ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

10.21437/Eurospeech.1999-224

FIRST EXPERIENCES OF THE GERMAN SPEECHDAT-CAR DATABASE COLLECTION IN MOBILE ENVIRONMENTS

Christoph Draxler, Dept. of Phonetics, University of Munich
Robert Grudszus, Robert Bosch GmbH, Stuttgart
Stephan Euler, Robert Bosch GmbH, Stuttgart
Klaus Bengler, BMW AG, Munich

ABSTRACT

In SpeechDat-Car, speech databases for speech driven devices and services for mobile environments are collected for nine European languages. The German SpeechDat-Car installation was the first fully equipped platform within the project. It has served as a testbed for the recording software for the entire project, and as an opportunity to perform technical and organizational feasibility tests for the German data collection. The main results are: i) drivers feel comfortable with the prompts displayed on an LCD screen attached to the dashboard, ii) the audio-visual mode of prompting allowed the elicitation of speech even under difficult traffic conditions, and iii) the original synchronization protocol could be made more efficient by 20%, so that the 129 items of one recording session can now be recorded in 45-50 minutes.

1 INTRODUCTION

Driver information systems are becoming increasingly complex as more and more functions are integrated in modern cars. Automatic speech recognition will be important for the safe operation of the future systems. In order to train robust speech recognizers for the automotive environment a large amount of speech material is required.

A large database provides a good representation of different phonetical contents from a wide range of different speakers in various driving situations. The use of such a database will allow an improved training of robust recognizers and provides a good data base for the validation of the recognition performance.

1.1 SpeechDat-Car

SpeechDat-Car is an EU-funded project (LE4-8334) for collecting speech data in mobile environments, especially cars. SpeechDat-Car is coordinated by Matra Nortel, France.

The SpeechDat-Car database languages and the project partners collecting them are listed in Table 1.

The vocabulary consists of 46 SpeechDat(II) items, 73 application words for car operation and teleservices, and 10 pseudo-spontaneous speech from simple vehicle control, navigation, and telecommunications tasks [1] (Table 2). To allow an exchange of data with the VODIS project [7], the VODIS application word list is also recorded in SpeechDat-Car.

Danish	University of Ålborg Sonofon
Dutch (Flemish)	Lernout & Hauspie
English	Vocalis
Finnish	Tampere Technical University Nokia
French	Matranortel Renault
German	Robert Bosch BMW University of Munich
Greek	Knowledge
Italian	Alcatel ITC-IRST
Spanish	Polytechnic University of Catalonia Seat-Volkswagen

Table 1: SpeechDat-Car Databases and Partners

Material	Item	Count
SpeechDat	isolated digits	4
	digit chains	8
	numbers	1
	money amounts	1
	date and time expressions	5
	read and spontaneous spellings	7
	geographical, person and company names	7
	phonetically rich words and sentences	13
SpeechDat-Car, VODIS	mobile phone	13
	teleservice	22
	car operation	32
	language dependent keywords	2
	voice activation sentences	2
	spontaneous sentences	10
	voice activation sentences	2

Table 2: SpeechDat-Car Database Contents

In each language, 600 sessions will be recorded. All databases will be validated according to uniform validation specifications [6] by SPEX in the Netherlands.

The SpeechDat-Car database specifications are publicly available [1], [2], [4]. They can be downloaded from the SpeechDat-Car web server at http://www.speechdat.org/

1.2 German SpeechDat-Car

The German SpeechDat-Car partners are Robert Bosch GmbH and BMW AG, with the Department of Phonetics of the University of Munich (IPSK) as a subcontractor. Bosch has been involved in the VODIS I and VODIS II projects. In these projects the feasibility of speech input for driver information systems has been demonstrated. The SpeechDat-Car data base extends the data bases collected in VODIS in the amount of material in each language and adds new European languages.

The IPSK has been subcontracted to record, annotate, and distribute the German SpeechDat-Car database.

2 TECHNICAL SETUP

SpeechDat-Car recordings consist of a multi-channel high bandwidth recording in the car, and a synchronized recording via GSM mobile phone.

Wherever possible, the driver should also be the speaker. An experimenter guides the speaker through the recording session and coordinates the progress of the recording with the outside traffic. In some countries however, using a mobile phone while driving is not allowed, even if a hands-free car kit is used. In these countries, the speaker is the co-driver.

2.1 Car installation

Bosch is responsible for the technical setup of the mobile platform. Based on the experience from the database collection in VODIS, Bosch specified the PC-based multi-channel speech recording system in the car.

For the German SpeechDat-Car data collection, a standard BMW 318i is used. The mobile recording platform (PLTM) is mounted on shock-absorbers on a 2cm thick plywood base in the car boot. A separate battery feeds the PLTM (Fig. 1).

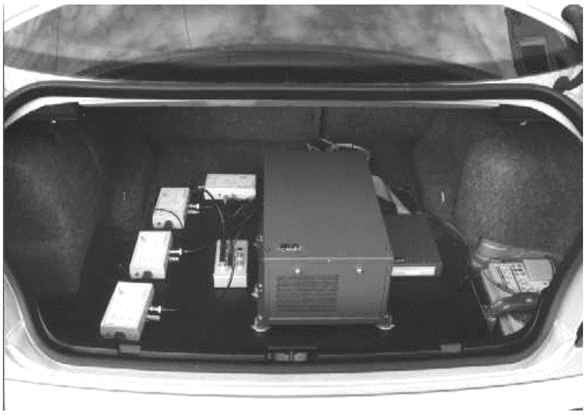

Fig. 1: PLTM Installation

The PLTM consists of a 266 MHz Pentium PC under Windows NT running on 12V. The PC is equipped with an internal 4 GB hard disk, an internal removable hard disk (2 GB Jaz) drive, and an external floppy drive. A color LCD screen is connected to the PC. This display is mounted on the dashboard to the right of the steering wheel (Fig. 2). The PC is operated via an infrared keyboard with an integrated mouse [4].

Audio data is recorded via L&H microphone preamplifiers and a 4-channel DataTranslation A/D board. Each audio channel is recorded at 16 KHz 16 bit.

Four microphones are used: a Shure headset microphone for the speaker, and 3 mouse microphones (AKG and Peiker) mounted on the A-pillar, over the speaker’s head behind the sunvisor, and in a mid-position near the rear view mirror (Fig. 2).

A hands-free mobile phone is mounted in a car kit; its microphone is located in mid position on the ceiling.

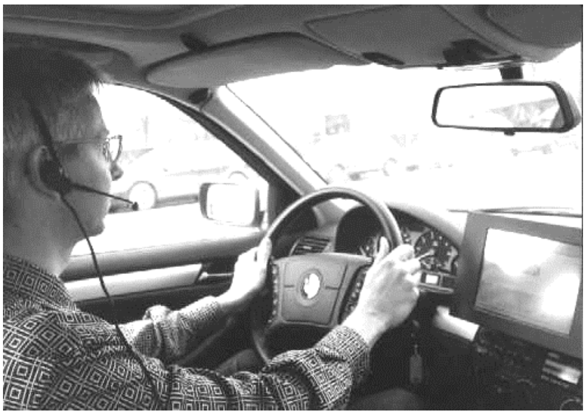

Fig. 2: Microphone and LCD-Display Positions

2.1.1 PLTM Software

The PLTM software was developed by Matra-Nortel. It is recommended that all SpeechDat-Car partners use this software.

The PLTM software features speaker and recording environment administration, recording and GSM tests, SpeechDat-Car recording procedures with synchronization protocol, and GSM phone connection monitoring.

A recording session consists of the recordings of each item. Recordings can be repeated, e.g. in case of errors, and aborted and continued later.

The software is operated by the experimenter. He or she enters the speaker (age, gender, dialect) and environment data (car equipment used, weather conditions, traffic and street conditions), and starts and stops the recording of every item.

For synchronization with the mobile phone recordings, the following protocol was originally specified:

- PLTM issues three DTMF tones that contain the current item code CCD.
- It then pauses for a predefined period Δ_1 (usually 0.5 to 1.0 s) to allow PLTF to start recording.
- PLTM then starts recording: first it issues two synchronization DTMF tones, then it records Δ_0 (minimum 1s) of environment noise, then it issues a beep to prompt the speaker.
- Recording is terminated either by the experimenter or a software timeout.

The synchronization protocol is shown in Fig. 3 A.

Each item recording is subdivided into three phases: initial display of prompt, recording preparation, and the recording proper. These phases are marked visually and auditively: During the initial display of the prompt the background is red, during the preparation it is yellow, and for the recording it is green. During the initial display of the prompt, the DTMF tones corresponding to the current item number can be heard, the beep marks the begin of the speech recording, and a sequence of three DTMF tones mark its end.

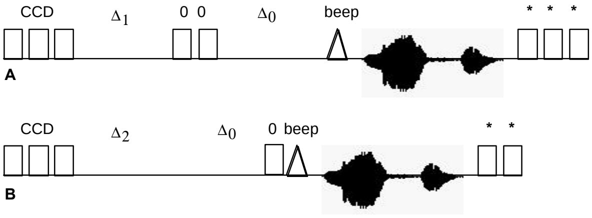

Fig. 3: Synchronization Protocols

2.1.2 Telephone recording installation

The German telephone recording platform (PLTF) consists of a 486 PC running SCO UNIX. It is connected to the telephone network via a primary rate ISDN interface. The proprietary PLTF software records the mobile phone speech signal from the car. Every item recording is stored in a separate file. File names are computed from the session and item code transmitted by PLTM. PLTF acknowledges the recognition of the item codes by a beep sent to PLTM, and overwrites previous files for repeated items.

3 ANNOTATION

The annotation of the SpeechDat-Car recordings is performed by the Department of Phonetics at the University of Munich. For this annotation, the WWWSigTranscribe software has been extended with a graphical display of both the PLTM and the PLTF signals to allow a quick visual control of the signal file contents. (Fig. 4) [3].

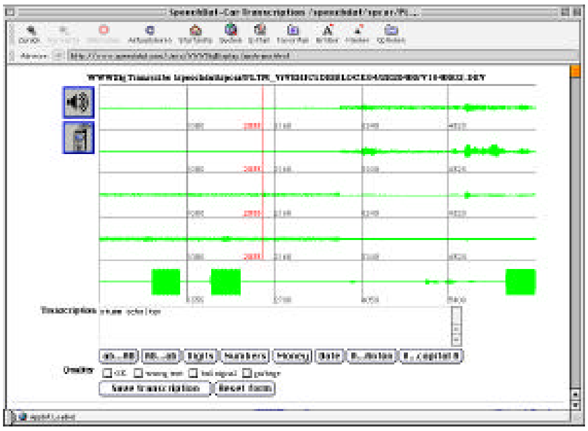

Fig. 4: WWWSigTranscribe Annotation Tool

4 EXPERIENCES

The German SpeechDat-Car recording platform was the first within the project to be fully functional. It has thus been used extensively as a testbed for the PLTM recording equipment in general and the for the German recording procedure in particular.

4.1 Platform installation and test

During the first phase of the test recordings (December 1998 to January 1999) the prefinal version of the PLTM software was installed and the hardware was tested. A few recording sessions with a small number of items were performed, both in a stationary and a moving car.

These first tests have uncovered two major problems:

- The operation of the PLTM software was slow and error-prone, and
- the headset microphone and its microphone amplifier did not function properly with the 12V power supply from the second car battery.

As a consequence, the PLTM software was redesigned, with the most important improvement being a streamlined and safe operation. Now stepping through the recording procedure is possible by pressing a single key, and more complex key sequences are required only for non-standard situations, e.g. recording repetitions, re-establishment of the GSM connection, etc.

The headset microphone amplifier worked well with a 9V block battery. However, because of the limited capacity of these batteries, this a potential source of problems.

4.2 Recording procedure tests

The new release of the PLTM software was installed early February 1999. With this software, the first full-length recordings were made.

For these recordings, students were recruited in the university cafeteria. As an incentive to participate, students were offered DM 20.- (10.- Euro) for a recording of approximately 60 minutes. It turned out that for the students money was less important than the opportunity to drive a high-quality car and to participate in a scientific project.

The recordings were carried out in a manner similar to the final production recordings: the experimenter briefly outlined the project and the necessity of the recordings. Then he instructed the test person. This instruction covered the operation of the car: seat and mirror adjustment, position of important switches (light, wipers, indicators, etc.), use of the automatic transmission, and a braking test.

A total of 15 recordings were made, with 7 male and 8 female speakers. Of these recordings, 10 took place in city traffic, both during light and rush-hour traffic, and 5 recordings were made while driving at high speed (~ 120 km/h) on a highway. During 10 recordings, the weather was dry, during 2 recordings the streets were covered with snow, and during 3 recordings it rained. 2 recording sessions were taped on video; these videos are intended as demonstration material, e.g. for journalists and also for project partners who have to convince insurance companies that having the driver read and speak while driving is not too distracting. It must be noted that in 14 recording sessions the GSM connection broke down at least once.

Initially, a complete recording session of 120 items took approximately 75 minutes; later on, the duration dropped to approximately 60 minutes. This reduction can be attributed primarily to the growing experience of the experimenter: The instruction of the driver was standardized, routes with a high risk of GSM connection failure (e.g. near large construction sites, iron bridges, etc.) were avoided, and the experimenter got more efficient in the operation of the software (quicker progression to next item, fast re-establishment of lost GSM connection, etc.).

The signals recorded during this phase showed to be of a low quality: The signal levels were quite low and the cutoff frequency was found to be near 1.5 KHz, effectively deleting high frequency parts of the speech signal. The low signal levels can be attributed to both the recording software which does not allow a precise metering of the recorded speech signal, and adjustment problems of the microphone and the driver's position.

The low cutoff frequency was caused by incorrect filter parameter settings; using correct setting solved this problem.

Finally, as a result of this recording phase, the synchronization protocol between PLTM and PLTF was simplified (Fig. 3B): the delay Δ2 to allow PLTF to start recording could be reduced substantially (down to 0 s); beep and synchronization DTMF tone are output immediately after each other instead of with a small pause in-between. Furthermore only one or two DTMF tones are needed to detect the end of recording. This new protocol has reduced the recording time by 3 to 5 seconds per item. As a consequence, the duration of an entire recording session is cut down by 6 to 10 minutes.

4.3 Prevalidation recordings

For the SpeechDat-Car data collection to begin, a prevalidation database must be created and validated by SPEX. This database consists of 6 complete recordings and the full set of documentation files.

The prevalidation recordings require a formal assessment of the recording platforms, both in terms of operational stability and user acceptability.

With the new synchronization protocol and the final version of PLTM installed, 9 recordings were carried out until 1st May. Both PLTF and PLTM passed the operational tests. All breakdowns of GSM connection could be attributed to external factors, such as bad signal reception or interference. However, in some cases re-establishing a broken GSM connection was not feasible from the PLTM platform because the software did not recognize the current status of the mobile phone. In these cases, either the mobile phone had to be reset manually, or PLTM had to be restarted to continue the recording session.

Due to the new synchronization protocol, the duration of a recording could be reduced to 45-50 minutes, a 20% improvement.

The nine speakers who participated in the prevalidation recordings were asked to assess the recordings. The results show that speakers feel comfortable with the technical setup, and that the duration of a recording session is acceptable (Table 3).

The prevalidation recordings are now being annotated. Results are expected by the end of May 1999.

How did appear to you the system response time?	3 long 5 medium 1 short
Did you at some point in the session want to end the recordings?	9 no
How well could you follow the displayed information during the session?	4 fair 4 good 1 excellent
How well did you appreciate the screen	3 fair

readability?	5 good 1 excellent
Were there any items that you found difficult to pronounce?	2 yes 7 no
What did you think of the actual length of the sessions?	1 long 6 medium 2 short
What is your general impression of the whole procedure?	5 fair 4 good

Table 3: Speaker Assessment

5 CONCLUSIONS AND OUTLOOK

The complexity of collecting multi-channel speech signals both in the car and via a GSM connection has led to a very long specification and testing phase for the entire project, with revisions as late as March 1999 of such central elements as the vocabulary to be recorded, the recording software functionality, and the synchronization protocol.

Now that the technical and procedural issues are solved, real production recordings can begin within the project. The German SpeechDat-Car recordings will effectively start in June 1999. Until the end of the year, roughly 100 recording sessions per month are scheduled. Annotation is expected to go on in parallel to the data collection, so that the database can be delivered for the final validation in April 2000.

6 REFERENCES

[1] Dufour, S.: Specification of the Speech Database (Definition of Corpus, Scripts, and Standards) Car Environments and Speaker Coverage; LE4-8334-SD1.12, 1999

[2] Draxler, Chr.: Specification of Database Interchange Format, LE4-8334-D1..33, 1999

[3] Draxler, Chr.: WWWSigTranscribe – A Java Extension of the WWWTranscribe Toolbox, Proc. LREC, 1998

[4] Grudszus, R.: Specifications of the Car Recording and Remote Fixed Platforms, LE4-8334-CD1.14, 1998

[5] Senia, F.: Specification of Database Interchange Format, LE2-4001 - SD1.3.1, 1997

[6] van den Heuvel, H.: Validation criteria, LE4-8334-CD1.31, 1999

[7] VODIS project: http://www.linglink.lu/le/projects/vodis/index.html