Improving Speech User Interface Performance in the Project54 System

Jennifer Carter, Andrew L. Kun, W. Thomas Miller, III
University of New Hampshire, Department of Electrical and Computer Engineering
{carter, andrew.kun, tom.miller}@unh.edu

ABSTRACT
In previous research we evaluated the Project54 system's speech user interface and found that it responded correctly 85% of the time. This paper presents initial implementations of two approaches to improving the SUI performance as well as one proposed approach that is currently being implemented.

Keywords
Project54, human/user interface

INTRODUCTION
The Project54 system integrates the police cruiser's standard equipment such as light bars and radar into a system where all equipment can be controlled through a GUI or a speech user interface (SUI) [1]. The purpose of the SUI is to allow the officer to control the in-car electronics while keeping his/her hands on the wheel and eyes on the road. The system runs under Windows on a PC which may be placed in a center console (see Figure 1).

Figure 1 Project54 System in a Cruiser

The Project54 system uses the Microsoft SAPI5 speech recognizer. Grammar files are used to identify the possible commands an officer will utter. This improves recognition accuracy. In order for an officer to issue a command to the speech user interface he/she must press the push-to-talk (PTT) button to indicate the beginning and the end of an utterance. This eliminates the need for the speech recognizer (SR) to continuously process incoming sounds and in turn also improves recognition accuracy.

LEAVE BLANK THE LAST 2.5 cm (1") OF THE LEFT COLUMN ON THE FIRST PAGE FOR THE COPYRIGHT NOTICE.

Evaluating the SUI
Between June 2003 and August 2004 we collected 49,177 utterances from twenty-seven police officers who were using the Project54 system in the field [2]. All officers were male. The utterances were saved as individual sound files and the recognition results corresponding to the utterances were saved. The average recognition rate of the SUI, over the entire corpus, was 85.34%, shown in Figure 2. Imperfect SUI recognition can be due to the SR engine's perfect performance and user error. In our corpus, 5.43% of the utterances were not recognized due to a speech recognition error and 9.23% were not recognized due to human error.

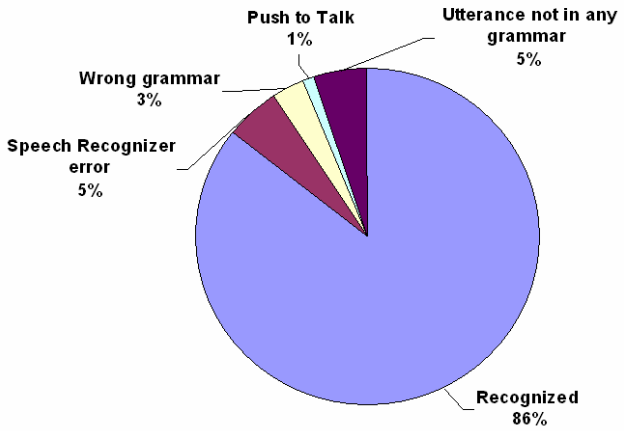

Figure 2 Overall Recognition Rate

We blamed recognition errors on the speech recognizer in all cases when the officer's utterance was clearly distinguishable by a human listener. However, this meant blaming the speech recognizer when, in addition to the officer's utterance, it also heard the voice output from the police radio or when traffic noise overshadowed the officer's utterance.

We encountered three types of user errors (Figure 2). Officers uttered commands that are not in any of the SUI's grammars (5% of the corpus). They also uttered commands that were not in the grammar that was loaded at the time of the utterance (3.1%). Finally, they cut off the utterance (at the beginning, at the end, or both) by pushing the PTT button too late and/or releasing it too early (1.13%).

IMPROVING SUI PERFORMANCE

In this paper we present our efforts to improve SUI performance by reducing system errors and user PTT errors. We decided to reduce system errors using two approaches. The first one was to use our corpus of speech samples to train the speech recognizer. The second one (currently being implemented) is to eliminate errors due to interfering voices from the in-car police radio. Our approach to reducing PTT errors is to send the recognizer sound input starting a little before the PTT is pressed.

Training the Speech Recognizer

System errors are due to the speech recognizer not recognizing or incorrectly recognizing clear valid enunciated utterances. As Project54 uses a general speech recognizer which does not require training, any voice can use the system. However the circumstances under which the officers use the speech user interface are not typical for a speech recognizer. There are many common and frequent background noises due to the fact that the system is used within a police car. The utterances in the collected corpus were used to train the speech recognizer. We will evaluate the results of this training using a new speech corpus that we have just started collecting.

Eliminating Interference from the Police Radio

In order for the speech recognizer to properly process officer's utterances, the utterances should not be contaminated with other sounds. If the microphone picks up multiple voices and sends them to be processed, the speech recognizer may be confused. The Project54 setup uses a directional microphone to filter out sounds not coming from where the driver's head should be (see Figure 1), but it cannot filter out all such sounds. To make matters worse in some cruisers the radio speaker is installed behind the driver's seat at the level of the headrest. In these cruisers, when the police officer's radio is active while the officer is talking to the Project54 system, both the officer's utterance and the voices on the radio will find their way into the speech recognizer. The police radio cannot be shut off or turned down as it is necessary for the officer's job.

One possible solution to this problem is to determine when the police radio is speaking and ignore any utterances that are issued while the radio is speaking. By monitoring the activity of the police radio the Project54 system can determine when the police radio is busy with voice traffic [3]. Therefore whenever the Project54 system detects that the police radio is speaking it will ignore any utterances sent to the speech recognizer. We are currently testing this solution in a laboratory environment.

Usage of the PTT button

In order for an officer to issue a command to the speech user interface he/she must press the PTT button to indicate he/she is going to issue an utterance so that the speech recognizer starts listening and can process the command. When the PTT button is pressed it sends a message to the speech recognition to start listening, and when the button is released it sends a message to the speech recognizer to stop listening. A common user error is the improper use of the PTT button. The user will either press the button after starting to talk or release it too soon while still speaking, so the speech recognizer only hears part of the utterance. In this case the speech recognizer is not likely to correctly interpret the utterance.

One possible solution to this problem is to implement a circular buffer for recording utterances. Our system constantly records sound input in short clips and stores the clips in buffers. These buffers are then sent for recognition when the PTT button is pressed. In our current system the user can select the option of sending to the recognizer the sound input that was recorded several tens of milliseconds before the PTT button was pressed (the length of time is adjustable). Therefore, if an officer often presses the PTT button too late he/she can enable this option and the recognizer will receive voice input starting before the PTT button was pressed.

CONCLUSION AND FUTURE RESEARCH

We implemented two approaches for improving the performance of the Project54 SUI and are currently experimenting with a third approach. All three approaches were designed based on our evaluation of the SUI based on a large corpus of utterances collected in the field.

We are currently evaluating our approaches. We are also refining the use of circular buffers. We will classify the content of the buffers before the PTT button is pressed as speech or non-speech, and only send the content to the recognizer if it is classified as speech. We will expand this approach and classify all input to the recognizer as clean input that is likely to be recognized correctly or as input with excessive noise (e.g. traffic noise) which is unlikely to be recognized correctly. Responses of the recognizer to input that is classified as noisy may then be ignored.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Justice under grants 2004CKWX0121 and 2005CKWX0426.

REFERENCES

1. Andrew L. Kun, W. Thomas Miller, III and William H. Lenharth, "Computers in police cruisers," IEEE Pervasive Computing, Volume 3, Issue 4, Pages: 34 - 41, October-December 2004
2. Laslo Turner, Andrew L. Kun, "Evaluating the Project54 speech user interface," Third International Conference on Pervasive Computing (Adjunct Proceedings), Munich, Germany, May 8-12, 2005
3. Nathan Purmort, "Measuring New Hampshire State Police radio usage," Technical report ECE.P54.2005.4, Electrical and Computer Engineering Department, University of New Hampshire, 2005