The U.S. SpeechDat-Car Data Collection

Peter A. Heeman, David Cole and Andrew Cronk

Center for Spoken Language Understanding
Oregon Graduate Institute of Science and Technology
20000 NW Walker Rd., Beaverton OR 97006
heeman@cse.ogi.edu

Abstract

The SpeechDat-Car data collection effort is an ambitious effort to collect data from multiple languages in an in-car setting. This paper describes the U.S. data collection effort. We discuss problems we had implementing the collection procedure; and changes we made to improve the procedure. This paper should benefit future in-car data collections.

1. Introduction

The U.S. SpeechDat-Car project is part of a European initiative to produce high-quality in-car corpora in a number of different languages [7, 11, 4, 10]. The corpora are designed to promote voice-based interfaces to existing in-car accessories, such as radios and telephones, as well as future accessories, such as navigation systems; and designed to capture the inherent noise in the car environment, by using seven different noise conditions, from parked car to highway driving with the radio on. The effort started in April 1998 with nine European Union languages: Danish, British English, Finnish, Flemish/Dutch, French, German, Greek, Italian, and Spanish. The partners behind the effort comprise car manufacturers (BMW, FIAT, Renault, SEAT-Volkswagen), companies active in mobile telephone communications and voice-operated services (Bosch, Alcatel, Knowledge, Lernout & Hauspie, Nokia, Sonofon, Tawido, Vocalis), and universities (CPK, Denmark; DMI, Finland; IPSK, Germany; IRST, Italy; SPEX, Netherlands; UPC, Spain; WCL, Greece).

The car is set up with a computer mounted in the trunk, and a flat panel display that the driver can see. The driver is prompted to say sentences, phrases, words, letters, or numbers; to answer a question or to make up a sentence. The interaction is controlled by the experimenter, who sits in the passenger seat. The experimenter starts and stops the recording. The driver is recorded through five microphones. One is a high-quality head-mounted close-talking microphone and the four others are far-talk microphones, mounted across the roof in front of the driver. Three of these and the close-talk microphone lead to a filter board in the computer, which in term leads to an analog-to-digital board. The car software communicates with the analog-to-digital board, and records the four channel signal on the hard drive. The fourth far-talk microphone is hooked up to a cell phone, which in term is recorded by a remote computer through the telephone network. The in-car computer also controls the cell phone: it dials the fixed server, and sends DTMF tones to tell the remote system the session and prompt numbers, and when to start and stop recording.

The SpeechDat-Car data collection is a very difficult one. The U.S. SpeechDat-Car project is the first corpus after the initial nine. Although we benefited from some of the experiences of the other participants, we had a number of difficulties. This paper discusses some of the implementation problems that we faced, and some improvements that we made, and software tools that we created to support or ease the data collection effort. We hope that our experiences will benefit other collection efforts for in-car data collections.

2. Corpus Collection work at OGI

The Center for Spoken Language Understanding at the Oregon Graduate Institute has been involved in corpus development for a number of years. Our corpus development includes a telephone corpus [1], a speaker recognition corpus with subject recordings over a two year time span [2], a corpus with recordings in twenty two different languages [6], and a corpus of speech from children in grades kindergarten through grade twelve [9].[1] These corpora are available for free to academic institutions and CSLU center members. CSLU also does custom corpus work on a contract basis. The U.S. SpeechDat-Car data collection is a custom corpus that we are doing through a contract from the European Language Resource and Distribution Agency. CSLU can use the corpus for internal research purposes but does not have distribution rights.

3. Problems Setting Up The Car

Even though our data collection effort followed the nine European ones, we still encountered a large number of difficulties.

3.1. Equipment no longer available

Some of the equipment, such as the recommended computer system, and the recommended monitor, were no longer available when we started the project. For instance, finding a flat panel monitor that could work in daylight and nighttime conditions was difficult, and led us to an innovative approach using two monitors (see Section 4.1). Also, the in-car computer had to run Windows NT and have two ISA slots for a filter board and the AD board. After spending a significant time searching for industrial computers for in-trunk mounting, we decided to build our own. We used an ordinary computer mounted in a cage used for rack mounted computers.

3.2. Lack of documentation

A lot of details in the setup were not documented. For instance, the fact that the filter board had to reprogrammed after each

[1] The first author was also involved in a corpus collection of human-human task-oriented dialogues [5], distributed through the Linguistics Data Consortium.

booting of the system was only discovered after we had sent in an audio file to be verified. Also, the exact setup of the in-car computer was not specified. In our setup, we connected the power inverter to the engine, and a Universal Power Supply (UPS) to the power inverted (to clean up the power), and hooked the computer to the UPS. The first power inverter that we bought was unable to give the UPS with a good power supply; this was solved by buying a more heavy duty power inverter, capable of outputting 800 watts. It was also unclear how to mount the computer in the trunk. A local store that specializes in installing stereos and other audio-visual equipment in cars suggested mounting it in a plywood box (with no front nor back), with 1 inch thick foam between the computer and plywood. Our setup is shown in Figure 1, where the power inverter is on the far left, the UPS is next, and the computer is on the right. Since our collection was done during the winter in Portland, the outdoor temperature ranged between 32 and 60 degrees Fahrenheit, and hence no special cooling nor heating was necessary. We did not experience any computer crashes with this installation.

Figure 1: In-Trunk Mounting

3.3. Incompatibilities with U.S. phone network

The in-car software works with the Nokia Data Suite to control the cellular phone. Unfortunately, the Nokia Data Suite in conjunction with the U.S. phones, did not support all of the required AT commands used by the in-car software.[2] After waiting three months for the new U.S. version 3.0 of the Nokia Data Suite, we found that even that version did not support the needed AT commands. We were finally able to get the source code for the in-car software, and had to rewrite it not to use the unsupported AT commands.

Not only is the cellular phone network different between the U.S. and Europe, so is the land-line digital network. The fixed platform software that the Europeans had used, which received the calls from the cell phone and recorded them, was designed to work with digital telephone cards for the European phone network. We did not even attempt to use this software; rather, we used the CSLU Speech toolkit [3, 8] to write a script to control the fixed platform. The toolkit provides a high level interface to using a Dialogic board for handling a T1 line. The script interprets the DTMF tones sent by the in-car recording software that indicate the session, the prompt, and when to stop recording.

[2]Our contract specified for us to use the GSM phone network and the Nokia 5190 phone to increase compatibility with the European corpora. This phone network operates at a different frequency than the European GSM network.

4. Equipment Improvements

We made two changes to the car installation: we used two monitors instead of one, and we built a box to better monitor and control the batteries on the microphone pre-amplifiers.

4.1. Dual Monitors

In the United States, there is an increasing negative attitude towards the use of cell phones. Some cities have in fact banned the use of non hands-free cell phones by drivers. There are also laws against drivers being to view television monitors. Due to the safety issues, we wanted to minimize the impact of having the driver view a computer monitor.

The SpeechDat project recommended the use of an eight inch flat panel monitor. The recommended monitor, however, was no longer being made at the time of our project, nor did the manufacturer have a replacement. Unfortunately, not any flat panel would do. Ideally, it should be capable of display during full sunlight and at night time, with minimal strain on the eyes. Hence, it should have as large of a brightness range as possible. We were only able to find a single complete flat panel that was suitable. It had a ten inch diagonal and made for use in police cars. Positioning such a large monitor in the car would be be difficult.

Instead, we opted for two 6” monitors: one for the driver and one for the experimenter. Our setup is shown in Figure 2, where the driver’s display is shown to the upper right of the steering wheel and the passenger’s is shown to the lower right of the steering wheel.[3] Each monitor could thus be mounted specifically for its recipient, rather than having to settle on a compromise position. The driver’s was mounted to the right of the dashboard instruments, and seemed to be an extension of them. Looking at the monitor was no more difficult than looking at the speedometer. The second monitor was mounted to that it could be easily seen by the passenger/experimenter, without causing any strain. An additional advantage of the two monitor setup was that we could revert to using a single one if we encountered problems with one of them.

The monitors, however, did not come fully assembled; rather they had to be assembled from a VGA input board, the flat panel, a control strip, and a potentiometer (to easily control brightness). We also had to custom build a case for them, which we built in clear acrylic. The flat panel was from Planar, designed for the transportation industry, with brightness ranging from 2 to 1000 nits.[4] Although this setup in the end was more expensive and more complicated than using the ready built 10” solution, the added pain was worth having better situated monitors.

4.2. Battery Monitor

One of the biggest problems was that the four channels recorded in the car (one closetalk and 3 far talk microphones) each

[3]Also visible is our cordless keyboard with builtin mouse pad from NetSurf. The mouse pad on the keyboard did not work adequately.

[4]Luckily, Planar is located next door to OGI. In fact, we were able to get one of their engineers to assemble create two complete monitors for us. Hopefully, future projects will have complete solutions available to them.

Figure 2: Two Monitor Setup

needed a pre-amp with their own nine-volt battery.[5] Each pre-amp needed to be turned on before each session and then off after it (the pre-amp can ware the battery done overnight). Although we had a voltmeter in the car, checking the batteries meant removing the pre-amp from its mounting and removing the battery cover. Hopefully, during this process, none of the wires had become loose, nor the pre-amp switched off. This procedure also caused wear and tear on the pre-amps, which are prone to breakdown.

The battery situation proved problematic, and a number of sessions suffered from decreased audio range or from the pre-amps not being switched on. One third of the way through, we built a box, shown in Figure 3, that has a mount on it for each battery, a single on-off switch that controls all four batteries, and a voltmeter with a four-way switch to select which battery to test. This box has a 8 foot cord so it can be in convenient reach of the experimenter, who can now even test the batteries during a session. This device drastically simplified testing and changing the batteries, and we feel that it resulted in improved audio range in the files.

5. In-Car Software

We built a number of software tools to aid the experimenter in collecting and managing the data.

[5]Each pre-amp needs its own power source separate from the others, otherwise you will get crosstalk.

Figure 3: Battery Setup for Pre-amps

5.1. Checking the Quality of Recordings

Most of the partners used software for controlling the collection procedure that was developed by Mantra/Nortel. The software had a test mode in which you could record and then play back the recorded speech for each of the four channels. During actual recording, it had monitors that were supposed to show audio activity on each of the four channels. These monitors were not adequate for determining if the channel was being properly recorded. Other software was supplied by SPEX for determining the range and signal to noise, but this was not built into the recording software.

Almost all problems with the recordings were due to whether the pre-amps being turned on, whether they were working properly, and whether the batteries were charged. To keep the recording quality as high as possible, we needed tools that the experimenter could run on the in-car computer to check the recording equipment.

We built a program in Perl that would run the signal to noise software on a set of audio files, build a web page of the results, including anchors to the demultiplexed audio files, and open up the page with a web browser. We had two versions of this tool. The first version runs the tool on the test audio file collected on the test screen of the recording software. This version of the tool allows the user to easily check the recording quality before starting a session, The user can see the signal to noise values, the recording range, the amount of clipping, and can listen to the audio files.[6] This tool is run off of the Windows ‘start’ button.

The second version of the tool runs on the audio files from a recording session. This tool can even be run during a recording session. Depending on the number of prompts recorded in the session, the tool would either run the signal analysis and demultiplexer on all of them, or would randomly pick a small sample of them.[7]

5.2. Speaker and Session Distribution

The corpus collection required certain distributions for road conditions, speaker age, speaker gender, and speaker accent. We wrote a Perl program that could be run in the car that would

[6]We also installed the CSLU speech toolkit, which meant that we could also look at spectrograms of the audio files.

[7]Running the signal to noise and demultiplexing on all of the audio files in the session would require too much time

report the current distribution. For each accent region, it reported the number of male and female subjects by age range, and it reported the number of road conditions. This allowed the experimenters to better understand the progress that was being made.

5.3. Transferring Data

Rather than using backup tapes to transfer data off of the car, we had an Ethernet adaptor board in the in-car computer, and simply plugged the car into our local network using a long cable.[8]. Uploading the audio files for 10 speakers took about one hour. We wrote a Perl program that would query the user as to which sessions to upload, gather up all of the audio files using 'tar', move this file onto a network mounted drive, and create a synchronization file, so that the program will no not to ask whether to upload the session again in the future.

Our in-car computer ran Windows NT, and although it had a twelve gigabyte hard drive, each partition was only four gigabytes. A partition could only hold about 18 speakers worth of data. We wanted to keep a copy of the sessions in the car for as long as possible, in case there was a problem with the upload (we did go back to these several times). To maximize how long audio files stayed on the in-car computer, we wrote a Perl script that would move the files from the partition that we recorded to, to one of the other partitions. When the other partitions were full, it would erase the session that had been upload most distantly, as determined by the date that the synchronization file was created.

6. Software for Managing the Corpus

We ended up building a number of other Perl programs for managing the corpus collection. These scripts had to allow sessions to be incrementally added to the corpus, as we needed to start transcribing and determining the quality of the recorded sessions as soon as possible. The scripts served the following functions.

- Download the collected sessions from the fixed platform into a tar file.
- Install the audio and log files from the fixed and in-car platform from the tar files onto our 70 gigabyte hard-drive.
- Check whether all files that we expected to be downloaded were found.
- Run the signal to noise software on the recently downloaded sessions, and save the results for creating the final deliverables, and for displaying the session quality.
- Create wave files and first approximation at the word transcription (based on what the speaker was prompted to say) for transcription with the transcriber tool in the CSLU speech toolkit. We modified the transcriber tool so that it could find simple transcription mistakes according to the SpeechDat-Car transcription standard.
- Look for more complicated errors (such as spelling mistakes) in the transcribed words.
- Create a list of wave files to aurally check for artificial noises.
- Create an HTML page that shows the quality of the sessions, in terms of signal to noise, audio range, clipping, and the results of the aural test.
- Create the final deliverable CD-ROM images, in which the transcriptions are merged in to the prompt files, and the sessions are divided into 650 megabytes partitions.

[8] This only worked because we did our collection locally

7. Conclusion

The SpeechDat car data collection was a very challenging data collection. Our worst fears when we started the collection was that of whether subjects could safely do the data collection while driving. Besides being pulled over by the police a half dozen times (no ticket was ever issued), the biggest safety problem was when another car drove into us when we were doing a parked-car session. In the end, our biggest problem turned out to be in getting the in-car recording software to work on with the U.S. phones. If we had the source code for this software to begin with, we could have shaved off many months from the project.

8. References

[1] R. A. Cole, M. Noel, T. Lander, and T. Durham. New telephone speech corpora at CSLU. In *Proceedings of the European Conference on Speech Communication and Technology*, 1995.

[2] R. Cole, M. Noel, and V. Noel. The cslu speaker recognition corpus. In *Proceedings of the International Conference on Spoken Language Processing*, 1998.

[3] D. Colton, R. Cole, D. Novick, and S. Sutton. A laboratory course for designing and testing spoken dialogue systems. In *Proceedings of the International Conference on Audio, Speech and Signal Processing*, 1996.

[4] C. Draxler, R. Grudszus, S. Euler, and K. Bengler. First experiences of the german speechdat-car database collection in mobile environments. In *Proceedings of the European Conference on Speech Communication and Technology*, 1999.

[5] P. Heeman and J. Allen. The Trains spoken dialog corpus. CD-ROM, Linguistics Data Consortium, 1995.

[6] T. Lander, R. Cole, B. Oshika, and M. Noel. The OGI 22 language telephone speech corpus. In *Proceedings of the European Conference on Speech Communication and Technology*, 1995.

[7] A. Moreno, B. Lindberg, C. Draxler, G. Richard, K. Choukri, J. Allen, and S. Euler. Speechdat-car: A large speech database for automotive environments. In *International Conference on Language Resources and Evaluation*, 2000.

[8] J. Schalkwyk, J. de Villiers, S. van Vuuren, and P. Vermeulen. CSLUsh: an extendible research environment. In *Proceedings of the European Conference on Speech Communication and Technology*, 1997.

[9] K. Shobaki, J. Hosom, and R. Cole. The OGI kids' speech recognizers and corpus. In *Proceedings of the International Conference on Spoken Language Processing*, 2000.

[10] H. van den Heuvel, A. Bonafonte, J. Boudy, S. Dufour, P. Lockwood, A. Moreno, and G. Richard. Speechdat-car: Towards a collection of speech databases for automotive environments. In *Proceedings of the Workshop for Robust Methods for Speech Recognition in Adverse Conditions*, 1999.

[11] H. van den Heuvel, J. Boudy, R. Comeyne, S. Euler, A. Moreno, and G. Richard. The speechdat-car multilingual speech databases for in-car applciations. In *Proceedings of the European Conference on Speech Communication and Technology*, 1999.