Kitchen Helper: Using A Gestural Interface for Procedural Learning – A Pilot Study

Joseph Binney and Brian Louie
Stanford University HCI Group
Stanford, CA 94305
{jbinney, blouie}@stanford.edu

ABSTRACT

In this experiment, we compare the effectiveness of gestural interfaces versus that of non-gestural ones for teaching users order-based procedural tasks. By effectiveness, we wanted to measure the time spent using the gestural interface, its ease of use, and its ability to ease memory retention. We implemented a gestural interface with the Kinect to teach people how to cook an omelet and found that, while it seems to slow down the completion of the task, it seems promising for teaching in a fashion that encourages ease of use and repeatability.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General terms: Design, Human Factors, Experimentation

Keywords: Kinect, gestural interface, motion detection, video, Kitchen Helper, cooking, instructions, learning

INTRODUCTION

Many tasks do not always lend themselves well to traditional instructions. For example, cooking from a recipe often demands multitasking and the processing of multiple cognitive threads. This, while in the domain of human mental capacity, tends to be inefficient and can produce quite a few errors. Our research tries to find ways to make cooking and similar tasks easier and faster. Ultimately, we developed an interface that provides video and audio instructions to teach users how to cook, in addition to implementing navigation via a gestural Kinect-enabled interface.

The Kinect is an especially fascinating device because it makes gestural and motion detection available in a cheap and fairly mainstream way. As the technology continues to get cheaper and more widely used, we can expect applications beyond its native gaming one to become more prominent. Thus our research also represents an opportunity to preliminarily explore the more practical uses of the Kinect and see where gesture and motion detections can fit in the scope of human-computer interaction and, more importantly, into our daily lives.

RELATED WORK

Our research was much inspired by *CounterActive*, an interactive cooking surface that demonstrates the effective use of video instructions that can be played and replayed during the cooking process [4]. One of the interesting takeaways from that project was that the interface was simple enough for a seven-year-old to prepare a dish. In developing our interface, we strived to achieve a similar goal with regard to ease of use.

At a more abstract level, our research was also inspired by *Portico*, a tablet system that can detect the presence of physical objects and then translate them to virtual output on the tablet [2]. The efficacy of that system lies in the interactivity of the gestural interface; the novelty and the ability to substitute user interface metaphors with physical objects makes certain tasks (e.g. playing tic-tac-toe) more interesting and intuitive.

Ultimately, an interface like ours might fit well in the bigger vision as explained by Bell and Kaye, who test a fully virtual interface that requires more cameras and motion sensors, which makes their idea more difficult to implement at a grander scale [3]. However, their interface is advantageous in that it can track the physical location of items on the counter, which would be an interesting next step for our research project.

KITCHEN HELPER

"Kitchen Helper" is an application we developed that combines audio and video information with the Kinect-enabled gestural interface to teach people how to cook a western omelet [1]. In the following sections, we will discuss the different features that we added to our interface.

Gestural Interaction

One of the biggest innovations in our research is the use of gestures to navigate around the application. We developed five gestures to use in the application; adding more would have made the gestures more difficult to remember, while using fewer would have severely limited the desired functionality of our application. For each gesture, we thought carefully about 1) most effectively limiting false positives and 2) creating gestures that intuitively suggest their functions. Before the experiment began, we provide a crash course for the gestures such that the user cannot advance past a step in the instructions until he successfully perform the gesture he is trying to learn. We also decided to add a "ding" sound, which played after a gesture was successfully performed. We felt that this was important so that the subject – who presumably has never used this interface before – would know when the Kinect successfully read his movements.

The first and second gestures were ways to move to the next and previous steps in the recipe. These gestures involved holding the right or left arm (respectively) horizontally to the side for one second, and then dropping them down to the user's side. The third gesture was to go back to the beginning of the current step, which entailed holding the user's arm out in front of him and then swinging it in two circles, in either a clockwise or counterclockwise fashion. The fourth gesture was to toggle the play/pause state of the video, which entailed forming an "L" with either arm out to the side of the user's body. And finally, the fifth gesture was used for a timer (in steps where a timer was required). To activate the timer, the user simply tapped his head.

Audio and Video Information

As noted in the "Related Work" section of this paper, previous studies have demonstrated the utility of video when teaching people how to cook. We also wanted to try adding an audio version of the instructions in each step so that the user could divert their visual attention to the actual cooking activity if such was needed. Thus, the video for each step was supplemented by a male voice paraphrasing the text instruction of the recipe in slow, clearly enunciated speech.

Graphical User Interface

To supplement the video, the monitor also showed a graphical user interface (GUI) to surround the video. We placed a progress bar and a text version of the instructions for the current step at the top of the screen. At the bottom of the screen, we added brightly colored buttons containing stick figures demonstrating how to use each gesture. Though we had a tutorial at the beginning of the program teaching users how to use the gestures, we wanted to ensure that we covered flukes from the tutorial and possible failure to remember the gestures. The lower-right corner also featured a timer for recipe steps that required timing. The timer's font started out white, and in the final ten seconds it turns red. At the end, a loud beeping sound occurs to alert the subject that the timer has finished.

IMPLEMENTATION

Our program was implemented in two distinct modules. In the first module, we used two open-source SDKs – OpenNI and Nite, both of which used C++ – to develop the Kinect-side code. (Presumably, the power of our application might improve with the recent release of Microsoft's official Kinect SDK, though such remains to be verified.)

In the second module, we used Java Swing to put together the GUI so that the user could be presented with an aesthetically pleasing product. We used VLC Player to show the video and audio content. Because VLC can be operated using remote control via a terminal, we were able to develop Java commands that navigate through VLC Player.

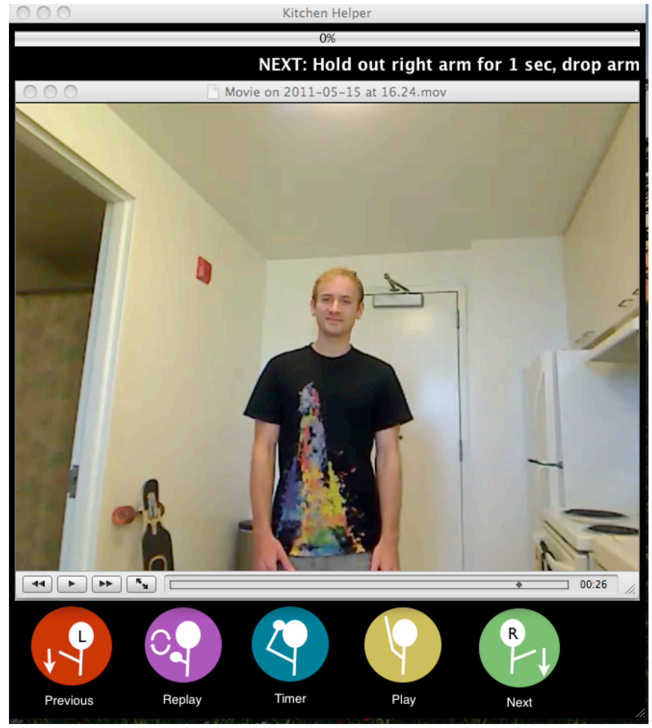

Figure 1: An Java Swing GUI that incorporates VLC to show video instructions. It also includes text instructions and a progress bar that line the top of the page. At the bottom, there are small stick figure icons that depict the gestures used to navigate around the program.

To connect the two modules, we opened a socket that lets the modules communicate. When a gesture was read using the first module, a unique signal was sent to the second module to tell it to manipulate the GUI and VLC interface in some specified manner.

METHODOLOGY

We wanted to compare the traditional recipe experience to our new Kitchen Helper application, so we divided subjects into two groups: one that followed a paper recipe, and one that used Kitchen Helper. For consistency, we used the same textual instructions for both condition groups. The recipe we selected was a 5-star Food Network western omelet recipe. We chose this recipe for its low ingredient cost and short cooking time.

In order to control for differences in cooking experience, we gave subjects a pre-survey asking how often they cook food, and evenly divided subjects into conditions based on experience. We also collected demographic information in the survey, and asked whether they had cooked an omelet before.

Before beginning the experiment, we placed all the necessary ingredients and utensils on the kitchen counter. We also labeled cooking temperatures on the stove knobs, and marked the area of the kitchen where Kitchen Helper could be used. Using a script, we gave instructions to all participants and answered questions before beginning the cooking experience. Then, they independently cooked an omelet while we timed the cooking speed.

After finishing cooking, we took pictures of each omelet, saved a log of gestures used, and administered a second survey. The survey used a 7-point Likert scale to evaluate how easy the cooking experience was, and how well participants thought they could repeat the recipe. We also asked what might make the experience easier, and asked Kitchen Helper users if they would recommend using the application over using a text recipe. After collecting data from all participants, we blindly scored each omelet photo on a scale of 1-5 (1 = worst, 5 = best) on three aspects: cooking amount (not overcooked or undercooked), shape (folded and solid), and filling quality (finely chopped and sufficiently sautéed).

Figure 2: One of the omelets that was cooked by a subject who used the Kinect to learn how to make it. In general, the quality of omelets was the same across both conditions.

RESULTS

We found no significant difference in omelet quality between the two groups. On average, using Kitchen Helper (K) took 5 minutes longer than the paper recipe (P), with average times K = 21m37s, P = 16m35s (p-value = 0.003, n = 10). However, despite taking relatively longer, 100% of Kitchen Helper users said they would recommend using the application over a paper recipe.

The most common complaint among paper recipe subjects regarded clarifying the recipe instructions, particularly "stir [eggs] constantly until just-cooked." None of the subjects who used Kitchen Helper complained about unclear recipe instructions, even though we used the same terminology and descriptions in the textual and audio instructions.

Our results suggest that Kitchen Helper may have provided an easier cooking experience (1 = easy, 7 = hard): K = 2.3, P = 3 (p-value = 0.242, n = 10). The results also demonstrate that Kitchen Helper might have helped subjects better repeat cooking the recipe (1 = poorly, 7 = well): K = 5.9, P = 5.5 (p-value = 0.497, n = 10). For subjects who had the least cooking experience, the margin of difference in repeatability was even greater: K = 5.83, P = 4.86 (p-value = 0.179, n = 6).

FUTURE WORK

In the future, we would like to look at the official Kinect SDK announced by Microsoft to see if we could make the gestural detection even more precise and powerful. We would also like to try using a different recipe, perhaps one that fewer people already knew how to make so that we could get a better grasp on how well our interface can actually teach people how to cook. And finally, we would simply like to test more people. As noted earlier in the paper, our p-values kept creeping lower as we tested more subjects; due to time considerations we were unable to do so, but in the future, more subjects could make our preliminary findings more significant.

We would also like to incorporate feedback from our subjects into the next iteration of the application. For example, subjects in the Kinect condition suggested that we more explicitly highlight the presence of the timer, or at the very least alert them before the process starts that they have a timer available to use during the cooking process.

Furthermore, we would like to see if we could find ways to make the interface faster to use, considering that time of completion was the only statistically significant difference between our interface and a standard paper one. As of now, we are not sure where we would be able to take shortcuts in our particular implementation, and thus we wonder whether the extra time is inherent to the type of interface we decided to build. Further research in this particular aspect of the interface might be useful, as the speed with which a task can be completed is always a big factor in evaluating effectiveness of an interface.

DISCUSSION

Our gesture recognition system was imperfect, occasionally registering false positives or failing to recognize intended gestures. Even so, all Kitchen Helper users liked the application and said they would recommend using it over a paper recipe. Users particularly appreciated the video instructions as they helped to conceptually clarify the steps of the recipe. Several users commented on the appropriateness of a gesture recognition system, since they could use the application with dirty hands.

We noticed that subjects who used the paper recipe had a greater tendency to perform instructions out of order, or ignore steps of the recipe. One paper recipe subject began cooking eggs two steps early, and ended up overcooking the omelet. Several paper recipe subjects also ignored instructions to constantly stir the eggs while cooking, in fear of scrambling the eggs. Kitchen Helper emphasized the importance of order in the cooking process, and demonstrated execution of each step. We believe this lead to greater consistency in the resulting omelets.

When we examined the self-reported repeatability for our subjects with the least cooking experience, we found that the average value for Kitchen Helper users was consistent with all users in this condition (5.83 versus 5.9), but inexperienced paper recipe users found it less repeatable (4.86 versus 5.5). This result suggests that Kitchen Helper might provide the greatest benefit for people with little cooking experience. We also extrapolate repeatability ratings to be a measure of how effective Kitchen Helper was in teaching a new recipe.

While some of our results were not statistically significant, we noticed that p-values decreased as we continued to test more subjects. We believe that these results would be significant if we continued evaluating Kitchen Helper on a larger subject pool.

CONCLUSION

This paper presents Kitchen Helper, a tool that uses audio, video, and gestures to teach people how to cook. In this particular instance, it teaches people to cook western omelets, though there is little reason to believe that it cannot be used to teach cooking of other food items. We hope that this starts a discussion about the use of gestures in broader applications of HCI.

Ultimately, we found that while our interface isn't perfect, our research begins to dig at the potential of widespread gesture detection using the Kinect in more practical applications. In future iterations of our research, we hope to see solutions that will circumvent the problems that we have, something that might be more possible with fewer time and resource constraints.

ACKNOWLEDGMENTS

We would like to thank Scott Klemmer and Jesse Cirimele for teaching CS376 at Stanford, giving us the framework to help us complete this project. We would also like to thanks Tico Ballagas from Nokia Research for helping us explore the design space for Kitchen Helper and helping us pinpoint the features we wanted to include in our application.

REFERENCES

1. "Western Omelette". http://www.foodnetwork.com/recipes/food-network-kitchens/western-omelette-recipe/index.html
2. Avrahami, Daniel. Portico. http://vimeo.com/21825173
3. Bell, Genevieve and Kaye, Joseph. Designing Technology for Domestic Spaces: A Kitchen Manifesto. In *Gastronomica: The Journal of Food and Culture*, Vol. 2, No. 2 (Spring 2002, Austin, TX), pp. 46-62.
4. Ju, Wendy, Hurwitz, Rebecca, Judd, Tilke, and Lee, Bonny. CounterActive: An Interactive Cookbook for the Kitchen Counter.