Paper Number IVSS-2004-UAS-01

Field Testing the Project54 Speech User Interface

Laslo Turner
Andrew L. Kun
Electrical and Computer Engineering Department, University of New Hampshire

ABSTRACT

An in-car speech user interface (SUI) is being tested in the field. Preliminary results, based on a corpus of about 19000 utterances, show that the SUI recognition rate is around 86%. Approximately one third of the SUI's errors are due to speech recognizer errors while about two thirds are due to user error. We have identified three types of user errors: issuing an invalid command (54%), issuing a command that is valid in some context but not in the current context (35%) and operating the push-to-talk button incorrectly (11%).

INTRODUCTION

Technological advances have introduced many after-market electronic devices in police cruisers. Today's cruisers are equipped with lights, sirens, digital radios, scanners, radars, mobile computers, GPS units, barcode readers, video recorders, etc. Most of these devices are installed in cruisers as stand-alone devices, meaning that each of them has its own user interface. We have created a system that integrates the devices into a single system which allows officers to interact with the individual devices using a speech user interface (SUI). Our system is currently installed in about 100 police cruisers in everyday use by various police departments.

We are currently conducting a study with the aim to quantify the performance of our SUI. To this end we are collecting speech interaction data in the field.

BACKGROUND

In-car speech user interfaces have been developed by a number of groups. Hunt (1999) describes work on in-car speech recognition at Dragon Systems UK R&D. Kun et al (2003) describe a system that integrates multiple in-car devices and lets the user interact with them using a SUI. Using a SUI to interact with in-car devices promises to make such interactions safe, however poor SUI design can result in unsafe interactions which in turn can lead to accidents (Green 2000).

Work on in-car SUIs has also dealt with the development of speech recognition (SR) engines for the in-car environment; see for example (Cristoforetti et al 2000). Developing such SR engines requires in-car speech corpora. While multiple in-car speech corpora exist (Heeman et al 2000) we are not aware of any corpora recorded in police cruisers during everyday operation.

THE SPEECH USER INTERFACE

The SUI uses a commercial SR engine (we have worked with several recognizers and are currently using the Microsoft SAPI 5 recognizer). It also uses a commercial text-to-speech (TTS) engine.

In our system human-computer interaction is relatively simple: the officer utters a phrase and the SUI reacts to this utterance. The SUI may execute a command, fill in a data field or initiate data retrieval. The SUI does not initiate interaction.

SR performance is improved by using a push-to-talk (PTT) button. The officer signals to the SR engine that an utterance should be recognized by pressing the PTT button. Recognition is started when the PTT button is pressed and it is stopped when the button is released. SR performance is also improved through the introduction of grammars. Grammars are switched by the SUI in response to officer utterances. For example, the officer may be filling in a form in which one of the fields describes the gender of a person. The officer may say

'Gender' to signal that he/she will tell the SUI the gender of the person next. At this point the SUI loads a simple grammar which only lists the two genders and one or two other utterances (such as "cancel").

QUESTIONS ASKED

In testing our SUI we asked four questions, discussed in this section.

What is the recognition rate of the system? From our prior experiences with the system we expected that the recognition rate will be around 90%.

What are the reasons for the imperfect recognition rate? We expected that there would be two groups of reasons for recognition to be imperfect: the SR engine's imperfect performance and human error.

Do officers use the SUI? We expected to find that the majority of the officers who volunteered to take part in our data collection effort actually use the system. However, we also wanted to see how the amount of SUI interactions changes over time.

What activities is the SUI most used for? The SUI allows the officer to interact with every device in the cruiser. However, we expected to find that the SUI will not be utilized to interact with all in-car devices.

TESTING PROCEDURE

Our in-car system allows recording the utterances that are fed to the SR engine. When recording is turned on a wav file is created every time the officer presses the PTT button. The wav file contains the sound picked up by the system's microphone. Wav files are named using the current date and time. The system also creates a file with all the recognition result. This allows us to pair up wav files and recognition results. All relevant files are retrieved when an officer brings in the vehicle for upgrades or repairs.

Officers are asked to volunteer to participate in our data collection effort. Participating officers use the system in their everyday work.

RESULTS AND DISCUSSION

RECOGNITION RATE

As of January 2004 we have just over 19000 utterance records on file. They were collected from ten officers. Table 1 shows the SUI recognition rate for each of the ten officers as well as the average over the ten officers.

As Table 1 shows, the average recognition rate of the system is 86.18%. For one of the officers it dips to 73.97% and for two officers it is over 90%. The average recognition rate is thus lower than we expected it to be.

We also quantified the mean number of commands that the user successfully issues before a command is not recognized by the SUI. For this purpose we introduced the MCBE, or mean-commands-between-errors variable, which we defined as the total number of voice commands that are recognized divided by the total number of commands that are not recognized:

$$MCBE = \frac{commands\_recognized}{commands\_not\_recognized}.$$

user	records	recognized	not recognized
1	4817	4259 (88.42%)	558 (11.58%)
2	1139	918 (80.60%)	221 (19.40%)
3	1220	1083 (88.77%)	137 (11.23%)
4	1866	1658 (88.85%)	208 (11.15%)
5	2820	2420 (85.82%)	400 (14.18%)
6	2351	2167 (92.17%)	184 (7.83%)
7	464	430 (92.67%)	34 (7.33%)
8	1610	1320 (81.99%)	290 (18.01%)
9	2496	2029 (81.29%)	467 (18.71%)
10	799	591 (73.97%)	208 (26.03%)
total	19582	16875 (86.18%)	2707 (13.82%)

Table 1 SUI recognition rate

From Table 2 we see that the MCBE ranges from just under 4 to a little over 20 for the ten officers.

user	recognized	not recognized	MCBE
1	4526	291	15.55
2	1022	117	8.74
3	1128	92	12.26
4	1742	124	14.04
5	2537	283	8.96
6	2229	122	18.27
7	442	22	20.09
8	1427	183	7.8
9	2172	324	6.7
10	638	161	3.96

Table 2 Mean commands between errors

REASONS FOR IMPERFECT RECOGNITION

By listening to all the records in our corpus we found that some utterances were clearly distinguishable by a human listener and that the command given was in the grammar that was loaded at the time of the utterance. This means that the SR engine did not recognize the utterance even though the officer did what he/she was

expected to do. In Table 3 we show the accuracy of the SR engine by taking into account only the records that were clearly distinguishable to a human listener and that were uttered with the correct grammar loaded. We can see that on average the SR engine performed very well with a 94.47% accuracy. In Table 3 we also show that the utterances that were not recognized were either misrecognized as another command or unrecognized. Misrecognized commands lead to the execution of the wrong command which usually has to be undone. Unrecognized commands are simply rejected by the SUI.

user	records	recog-nized	misrecog-nized	unrecog-nized
1	4526	4259 (94.10%)	169 (3.73%)	98 (2.17%)
2	1022	918 (89.82%)	79 (7.73%)	25 (2.45%)
3	1128	1083 (96.01%)	22 (1.95%)	23 (2.04%)
4	1742	1658 (95.18%)	46 (2.64%)	38 (2.18%)
5	2537	2420 (95.39%)	63 (2.48%)	54 (2.13%)
6	2229	2167 (97.22%)	49 (2.20%)	13 (0.58%)
7	442	430 (97.29%)	5 (1.13%)	7 (1.58%)
8	1427	1320 (92.50%)	53 (3.71%)	54 (3.78%)
9	2172	2029 (93.42%)	118 (5.43%)	25 (1.15%)
10	638	591 (92.63%)	21 (3.29%)	26 (4.08%)
total	17863	16875 (94.47%)	625 (3.50%)	363 (2.03%)

Table 3 SR engine accuracy

As shown in Table 4, we also found that the remaining utterances that were not recognized belonged to one of the following three groups:

- the utterance was cut off (at the beginning or at the end or both) because the user pushed the PTT button too late and/or released it too early;
- the utterance was clearly distinguishable by a human listener however, the command uttered was not in any of the SUI's grammars;
- the utterance was clearly distinguishable by a human listener however, the command uttered was not in the grammar that was loaded at the time of the utterance.

Table 1, Table 3 and Table 4 show that out of the 2707 utterances in our corpus that were not recognized by the SUI, 625+363=988, or 36.5%, were not recognized due to the SR engine making a mistake and 196+922+601=1719, or 63.5%, were not recognized due to human error. The most common human error was issuing a command that was not in any of the SUI's grammars – fully 922, or 34%, of utterances that were not recognized were not in any of the grammars. This result tells us that officer training needs to improve and that we can expect dramatic improvements in SUI performance if training is indeed improved.

user	records	push-to-talk	not in any grammar	not in loaded grammar
1	291	48 (16.50%)	160 (54.98%)	83 (28.52%)
2	117	13 (11.11%)	68 (58.12%)	36 (30.77%)
3	92	6 (6.52%)	56 (60.87%)	30 (32.61%)
4	124	42 (33.87%)	37 (29.84%)	45 (36.29%)
5	283	34 (12.02%)	156 (55.12%)	93 (32.86%)
6	122	12 (9.84%)	58 (47.54%)	52 (42.62%)
7	22	8 (36.36%)	12 (54.55%)	2 (9.09%)
8	183	14 (7.65%)	98 (53.55%)	71 (38.80%)
9	324	16 (4.94%)	154 (47.53%)	154 (47.53%)
10	161	3 (1.86%)	123 (76.40%)	35 (21.74%)
total	1719	196 (11.40%)	922 (53.64%)	601 (34.96%)

Table 4 User errors

SUI USAGE

Table 5 shows how each of the ten officers used the SUI over the period of time they participated in the data collection effort. On average the officers issued from just under six commands an hour to about 18 commands an hour. From these numbers we conclude that the officers do use the SUI and that some of them use it very frequently. We also noticed that new users tend to use the system more than users who have had the system for a while. Figure 1 shows a representative pattern of usage over a period of about three months. At first the user issued around 200 commands per day. Later the usage was reduced significantly – on some days the user barely issued one or two commands, while on others around 50 commands were issued. A question we have

not answered yet is why this change takes place. It is likely that there are two reasons for this. One of them is probably that users are at first excited about the system and spend considerable time learning how to use it, but with the passage of time they experiment less. The second reason is probably that as time passes the officers decide not use the SUI for certain activities for which they find it to be of limited value.

It is worth noting that the user with the lowest number of utterances per hour was user 10, the same user who had the lowest SUI recognition rate (just over 73%). Presumably this user got discouraged and did not use the system much.

user	records	days	hours	per day	per hour
1	4526	61	418	74.20	10.83
2	1022	17	55	60.12	18.58
3	1128	16	91	70.5	12.4
4	1742	55	178	31.67	9.79
5	2537	56	249	45.3	10.19
6	2229	24	127	92.88	17.55
7	442	17	57	26	7.75
8	1427	25	103	57.08	13.85
9	2172	45	245	48.27	8.87
10	638	37	107	17.24	5.96

Table 5 SUI usage

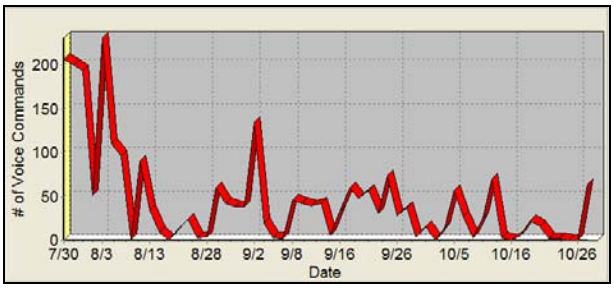

Figure 1 Change in SUI usage over time

WHEN IS THE SUI USED?

We determined how the officers used the SUI by looking at which grammar was loaded by the SR engine at the time of an utterance.

Table 6 shows the three most common grammars used. We can see that the Patrol Screen grammar was used most often. The Patrol Screen is an application that controls multiple electronic devices: the radar, the radio and the lights and sirens. The next two common grammars are the Registration Inquiry grammar and the License Inquiry grammar. These grammars are used to perform registration and driver's license inquiries.

We should note that since the Patrol Screen application is used to control multiple devices, we will need to evaluate what activities are actually performed when this grammar is loaded.

rank	grammar	usage [%]
1	Patrol Screen	40.11
2	Registration Inquiry	10.94
3	License Inquiry	8.95

Table 6 Grammars used

CONCLUSION AND FUTURE WORK

Investigating our small corpus we were able to find preliminary answers to the four questions posed in the "Questions Asked" section. We found that the recognition rate of the system is 86%. We also found that recognition is imperfect primarily because of user errors. Approximately one third of the system's errors are due to speech recognizer errors, while about two thirds are due to user error. We have identified three types of user errors: issuing an invalid command (54%), issuing a command that is valid in some grammars but not in the current grammar (35%) and operating the push-to-talk button incorrectly (11%). These results will be used to improve officer training. If user errors are reduced by only 50% we will improve the system recognition rate to over 90%.

We also found that officers who volunteered for the data collection effort do indeed use the SUI although there are significant differences between individual users. We also found that the most common activity performed using the SUI is to interact with the Patrol Screen application, which controls three in-car devices (radar, radio and lights and siren).

Our data collection effort is now picking up speed. Our system is currently installed in about 100 cars and about half of the officers using these cars have volunteered to participate in our study. We expect to increase the size of our corpus significantly over the course of the next year. We will use the corpus to get a more accurate view of how our SUI is used by officers. The corpus will also be made available to the research community. Our data collection work will be augmented by asking officers to participate in written surveys (the first one is planned for March 2004). We will also spend time with officers on the road and observe how they perform their everyday work with the help of our SUI. Over the course of the next several months we plan to spend four hours (half a shift) with 10-12 officers each.

REFERENCES

Cristoforetti, Luca et al. 2003. In-car speech interaction by means of multiple recognition units. Proceedings of the Workshop on DSP in Mobile and Vehicular Systems.

Heeman, Peter, et al. 2001. The US SpeechDat-Car data collection. Proceedings of the 7th Eurospeech.

Hunt, Melvyn J. 1999. Some experience in in-car speech recognition. Proceedings of the IEE Colloquium on Interactive Spoken Dialogue Systems for Telephony Applications.

Green, Paul A. 2000. Crashes induced by driver information systems and what can be done to reduce them. Proceedings of Convergence 2000.

Kun, Andrew L., et al. 2003. Project54: Standardizing electronic device integration in police cruisers. IEEE Intelligent Systems, 18(5): 10- 13.

CONTACTS

Laslo Turner
Graduate Research Assistant
University of New Hampshire
Electrical and Computer Engineering Department
Kingsbury Hall
Durham, NH 03824
Phone: 603-862-4326
Email: lturner@unh.edu

Andrew L. Kun
Assistant Professor
University of New Hampshire
Electrical and Computer Engineering Department
Kingsbury Hall
Durham, NH 03824
Phone: 603-862-4175
Fax: 603-862-1832
Email: andrew.kun@unh.edu