Using Eye Contact and Contextual Speech Recognition for Hands-Free Surgical Charting

G. Julian Lepinski
Human Media Lab
Queen's University
Kingston, Canada
lepinski@cs.queensu.ca

Roel Vertegaal
Human Media Lab
Queen's University
Kingston, Canada
roel@cs.queensu.ca

***Abstract*— In this paper we discuss ongoing research into applications for multimodal Attentive User Interfaces in hands-free charting during surgical procedures. Although speech recognition has matured enough that it can now be used for some software and hardware control, speech recognition solutions still have trouble filtering "command speech" from "ambient speech." Our research builds on previous research that couples eye contact sensing with speech recognition to gauge intent. Users enable a voice activation system used for surgical time charting by fixing their gaze on a small camera before speaking command words.**

Keywords—Attentive UI, charting, gaze, voice recognition

I. INTRODUCTION

While a surgical team focuses on their immediate surgical tasks, there are a number of auxiliary tasks which also need to be performed. One frequent secondary task is the charting done by surgical nurses. A new healthcare initiative in hospitals in Ontario, Canada requires the charting of numerous surgical time benchmarks [3]. Charting this information is valuable; however, nurses may have trouble recording exact times while occupied with other tasks.

This problem was identified to us by a surgical nurse and was the starting point for our research. We have worked to develop software which allows the surgical team to record the times that specific surgical benchmarks were reached without a significant interruption in other ongoing tasks. In the current iteration of this project, the team will transfer these time records later to the permanent chart.

Traditional voice recognition software works well in environments where the user is able to manually enable or disable voice recognition, or keep silent when not commanding the system. This poses problems in an operating theatre because of the need for constant communication between members of the surgical team. We have addressed this problem by using techniques discussed in [1] to couple gaze information with voice recognition.

II. EXISTING WORK

Voice activation systems have been used in health-care applications for some time now [4]; however, these systems have generally relied on microphones placed physically on users, the use of activation switches, and the employment of extremely sophisticated voice recognition software to provide reliable operation. We have found this to be unsuitable for a system which must be inexpensive and used unobtrusively by any member of a surgical team.

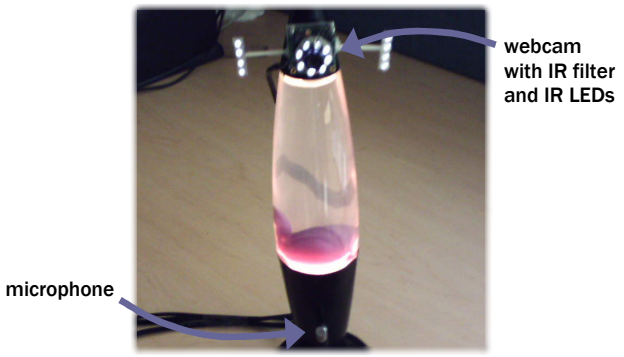

Figure 1. The AuraLamp detects gaze to give context to speech

Previous research has shown that contextual information such as gaze – when coupled with another input mode, such as voice recognition – allows software to better gauge user intent [5]. This has remedied one of the most significant problems with voice activated systems: such systems are often unable to distinguish between command speech and speech unintended for the system. Shown in Fig. 1, and discussed in [1], the AuraLamp uses a webcam equipped with an infrared (IR) filter, on- and off-axis IR LEDs and a small microphone to receive voice commands when it detects user gaze.

III. OUR WORK

A. *Hardware*

Our research work was initially conducted using face detection as a temporary stand-in for true gaze detection. We have used inexpensive webcam and microphone hardware.

For true eye contact sensing, the hardware must be capable of exploiting the bright pupil effect in users [1]. This requires webcams modified with IR LEDs and IR filters. In the upcoming phase of this project we will deploy these cameras along with updated software to recognize true gaze in real-time.

B. *Software*

Developed as a Windows application, our software runs on desktop computers, which are frequently present in modern

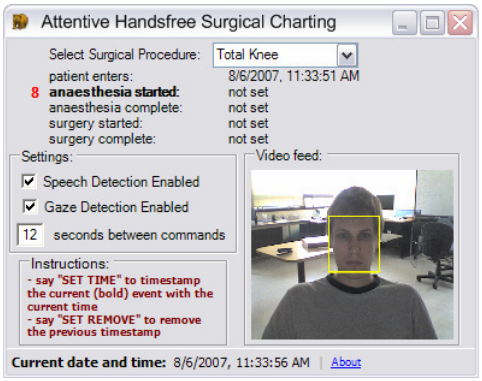

Figure 2. The attentive surgical charting software

operating rooms. Users are able to see the camera's video feed and tracking output, as well as all previous and upcoming time-stamped surgical benchmarks (shown in Fig. 2).

We have used the open source OpenCV library (*http://sourceforge.net/projects/opencvlibrary*) for image capture and processing and we have also made use of the OpenCVDotNet wrapper to interface with OpenCV (*http://code.google.com/p/opencvdotnet*). Our software employs gaze information as an indicator of user intent at all times. False positives do occur, and we have included a threshold requirement for uninterrupted gaze which is user specified.

Once this threshold is met, the speech recognition engine is enabled until a command is detected or until the user stops fixing their gaze on the camera, as indicated in Fig. 3. We use the MS Speech SDK 5.1 (*http://microsoft.com/speech*) for voice recognition.

Our voice commands are: "set time," which stamps the active benchmark with the current time and sets the next benchmark to active; and "set remove," which removes the most recently time-stamped benchmark. This allows users to remove erroneous timestamps.

IV. Initial Results

We have conducted preliminary tests of our software in a noisy and crowded environment in an attempt to simulate the conditions of an operating theatre.

Initial results have been promising, with the software providing sufficient accuracy to complete basic test scenarios. Charting was not accomplished effortlessly, however a drastic improvement was observed over using the voice recognition system without gaze as an additional input mode.

Users reported the system to be easy to use, including users with little or no familiarity with voice recognition or gaze

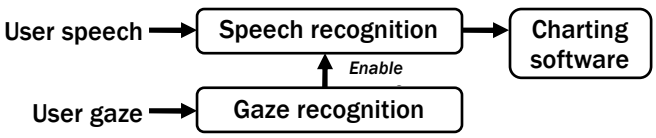

Figure 3. The design of the system

based systems prior to using our software.

As anticipated, face detection is not a sufficient substitute for gaze. The system was often triggered by erroneous face detection when a user was not fixating on the camera with their gaze. Gaze detection in our next iteration should remedy this.

In situations where we enforced uninterruptible tasks, we observed that the accuracy of user time reporting increased when using our software versus no aid, as users were forced to rely on their recall or on rough time estimates otherwise.

Further experimental results, including video recordings of the system in operation and the test environment may be viewed at *http://hml.queensu.ca/surgerytracker/*.

V. Future Work

Our lab is currently working to set up a real-world pilot of our software in hospitals around Ontario, Canada. We aim to collaborate with surgical teams to determine how our system can benefit users in real-world situations, and what aspects of the system require development or modification to improve usefulness and functionality.

We have also identified a problem with using the OpenCV Haar filter for face detection in our implementation, as real-world participants will be wearing surgical masks, and will not be well recognized. We are in the process of training a new Haar filter to recognize masked faces so it is better able to recognize a surgical team.

We see numerous possibilities for context-aware speech detection in medical settings. Providing speech recognition software with context awareness allows it to surmount many of the problems which have kept it out of more widespread use. We see the need for further research into more complete voice activated charting software and hardware, which we hope will help free up nurses and other health-care professionals to focus on more pressing health care tasks.

Acknowledgment

The authors would like to gratefully acknowledge Nancy Lepinski at Markham Stouffville Hospital for proposing the original idea for this research.

References

[1] A. Mamuji, R. Vertegaal, J. Shell, T. Pham, and C. Sohn, "AuraLamp: contextual speech recognition in an eye contact sensing light appliance," In Extended Abstracts of Ubicomp, 2003.

[2] C. H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Real-time detection of eyes and faces," In Proc. ICMI (1998).

[3] J. Trypuc, A. Hudson, and H. MacLeod, "The pivotal role of critical care and surgical efficiencies in supporting Ontario's Wait Time Strategy: Part 3." Healthcare Quarterly, 9(4) pp. 37-45, 2006.

[4] L. Rossi, D. Sacerdoti, B. Billi, G. Lesnoni, M. Orciuolo, T. Rossi, D. Sacerdoti, and L. Bertollini, "Automatic speech recognition in vitreo-retinal surgery. A project for a prototypal computer-based voice-controlled vitrectomy machine," European journal of ophthalmology vol. 6, no. 4, pp. 454-459, 1996.

[5] T. Maglio, C. S. Matlock, S. Campbell, S. Zhai, and B. A. Smith, "Gaze and speech in attentive user interfaces," Proceedings of the Third International Conference on Advances in Multimodal Interfaces, pp.1-7, 2000.