Hands On Cooking: Towards an Attentive Kitchen

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

ABSTRACT
To make human computer interaction more transparent, different modes of communication need to be explored. We present eyeCOOK, a multimodal attentive cookbook to help a non-expert computer user cook a meal. The user communicates using eye-gaze and speech commands, and 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
For a person not skilled in the culinary arts, the use of a cookbook is often a necessity. A cookbook, however, requires visual and physical attention, which distracts from the process of cooking. Additionally, to use a cookbook, users have to understand cooking terminology and may also have to manage many time sensitive processes. Due to many conflicting demands for user attention, preparing a meal is often a stressful experience.

The attentive cookbook directs traditional cookbook instructions and glossary definitions of ingredients and processes to the unoccupied auditory channel. Other approaches to automating and enhancing the cooking experience include Cook's Collage [6], 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 cooking context. eyeCOOK qualifies as both an *information appliance* [2], and a *context aware* system [4], but because user attention drives the human interface scenario, 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, to work within the user's *attention space* [5].

CHI 2003, April 5-10, 2003, Ft. Lauderdale, Florida, USA.
ACM 1-58113-637-4/03/0004

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. These were chosen instead of devices that require physical contact (i.e., mouse, touch screen), which may be inconvenient and unsanitary while preparing food.

Once a recipe has been selected from the database, the system shows the recipe in one of two attention sensitive display modes. The user can have the system read aloud ingredients and instructions, access additional information such as pictures and definitions of ingredients, cooking terminology, and cooking instruments, nutritional information, the history of the dish, and suggestions of what other food items it could be served with. Once the user begins the actual process of cooking, the system automatically sets timers and reminders. The relationship between current, previous and future cooking steps and the ingredients they involve is implicit via a simple color-coding scheme. This dynamic colouring provides a cue situating the user's current cooking task within the recipe as a whole.

Attentive Features
Features that help manage the user's attention include:

- **Automatic Timer** – system will automatically set timers based on context (i.e., cooking mode + current recipe step). Timers will alert the user when a task is complete or requires attention.
- **Adaptive Display** – Optionally, the user can allow the system to choose the appropriate display mode based on presence reasoning. If eye gaze is detected, the Cookbook Page mode is activated (see Figure 1), where the complete information for one recipe is displayed on one screen. If eye gaze is not detected then the user is not in front of the screen, so the display is set to Recipe Card mode. This increases the size of the text, breaking the recipe description into multiple screens, helping a user who is further away from the display read the recipe.
- **Adaptive Input Channels** – If the user is within the range of the eye tracker, voice commands are used in conjunction with eye gaze. Otherwise, more detailed voice commands are activated and eye gaze input is dormant until a user is once again in range.

Figure 1. eyeCOOK in Page Display Mode

NATURAL INPUTS

eyeCOOK is designed specifically to use *natural* input modalities: those that humans use in human to human, non mediated communication [5]. To reduce the need for users to provide explicit input, or change their behavior to accommodate interface constraints, implicitly provided attentional cues are observed and interpreted. We believe that this approach improves the learnability and intuitiveness of interfaces designed for novice users.

Voice Commands

eyeCOOK uses context-sensitive, localized grammars. This allows more synonyms for a given speech recognition command, reducing the chance of misinterpreting a word.

Eye Gaze Commands

When the user is in range of the eye tracker, eyeCOOK substitutes the object of the user's gaze for the word '*this*' in a speech command. For example, '*Define this*' will trigger the define operation on the current eye gaze target. Since current eye trackers are spatially fixed and offer limited mobility to users, the user will not always be in a location where eye tracker input is available. Our speech grammar is designed so that system functionality is not reduced when using speech without eye tracking. This is accomplished by the user stating "define *boil*", instead of saying "define *this*" while looking at the word *boil*.

TOWARDS AN ATTENTIVE KITCHEN

Interfaces that recognize and respond to user attention, and understand how it relates to the overall task within the kitchen environment can help the user efficiently engage in her/his task. To achieve this, we must increase sensing capability [3], improve coordination among appliances [5], and give appliances the ability to affect the environment [3,5].

Environmental Sensors

Temperature sensors used to keep track of the status of the oven and the elements of the stove can increase the system's ability to guide the user's cooking experience and could be synchronized with electronic timers.

Appliance Information Integration

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. As well, selecting a recipe can result in the addition of necessary ingredients to an electronic shopping list stored on the user's Personal Data Assistant (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. Tran, Q., et al. (2002). Cook's Collage: Two Exploratory Designs. Position paper for Families Workshop at CHI 2002. Minneapolis, MN.