eyeCOOK: A Gaze and Speech Enabled Attentive Cookbook

Jeffrey S. Shell, Jeremy S. Bradbury, Craig B. Knowles, Connor Dickie, Roel Vertegaal
Human Media Lab, Queen's University
Kingston, ON, Canada, K7L 3N6
{ shell, bradbury, knowles, connor, roel }@cs.queensu.ca

ABSTRACT
While preparing food, cooks often have to manage many time sensitive processes. Because cookbooks require visual and physical attention to use, they may distract, rather than focus the cook on executing the recipe. The knowledge requirements of cooking, concurrent demands for attention and the sensitivity of recipes to proper procedure conspire to make cooking a stressful experience, particularly to novices. We present eyeCOOK, a multimodal attentive cookbook which allows users to communicate using eye-gaze and speech. eyeCOOK responds visually and/or verbally, promoting communication through *natural* human input channels without physically encumbering the user. Our goal is to improve productivity and user satisfaction without creating additional requirements for user attention. We describe how the user interacts with the eyeCOOK prototype and the role of this system in an Attentive Kitchen.

Keywords
Attentive User Interfaces, Gaze, Eye Tracking, Speech, Context-aware, Information Appliance, Sensors.

INTRODUCTION
The attentive cookbook reduces the burden caused by competing requests for physical attention by directing instructions and glossary definitions to the unoccupied auditory channel. Gaze and speech were chosen as input modalities because they do not require physical contact (i.e., mouse, touch screen), which may be inconvenient and unsanitary while preparing food. The dynamic display window automates page turning, allowing users to focus their hands on the process of cooking. Alternative approaches to enhancing the cooking experience include Cook's Collage [7], a passive memory recovery aid, and CounterActive [1], a touch enabled multimedia cookbook. Unlike these approaches, eyeCOOK adapts its behavior and interface presentation based on the user's eye gaze, proximity, and the current cooking task. eyeCOOK qualifies as both an *information appliance* [2], and a *context aware* system [4], but because user attention drives the human interface, it is most precisely described as an *Attentive User Interface* (AUI) [5]. Using knowledge of the user's attentive context, the system applies contextual reasoning to activate localized grammars, working within the user's *attention space* [5].

SYSTEM DESCRIPTION
Our attentive cookbook prototype consists of an electronic recipe database with a hypertext-style interface. eyeCOOK receives input from an LC Technologies eye tracker and a wireless microphone using the Microsoft Speech API (SAPI) for speech recognition and production. The eyetracker is calibrated only once for each user. Since eyeCOOK uses a small context-sensitive grammar, speech recognition calibration is rarely required.

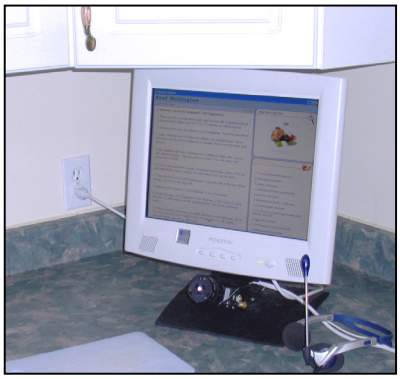

Figure 1. The eyeCOOK System

eyeCOOK can read aloud ingredients and instructions, access additional information such as pictures and definitions of ingredients, terminology, cookware, nutritional information, the history of the dish, and suggest other food items it can be served with. Throughout the process of cooking, the system automatically triggers timed notifications, to remind the user when steps are completed, or require attention. The relationship between the current, previous and future cooking steps and the ingredients they involve is implicit via a simple color-coding scheme. The dynamic colouring, shown in figure 2, situates the current cooking task within the recipe as a whole.

Adaptive Inputs and Display
The display adjusts its presentation based on user presence reasoning. If eye gaze is detected, then the user must be in front of the screen. Thus the entire recipe is shown on one page in Cookbook mode (see figure 2), and deictic references are interpreted using gaze and speech. Alternately, if eye gaze is not present, the display is set to Recipe Card mode. This breaks up the recipe into multiple cards and enlarges the text, thereby allowing a user who is further away from the display to easily read the recipe. Because the user is not providing eye gaze input in Recipe Card mode, the system compensates by adjusting the vocabulary of the speech recognition engine, providing answers to more detailed queries.

Figure 2. eyeCOOK in Page Display Mode

NATURAL INPUTS

eyeCOOK is designed to use *natural* input modalities, or those that humans use in human to human, non mediated communication [5]. Observing and interpreting this implicit behavior reduces the need for users to provide explicit input. Using these cues, interfaces can be designed such that the difficulty lies in the intended task, not the technological tool.

Gaze and Speech

When the user is in range of the eye tracker and looking at the display, eyeCOOK substitutes the target of the user's gaze for the word '*this*' in a speech command. For example, eyeCOOK responds to the spoken command *'Define this'* by defining the word the user is currently looking at. However, because eye trackers are spatially fixed and have a limited range, the user will not always be in a position where eye tracker input is available. Thus, our speech grammar is designed such that system functionality is maintained when users are not in front of the eye tracker. Instead of saying "define *this*" while looking at the word *sauté*, the user simply states "define *sauté*." The active vocabulary is dynamically generated using context-sensitive, localized speech grammars, allowing more synonyms to be included for a given word. Real world performance may be improved by adding partial terms and colloquialisms that may only be relevant in specific circumstances.

TOWARDS AN ATTENTIVE KITCHEN

Interfaces that recognize and respond to user attention, and understand how it relates to the overall activity can help the user efficiently engage in tasks. To achieve this, we must augment the kitchen with attentive sensors that monitor human behavior [5,6], augment appliances with functional sensors [3,6], improve coordination among appliances [5], and allow appliances to affect the environment [3,5].

Attentive and Environmental Sensors

Increasing the knowledge of users' activities may allow interfaces to engage in less interruptive, and more respectful interactions with users. Visual attention, a prime indicator of human interest, can be deduced by adding eye contact sensors [5,6] to items in the environment. This information can be used to determine the appropriate volume and timing of notifications to the user. Additionally, temperature sensors can be used to keep track of the status of the oven and the elements of the stove and could be synchronized with electronic timers to increase the system's ability to guide the user's cooking experience

Appliance Coordination

Integrating knowledge of the environment can result in improved functionality, taking up less of the user's time and effort. For example, user recipe preferences, timing constraints, as determined by the user's electronic schedule, and currently available ingredients, communicated by food storage areas, can be combined to suggest recipes. Once selected, the ingredients from the recipe can be added to an electronic shopping list stored on the user's PDA.

Active Environmental Actions

The kitchen should not only be aware of its environment, but it should also be able to affect it. Thus, it should be able to take actions which increase efficiency, and reduce the user's action load, like automatically preheating an oven.

CONCLUSIONS

We have presented eyeCOOK, a gaze and speech enabled multimodal Attentive User Interface. We have also presented our vision of an Attentive Kitchen in which appliances, informed by sensors, coordinate their behavior, and have the capability to affect the environment. This can reduce the user's workload, and permit rationalizing requests for user attention.

REFERENCES

1. Ju, W. et al. (2001). CounterActive: An Interactive Cookbook for the Kitchen Counter. *Extended Abstracts of CHI 2001* (Seattle, April 2001) pp. 269-270
2. Norman, D. A. *The Invisible Computer*, MIT press, 1999
3. Schmidt, A. et al. How to Build Smart Appliances. *IEEE Personal Communications* 8(4), August 2001. pp. 66-71.
4. Selker, T., et al. (2000). Context-Aware Design and Interaction in Computer Systems. *IBM Systems Journal* (39) 3&4, pp. 880-891.
5. Shell, J. et al. Interacting with Groups of Computers. *Commun. ACM* 46(3) March, 2003.
6. Shell, J. et al. EyePliances: Attention Seeking Devices that Respond to Visual Attention. *Extended Abstracts of CHI 2003* (Ft.Lauderdale, April 2003) pp. 770-771
7. Tran, Q., et al. (2002). Cook's Collage: Two Exploratory Designs. Position paper for Families Workshop at CHI 2002. Minneapolis, MN.