. has recently been diagnosed with diabetes. She uses a blood glucose meter daily to monitor her glucose levels. She sits at her kitchen table to perform the glucose check, and an automated system records her activities, recognizes when she has made an error, and provides her with corrective feedback to ensure that she performs the procedure correctly. This automated coach will help her learn to calibrate the device and properly check her glucose levels. The system will also provide her with guidance in interpreting the results and determining whether she should eat, take medicine, or exercise more to regulate her glucose.

Technology in the home can support the activities of its older residents as a virtual assistant or technology coach. In an "aware home," technology is designed with intelligence to support the activities of people living there (*http://www.awarehome.gatech.edu*). Given existing sensing technology, one can recognize the activities of an individual (Moore, Essa, & Hayes, 1999) and use that information to provide guidance, much as a human caregiver can. However, to design an effective coach, one must first obtain the answers to a number of questions. What does the technology have to "know" about the human's action? How can advanced technology provide instruction and support the user's performance? What types of feedback can older adults successfully use?

FEATURE AT A GLANCE: Technology in the home environment has the potential to support older adults in a variety of ways. We took an interdisciplinary approach (human factors/ergonomics and computer science) to develop a technology "coach" that could support older adults in learning to use a medical device. Our system provided a computer vision system to track the use of a blood glucose meter and provide users with feedback if they made an error. This research could support the development of an in-home personal assistant to coach individuals in a variety of tasks necessary for independent living.

KEYWORDS: home technology, medical devices, support for learning

Many health-related tasks – especially those involving medical devices – consist of sequential steps. One error in the process may invalidate the entire sequence. Planning these steps can be cognitively intensive, and it may be especially difficult for older adults to monitor where they are in a sequence of actions or if they have made an error. The system itself provides little feedback about performance accuracy, yet the consequences of errors are high. Devices of this type include blood glucose meters, blood pressure monitors, heart rate monitors, oxygen tanks, and infusion pumps.

We selected the blood glucose meter as an exemplar technology for developing our coaching system because of the complexity of the glucose monitoring task and the limited instructions provided by manufacturers of these devices (Rogers, Mykityshyn, Campbell, & Fisk, 2001). Experienced users can make errors using these meters (Colagiuri, Colagiuri, Jones, & Moses, 1990; Hancock, Fisk, & Rogers, 2001), and even with well-designed training programs, both younger and older adults make errors using and calibrating the device (Mykityshyn, Fisk, & Rogers, 2002).

An Interdisciplinary Approach

Older users have unique needs, capabilities, and limitations that must be considered throughout the design process (Fisk, Rogers, Charness, Czaja, & Sharit, 2004). In addition, significant engineering challenges had to be addressed before we could implement a technology coach. The development of our technology coach required guidance from human factors/ergonomics as well as computer science (see Figure 1). We had to (a) model user needs through identification of task demands and understanding of needs and capabilities of the target user population– namely, older adults; (b) develop a system that could capture information from the environment in an unobtrusive manner; (c) use this information to recognize actions; (d) interpret activities being performed and identify if an error has been made; and (e) provide feedback to support task performance and learning. These steps provide the framework for our discussion.

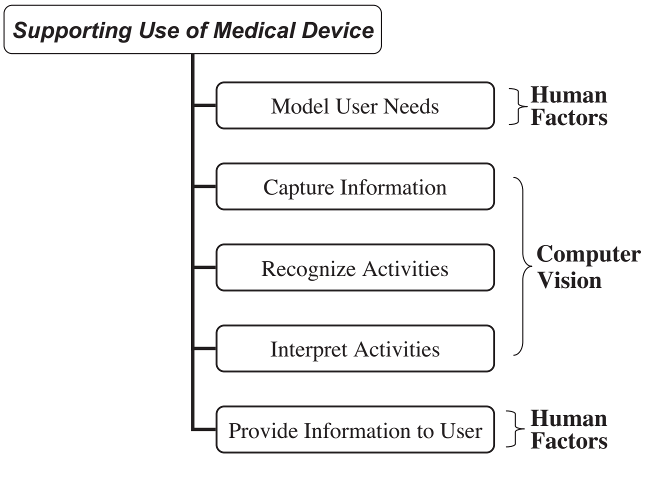


Figure 1. Overview of the interdisciplinary research approach.

Model User Needs

Identify task demands. We conducted an in-depth task analysis of several blood glucose meters. These devices are often advertised in terms such as, "It's as easy to use as 1, 2, 3. Just set up the meter, check the system, and test your blood." Yet, our analysis revealed that more than 50 substeps are required to perform the three basic steps (Rogers et al., 2001). We defined each task in terms of the information the user would need to complete it (task/knowledge requirements), the feedback provided by the system, and the potential problems that might arise if the task was not carried out properly. Many tasks required knowledge of the correct procedure and of the location and function of the control buttons.

We specifically designed our technology coach to increase awareness of the sequence of steps (as determined by the task analysis), to detect errors, and to provide immediate feedback to correct the error.

The most striking finding from our task analysis was the relative complexity of this supposedly simple medical device. More than 70% of the users in our survey reported difficulties learning to use their specific device (Rogers et al., 2001): They had trouble remembering steps, setting up and calibrating the meter, using the lancet, getting a blood sample, and reading the display.

The sequential nature of the task might contribute to problems because an error in an early step carries through to the other steps (e.g., inserting the strip incorrectly creates a lack of proper calibration, which may lead to an incorrect reading). Thus, it is crucial that users receive appropriate instructions about how to use the system safely and effectively. The task analysis provided detailed information about the task demands that would be required for the development of a successful technology coach. The steps identified in the task analysis served as the model for identifying and interpreting the users' actions.

Understand user needs and capabilities. Understanding user needs and capabilities is crucial to system design. We focused on cognitive capabilities such as working memory (the ability to keep information active while processing it); in this case, remembering the multiple steps involved in correctly using the blood glucose meter while performing them. Research has shown that working memory declines through life, beginning as early as age 30 and becoming more severe after age 65 (Wilson et al., 2002). Consequently, we specifically designed our technology coach to support working memory – to increase awareness of the sequence of steps (as determined by the task analysis), to detect errors, and to provide immediate feedback to correct the error.

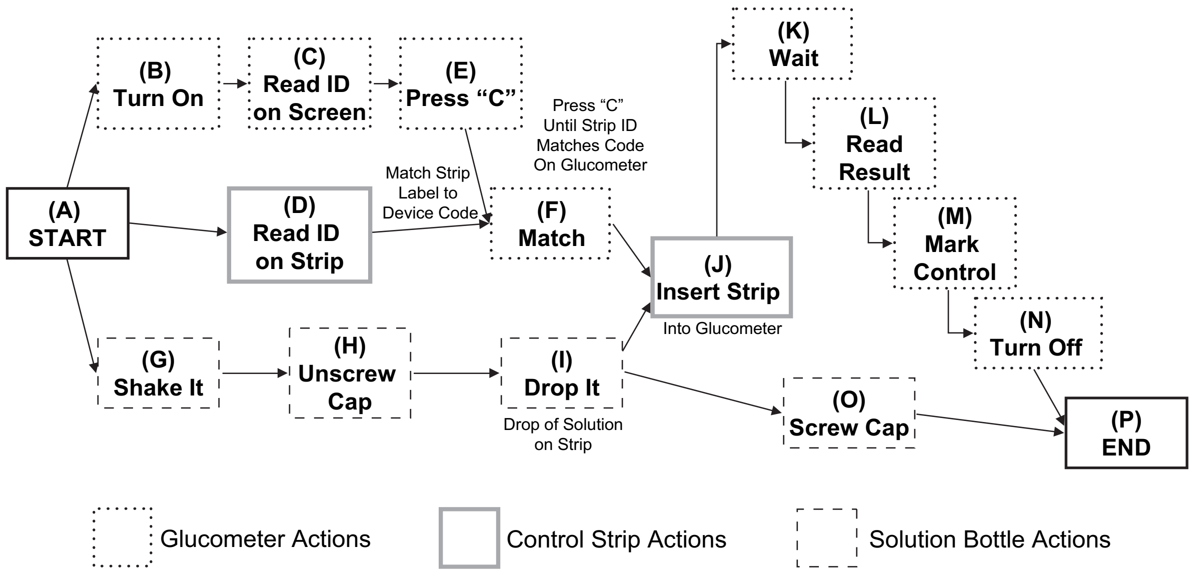


Figure 2. Conceptual diagram for the glucose calibration task.

We relied on general knowledge about age-related perceptual capabilities to ensure that the older adults in our study could adequately see and hear the feedback, but the more complicated issue of whether they could *comprehend* the feedback and use it effectively was the focus of study.

Information Capture, Recognition, and Interpretation

Developing a noninvasive system to observe and accurately recognize activities is a challenging but worthwhile research activity for computer vision researchers. Automated recognition of daily activities provides the basic contextual information one must have to implement a range of assistive technologies, smart appliances, and aware environments.

Consider the example of reading a book. The primitive intervals, as well as the temporal relationships, include the following: "First, fetch the book; next, look at the book while occasionally flipping the pages; finally, put down the book." Even this relatively trivial example suggests that one needs a variety of relationships to represent activity:

- *Sequential streams:* There is a natural partial ordering of components.
- *Duration of elements:* The primitives are not events but have temporal extent.
- *Multiple parallel streams:* Many intervals may occur in parallel.
- *Logical constraints:* Some intervals can be satisfied by a disjunction of subintervals.
- *Nonadjacency:* Sequenced intervals may not meet but may only be ordered.
- *Uncertainty of underlying vision component:* Extracted features will always be noisy.

Extensive research has been conducted on developing systems that recognize, annotate, or respond to user activity (e.g., Aggarwal & Cai, 1999; Vaswani, Roy-Chowdhury, & Chellappa, 2003). However, most approaches consider activity as a single stream of events.

We developed a new approach for modeling and recognizing activities for the sole purpose of coaching blood glucose monitor use. First, we presumed that elemental or primitive intervals are basic units that are sequenced to define higher-level activities. Second, we assumed temporal and logical constraints (for example, one must remove the litmus paper from its container before installing it into the meter).

It was clear that older adults could benefit from feedback, both immediately and after a retention interval, but that the feedback must be specific to be effective.

We devised a representational mechanism and interpretation method that explicitly encoded the glucose meter task. We propose a new representation schema, *Propagation Networks* (P-Nets), and a corresponding inference algorithm called *D-Condensation*. A P-Net represents an activity by associating one event node in the network with each primitive event in the activity (see Figure 2 and Shi, Huang, Minnen, Bobick, & Essa, 2004, for details). Two dummy nodes represent the start and end of the activity. Links in the network correspond to partial order constraints between pairs of actions. The nodes in the network were based on task analysis findings of the proper sequential steps for performing tasks such as calibrating the blood glucose meter.

Sensing system for information capture. We constructed the system in a layered framework (see Figure 3, next page). At the bottom layer is an input stream of raw sensor information – in the case of our study, both video frames and a data stream from the glucose meter itself. The tracking information and the device state served as input to Bayesian networks that asserted instantaneous primitives such as unscrewing the

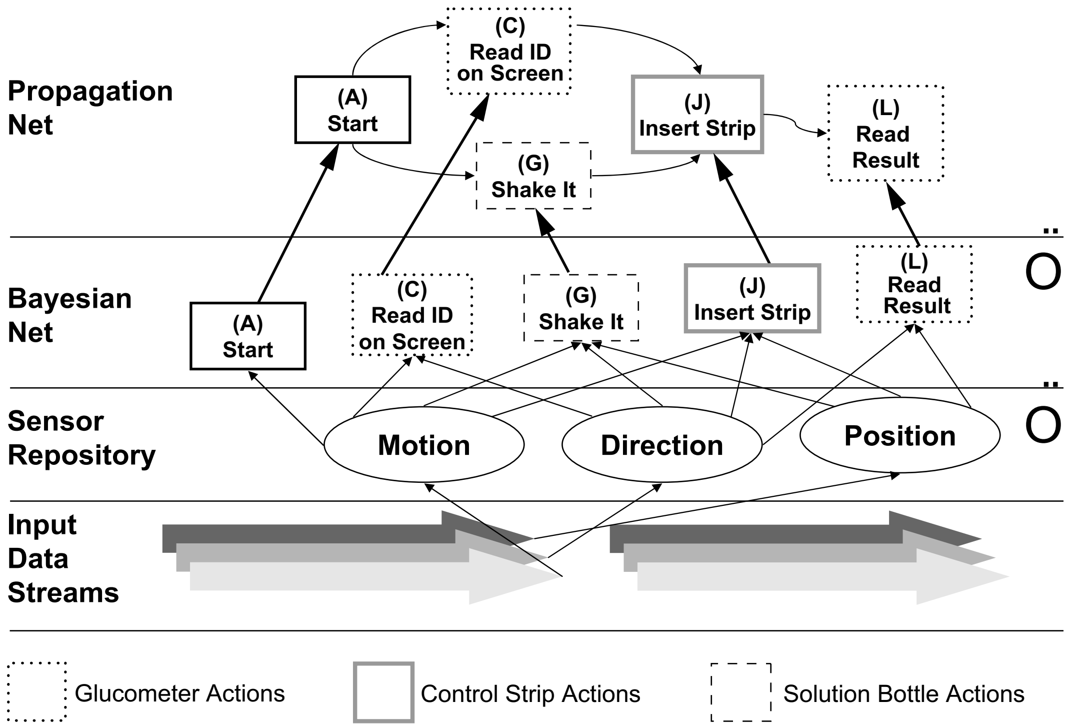


Figure 3. System architecture.

cap or reading results from the meter's screen. These networks served as the observation models for the P-Net (represented at the top of the system architecture diagram).

We constructed a vision-tracking system that used particle filters to track multiple objects, including hands, the testing strip, the liquid bottle, and the glucose meter. We created one tracker for each object and randomly initialized its particle locations. Two statistical features – color histograms and orientation histograms – measured the similarity between the image and the template corresponding to the particle state and thus allowed computation of the particle likelihood. Both features were computationally simple and insensitive to variations in image scaling and rotation. Figure 4 (page 21) provides tracked key frames of the sequence, with each representing one salient event node in the P-Net.

Recognize and interpret activities. We built a 16-node P-Net representation for the blood glucose meter calibration procedure. Three participants performed 41 sequences: 21 were correct, 10 were missing one step, and 10 were missing six steps. For training, we used six correct sequences and saved

Table 1. Overall Computational Performance

Measure	Data
Sequence length range	[232, 928]
Average speed (frames/s)	122.7
Maximal distinctive particles	238
Maximal subsequent states	1967

the rest for testing. The middle-level output from the Bayesian networks was poor because the low-level detectors generated too many false alarms. The temporal constraints encoded in the P-Net, however, caused the final labeling to be much better than earlier indicators suggested.

The D-Condensation algorithm we used is very fast and more than sufficient for real-world applications. The computational statistics are summarized in Table 1, and final results are listed in Table 2 (page 21). All correct sequences were recognized. Eight of the 10 missing-one-step sequences were identified, whereas the other two were labeled as correct. Identification errors were caused by insertion errors in the vision module that made the sequences statistically indistinguishable from correct sequences.

We evaluated the labeling of each frame, specifying whether a particular node in the P-Net was active or inactive (see Table 3, page 22). Although there was a range of individual labeling ratios, the overall correct ratio for any sequence was very high (over 98%), and the average correct positive ratio was higher than 87%.

In summary, the computer vision system was able to capture, recognize, and interpret the activities of the person who was using the blood glucose meter. The next step was to provide users with corrective feedback if they omitted an action, performed an action at the wrong time, or performed an incorrect action.

Provide Effective User Feedback

Feedback, simply defined, is action taken by an external agent to provide information with regard to some aspect of task performance (Kluger & DeNisi, 1996). Findings in the literature have been mixed on virtually every aspect of feedback delivery, including the ideal timing or content (McLaughlin, Rogers, & Fisk, 2006).

We assessed whether older adults could benefit from feedback from the technology coach and whether feedback effectiveness varied based on the content of the information provided. We used a "Wizard of Oz" technique to simulate what the system would ultimately do (e.g., Dahlbäck, Jonssön, & Ahrenberg, 1993). An experimenter played the role of the technology coach (unbeknownst to the participant) so we could assess feedback benefits while developing the recognition system. (Portions of these data were presented in Fiesler, McLaughlin, Fisk, & Rogers, 2003.)

Future efforts must address the engineering challenges that remain, as well as psychological issues related to the design of effective feedback systems.

Participants received feedback only when they made mistakes. The eventual goal of this feedback training would be for users to operate the meter on their own and feedback would no longer be necessary for accurate performance. The idea is that when the medical device is used in the home setting, performance would be monitored for errors and corrective feedback would be provided. However, the feedback should be designed to support learning so that participants can use the medical device when they are away from home as well.

We manipulated each type of feedback in the following ways:

1. *Action feedback:* Participants received instructions on exactly what action to take to correct the mistake. For example, if they inserted the strip upside down into the meter, the feedback might be, "Turn the strip over."
2. *Concept feedback:* Participants received instructions on what to do to correct their mistakes without being told how to do it. With this same error, the feedback might be phrased as, "The strips should be inserted pink side up."
3. *Nonspecific feedback:* Participants were told if they had made an error but not how to correct it. Action and concept training are important for training older adults to use computer-based systems (e.g., Mead & Fisk, 1998). The nonspecific feedback condition provided a comparison to assess the benefits of general feedback.

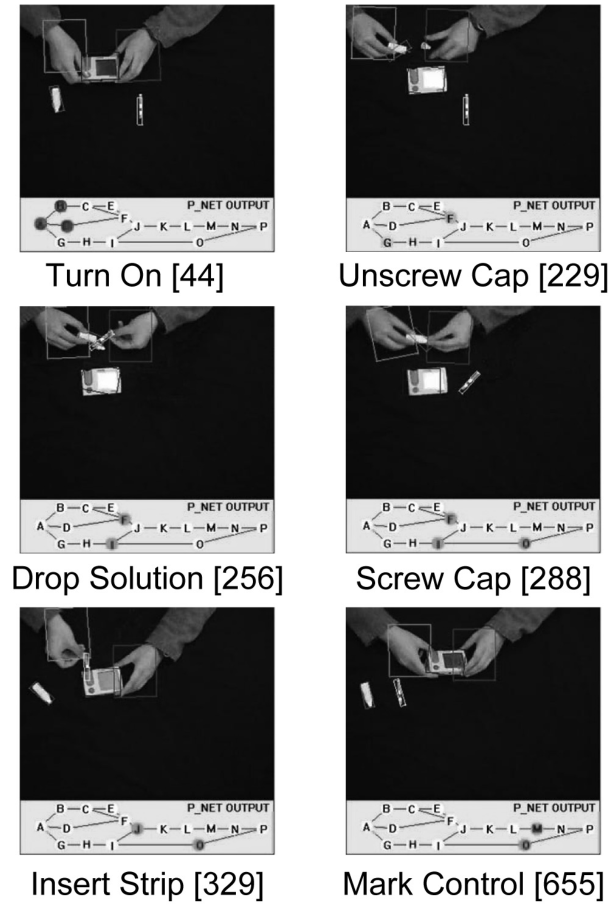


Figure 4. One of our test data sequences with the P-Net shown below it for various actions. The bracketed numbers show the frame number in the sequence. The grayed action nodes in the P-Net output shows P-Net belief on whether the action is occurring. (Refer to Figure 2 for the actions represented by the nodes.)

Participants were 30 older adults ranging in age from 65 to 75. Ten were assigned to each feedback condition. They received initial instructions via the manufacturer's video and were then asked to perform a glucose control calibration test. They received computerized feedback if they made an error as they tried to use the system. We measured trials to criterion, which was performing the calibration without making a single error. We measured performance immediately after training and again after a 2-day retention interval to provide insight into learning.

TABLE 2. OVERALL EVALUATION

Sequence Category	Total	Correct (%)	Almost Right (%)	Negative (%)
Training	6	100	0	0
Correct	15	100	0	0
Missing one step	10	20	80[a]	0
Missing six steps	10	0	50[b]	50

[a] All eight claim missing that step; two of eight claim missing an extra step; one claims missing two extra steps.
[b] Three claim missing five nodes; two claim missing six; all five claim at least three actual missing steps.

Table 3. Labeling Individual Nodes

Individual Node	Overall Success[a]	Correct Positive[b]	Correct Negative[c]
B: Turn on	0.9999	1.0000	0.9999
C: Read ID screen	0.9901	0.9956	0.9897
D: Read ID strip	0.9893	0.9333	0.9909
E: Press "C"	0.9787	0.2344*	0.9998
F: Match	0.9847	0.9267	0.9908
G: Shake it	0.9590	0.6003	0.9738
H: Unscrew cap	0.9563	0.5041	0.9857
I: Drop it	0.9827	0.8584	0.9941
J: Insert	0.9878	0.8643	0.9961
K: Wait	0.9964	0.9987	0.9958
L: Read result	0.9966	0.9847	0.9991
M: Mark control	0.9983	0.9720	0.9993
N: Turn off	0.9967	0.8997	0.9997
O: Screw cap	0.9476	0.6629	0.9617

[a] Overall success is the average of all nodes computed as the correct positive plus the correct negative divided by all frames.
[b] Correct positive is the number of correctly labeled positive frames divided by the number of all positive frames for node *I*.
[c] Correct negative is the number of correctly labeled negative frames divided by the number of all negative frames for node *I*.
* The score is anomalously low because this event is comparatively very short; the other events were much longer and could be more positively matched to the ground truth.

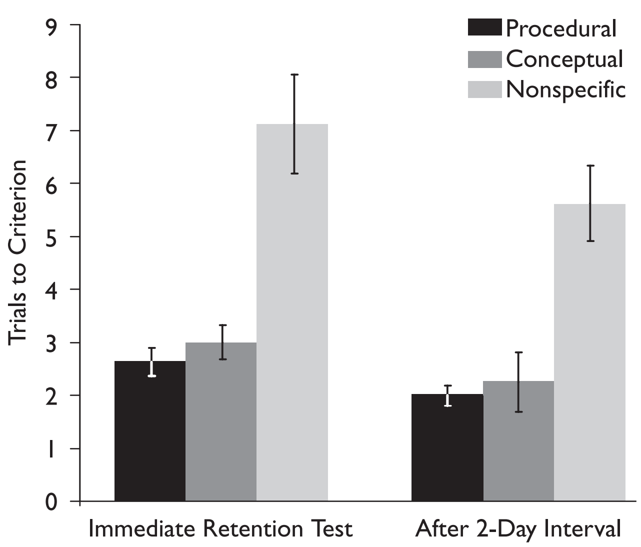


Figure 5. Data from the feedback study (means with standard error bars).

The data in Figure 5 illustrate that performance was worst for the nonspecific feedback condition. Simply being told an error had been made did not provide much support for learning: It took participants nearly seven trials to perform the glucose control calibration test correctly. Moreover, after a 2-day retention interval, they required almost six more trials to perform the same procedure correctly. The two feedback conditions with content – action or concept – did not differ from each other and supported immediate performance as well as retention after two days. Participants in those conditions required fewer trials after the retention interval, which suggests they had successfully learned the procedure.

For the calibration task we assessed, feedback with content supported performance regardless of whether participants were told what action to perform or the concept of what they should do. For other tasks, however, the nature of the information provided during training does differentially affect performance (e.g., Mead & Fisk, 1998). Therefore, we need to assess the relative benefits of different feedback content for a broader range of tasks before drawing conclusions. Nevertheless, it was clear that older adults could benefit from feedback, both immediately and after a retention interval, but that the feedback must be specific to be effective.

Conclusion

As noted earlier, self-care in the domestic environment is an important aspect of health care. Although there is a benefit of empowering individuals to be involved in their own health care and maintenance, it is also a challenge because inappropriate use can be harmful. We focused on a medical device used in a home environment that was representative of the cognitive requirements of many devices currently prescribed for home use. The significance of this particular device is far from trivial because of the fact that many older adults are diagnosed with late-stage diabetes. Moreover, this research should lead to the development of principles and guidelines that would be applicable to other technology coaches.

Future efforts must address the engineering challenges that remain, as well as psychological issues related to the design of effective feedback systems. With respect to the computer vision issues, we must continue to improve the accuracy of the system and integrate multimodal sensing sources. It will be crucial to extend this effort to other systems that range in complexity and sequential form. We also recognize the importance of activity discovery; that is, the ability to automate the task decomposition and to be able to learn a model of "normal" activity that can be used as the basis to detect misuse.

With respect to the coaching component, many feedback issues remain: content, timing, amount, and optimal display format. In addition, we must understand whether and how feedback effectiveness differs as a function of user capabilities and experience.

The specification of the components in our computational vision system, based on the glucose meter, can be used to inform the development of technology aids for other medical devices (e.g., blood pressure monitor, infusion pumps), and other activities (e.g., preparing a meal, performing specific exercises prescribed as part of a rehabilitation program). Technology coaches have the potential to support a variety of activities for a range of users.

References

Aggarwal, J. K., & Cai, Q. (1999). Human motion analysis: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(11), 1241–1247.

Colagiuri, R., Colagiuri, S., Jones, S., & Moses, R.G. (1990). The quality of self-monitoring of blood glucose. *Diabetic Medicine*, *7*, 800–804.

Dahlbäck, N., Jonssön, A., & Ahrenberg, L. (1993). Wizard of Oz studies: Why and how. In *Proceedings of the 1st International Conference on Intelligent User Interfaces* (pp. 193–200). New York: ACM Press.

Fiesler, C. L., McLaughlin, A. C., Fisk, A. D., & Rogers, W. A. (2003). Conceptual versus procedural feedback in the training of a home medical device. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 1810–1814). Santa Monica, CA: Human Factors and Ergonomics Society.

Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2004). *Designing for older adults: Principles and creative human factors approaches*. Boca Raton, FL: CRC Press.

Hancock, H. E., Fisk, A. D., & Rogers, W. A. (2001). Everyday products: Easy to use . . . or not? *Ergonomics in Design*, *9*(4), 12–18.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284.

McLaughlin, A. C., Rogers, W. A., & Fisk, A. D. (2006). How effective feedback for training depends on controlled learner resources and task demands. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 2624–2628). Santa Monica, CA: Human Factors and Ergonomics Society.

McLaughlin, A. C., Rogers, W. A., Fisk, A. D., & Essa, I. (2005). Designing a technology coach for older adults. In *Proceedings of Eleventh International Conference on Human-Computer Interaction.* Mahwah, NJ: Erlbaum.

Mead, S. E., & Fisk, A. D. (1998). Measuring skill acquisition and retention with an ATM simulator: The need for age-specific training. *Human Factors*, *40*, 516–523.

Moore, D., Essa, I., & Hayes, M. (1999). Exploiting human actions and object context for recognition tasks. In *Proceedings of Seventh International Conference on Computer Vision - ICCV 99* (pp. 80–86). Los Alamitos, CA: IEEE Computer Society.

Mykityshyn, A. L., Fisk, A. D., & Rogers, W. A. (2002). Learning to use a home medical device: Mediating age-related differences with training. *Human Factors*, *44*, 354–364.

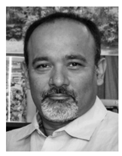
Rogers, W. A., Mykityshyn, A. L., Campbell, R. H., & Fisk, A. D. (2001). Analysis of a "simple" medical device. *Ergonomics in Design*, *9*(1), 6–14.

Shi, Y., Huang, Y., Minnen, D., Bobick, A., & Essa, I. (2004). Propagation networks for recognition of partially ordered sequential action. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition – CVPR 2004* (pp. 862–869). Los Alamitos, CA: IEEE Computer Society.

Vaswani, N., Roy-Chowdhury, A., & Chellappa, R. (2003). Activity recognition using the dynamics of the configuration of interacting objects. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition – CVPR 2003* (pp. 633–640). Los Alamitos, CA: IEEE Computer Society.

Wilson, R. S., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., & Bennett, D. A. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging*, *17*, 179–193.

Wendy A. Rogers is professor of psychology at Georgia Institute of Technology. She received her B.A. from the University of Massachusetts – Dartmouth and her M.S. and Ph.D. from the Georgia Institute of Technology. Her research interests include skill acquisition, human factors, training, and cognitive aging. She is codirector of the Human Factors and Aging Laboratory. She may be reached at School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332-0170, wendy@gatech.edu.



Irfan A. Essa is an associate professor of computing in the School of Interactive Computing in the College of Computing at Georgia Institute of Technology. He received his B.S. in engineering from IIT and his S.M. and Ph.D. from Massachusetts Institute of Technology. His research interests include computer vision, computer graphics, and machine learning, in which he specifically focuses on building intelligent machines and environments. He codirects the Computational Perception Laboratory and is a member of the GVU Center and Robotics and Intelligent Machines Centers.

Arthur (Dan) Fisk is professor and coordinator of the Engineering Psychology Program in the School of Psychology at Georgia Institute of Technology. In 2003, he received the Human Factors and Ergonomics Society's Paul M. Fitts Education Award, and in 2004, he was awarded the Georgia Institute of Technology Doctoral Thesis Advisor Award. His first novel, the tell-all of his hiking adventures, will be published as soon as someone pays him to write it.

Research reported in this article was supported in part by the following grants: Grant P01 AG17211 from the National Institutes of Health (National Institute on Aging) under the auspices of the Center for Research and Education on Aging and Technology Enhancement (CREATE); Award 0121661 "The Aware Home: Sustaining the Quality of Life for an Aging Population" from the National Science Foundation; and the Aware Home Research Initiative industrial partners. We thank Aaron Bobick, Casey Fiesler, Yan Huang, Anne McLaughlin, David Minnen, and Yifan Shi for their contributions to and help with this work. Portions of this research were presented at the Human Factors and Ergonomics Society 47th Annual Meeting (Fiesler et al., 2003) and the HCI International Conference (McLaughlin et al., 2005). EID