SPEECHDAT-CAR: SPEECH DATABASES FOR VOICE DRIVEN TELESERVICES AND CONTROL OF IN-CAR APPLICATIONS

Marc Sala (1), Francisco Sánchez (1), Heino Wengelnik (2), Henk van den Heuvel (3), Asunción Moreno (4), Enrica Deregibus (5), Gael Richard (6), Eric Le Chevalier (7)

(1) SEAT S.A., Martorell, Spain
(2) Volkswagen AG, Wolfsburg. Germany
(3) SPEX, Nijmegen, Netherlands
(4) UPC, Barcelona, Spain
(5) FIAT CRF, Orbassano, Italy
(6) Matra Nortel Communications, Bois d'Arcy, France
(7) Renault, Guyancourt, France

ABSTRACT

The SpeechDat-Car project included in the 4th framework of the European Community's Language Engineering Programme, started in April 1998 with a duration of 30 months. It is a common initiative of car manufacturers, telephone communications operators, companies active in voice operated services and Universities that aims at collecting a set of speech databases in nine different languages to support training and testing of robust multilingual speech recognition for in-car applications.

This paper describes the database requirements, the background of the project, the design and validation of the databases, the definition of recording platforms, and main goals from the automotive exploitation point of view.

1. INTRODUCTION

In order to increase safety and comfort levels in cars and taking into account the high number of new systems that could distract the attention from the driving task, Speech Recognition is becoming an essential field in the car environment. In-car speech recognition allows the driver to control car accessories as telephone, radio, CD player, navigation systems, windows, etc. Furthermore, some voice teleservices are already being offered in the present and it is expected that other systems will appear in the near future.

The SpeechDat-Car project, included in the 4th framework of the European Community's Language Engineering Programme, aims at providing voice-operated services for in-vehicle applications. The current speech recognition technologies need very large speech databases recorded in the same environment in which the recognition system will operate. The objective of the project is to produce such databases for the car environment in different situations (cities, highways, etc.). As final result, 300 speakers per language and nine European languages are to be recorded in two different car conditions selected among a sample of seven scenarios, frequently met in car environments (see 3.3.2), in order to generate robust databases to support speech recognition for in-car applications. The speech is recorded in a platform installed in the car (mobile platform) and it is simultaneously transmitted over a radiotelephone (GSM) and recorded in a far-end platform to support teleservices accessed from cars.

The speech databases intend to support training of robust multi-lingual speech recognition for in-car applications:

- Voice dialling for radiotelephones in cars
- Accessing remote teleservices and voice driven services from car telephones
- Controlling car accessories

The Consortium participants are car manufacturers (Renault, BMW, Fiat, Seat-Volkswagen), Telecommunications terminals and Networks Manufacturers (Matra Nortel Communications, Nokia, Alcatel), Mobile Telephones Networks Operators (SONOFON in Denmark) and companies active in voice-operated services (Lernout & Hauspie, Robert Bosch Gmbh., Vocalis, Knowledge S.A.) and universities (University of Munich, University of Patras-WCL, Universitat Politècnica de Catalunya, University of Aalborg-CPK, DMI, IRST, SPEX). SPEX, as Validation Centre, is responsible to perform the validation of each produced Speech database according to the pre-defined SpeechDat-Car standards. Matra Nortel Communications does the project management.

2. REQUIREMENTS TO THE FINAL DATABASE FROM THE AUTOMOBILE MANUFACTURERS POINT OF VIEW

Convenient control of car electronics like car stereo, mobile phone, navigation system and some car-body functions with voice is, without any doubt, a reasonable way to improve traffic safety by freeing the hands and the attention of the driver from looking for the right keys somewhere in the cockpit.

Using speech recognition to dial a number with the mobile phone or choosing the destination desired in the navigation system, relaxes significantly the driving stress and is a comfort option that can also avoid accidents caused by distraction of the driver.

At the same time, a voice-controlled device has to be ergonomic, have a reasonable price, be easy to be used and have a logical structure that do not force the user to keep thousands of command words in mind.

To achieve a good recognition rate is a very important point in order to give sense to these devices and to obtain the full satisfaction of the user. Although this is an important goal that has to be accomplished in any speech recogniser, in the case of the car environment the reliability of the system is especially significant. Problems caused in case of misunderstanding of a command could be more distracting than operating the devices with the hands. In this case, the driver does not feel himself comfortable when using a system that he can not completely control.

A speech recogniser operating in an in-car environment has to be robust against additional noises or factors that could distort or affect the recognition of the pronounced command. Some of the problems that have to be faced are:

- Speaker characteristics (support different pronunciations, accents, voice tones...)
- Driving stress
- Engine noise at different driving conditions and speed
- Traffic noise caused by other vehicles, horns, etc.
- Climate factors (rain with different intensities)
- Type of road
- Open windows
- Problems in transmission of GSM telephony speech from moving vehicles
- Similar commands (in case of names)
- Microphone mounting position
- Different car types
- Other noisy car devices (audio, fan...)
- Etc.

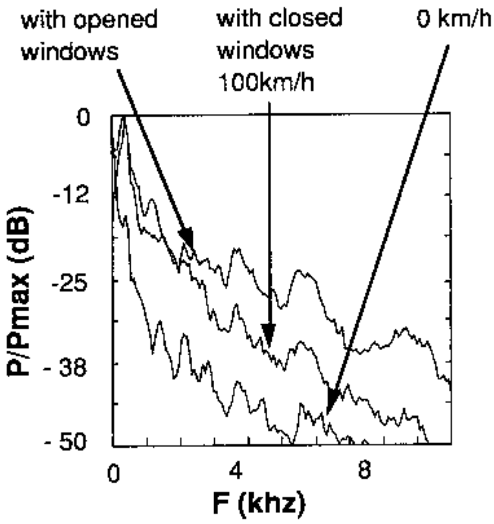

Figure 1. Example of power distribution of noise [12]

A representative example of the variability of noise in the car environment can be seen in **Figure 1**, in which a spectral analysis under different noise conditions is presented. The signals correspond to 1 second of noise. A peak in low frequencies (about 150 Hz) is detected, corresponding to the fundamental frequency of the engine. The spectrums decrease quasi exponentially, and after 4 kHz the power is very low. When the speed increases or when we open the windows, the noise power in higher frequencies (above 150 Hz) increases. Also the signal/noise ratio decreases (from 15 dBs in the case of 0 km/h and engine running to –10 dBs in the case of 90 km/h with opened windows).

Since multi-style training increases the amount of training data by introducing more variability in speaking style or environmental conditions to obtain more robust models, the databases resulting from the project are expected to be a good material to develop robust speech recognisers when operating in all these above mentioned conditions.

3. DATABASE DESIGN

3.1. Database specifications

Since best performance for recognition systems can only be obtained when the recording is made in context, the database will be directly recorded in the car.

As the specifications have been developed in consultations with all partners, and are intended to provide a common resource for developing general speech services, the specifications are identical for each language database.

3.2. Recordings

Three hundred speakers per language and 9 European languages will be recorded directly in the car, in the driver's position (except for those countries where this is not legal). The co-driver will operate the recording system, which includes a large display screen. The speech material is specified to cover several application needs (telephone, navigation, autoradio, commands and control). Multi-channel recordings will be performed simultaneously in the car and through the GSM network (see section 4).

3.3. Database content

3.3.1. Main features

The databases are intended to provide material for both training and testing of several classes of speech recognisers, including isolated word systems, word-spotting systems and vocabulary independent systems which use either whole word or subword modelling approaches. The databases aim to provide sufficient data to train basic speaker independent recognition systems to support common teleservices in each language represented.

The following targets motivate the SpeechDat-Car specifications:

- Telecommunications systems (IVR, dialling, remote access to teleservices and servers)
- Car equipment (radiotelephones, car radio, car accessories, navigation)

A set of application words has been selected with equal repartition for car and specific Telecom applications.

The main features of the databases includes common application words for system and menu navigation, directory assistance names, route and navigation words (cities, regions main roads, roads names), spelling for specifying geographical names or places, digits and numbers for dialling, accessing IVR (Interactive Voice Response) services and option selection, phonetically rich material to train vocabularies for future services using voice interfaces, and spontaneous sentences for future dialogue based applications.

3.3.2. Language, Speaker and environment coverage

The full database specification has as its goal to provide recordings of 300 different speakers in two environments each for nine of the EU languages (namely French, Flemish/Dutch, German, British English, Greek, Spanish, Danish, Finnish and Italian).

The speaker coverage is set to uniformly representative speakers of all ages (classified in 3 categories between 18 and over 60 years old, speakers of older age are though permitted) and to represent all dialectal regions of a language (Typically 5 regions are defined and at least 50 speakers for each regions need to be recorded). Additionally, the distribution between male and female should be respected with a possible 5% deviation (Typically this means that the number of male speakers and the number of female speakers have to be included in the interval of 45 – 55 %).

Each recording will be made in two different environments taken among 7 possibilities:

- Car stopped with motor running
- Car in town traffic
- Car in town traffic with open windows
- Car moving at a low speed with rough road conditions such as freeway or out of towns' roads
- Car moving at a low speed with rough road conditions such as freeway with noisy conditions
- Car moving at a high speed with good road conditions (smooth asphalt such as highway)
- Car moving at a high speed with good road conditions (smooth asphalt such as highway with audio equipment on)

In addition, some important information such as weather conditions, accessories used during recordings (windscreen wipers, ventilation...), size and brand of the car will be collected during the recordings.

3.3.3. Targeted application products

The different goals of this project are applications for three kinds of products: the mobile phones, the IVR applications and the car products.

For mobile phones, the targeted functions include number dialling, agenda (use of name to dial frequently called persons), access to local remote teleservices by phone. The corresponding vocabulary therefore includes telephone numbers, personal names, common application words for menu navigation and specific application words for the use of mobile phone.

For IVR products, the targeted functions based on advanced speech recognition technologies that match with the customer needs include message forwarding, voice mail, email box, fax services, telematic applications, telephoning banking, directory assistance and customer support services. The corresponding vocabulary therefore includes assistance words for interactive voice services, digits and numbers, credit card number, yes/no questions and common application words for IVR functions.

For car equipment products, the targeted functions include navigation and route guidance systems, needs for speech input system, and possible extension to future applications. The corresponding vocabulary includes phonetically rich words and sentences, navigation, route guidance words and commands, directory assistance names and common words for car accessories and car radio.

3.3.4. Item distribution of a recording session

The item distribution for each recording session is indicated in the table below, with an indication of the repetition of such item in the whole database:

Items	Nb/call	# per session	Remarks
Isolated digit	4	200	
Digit strings			
10 digits in isolation	1	600	
Telephone number	3		Triples, pairs, etc.
Spontaneous telephone number	1		
Credit card number	1	4	Out of a list of 150
PIN code	1	4	Out of a list of 150
Sheet number	1		For processing purposes (option: DTMF)
Natural numbers	1	600	100000<N<1000000, maximise –teen, -ty
Money amounts	1	600	Range from newspaper to car prices in Euro and Cents
Dates			
Spontaneous	1		
Absolute	1	*	
Relative	1	60	Set of approx. 10 language dependant
Times			
Spontaneous	1	**	
Analog	1	100	Given in text form (***)
Names			
City of birth (spontaneous)	1		
Most important cities	2	8	50 common names to all partners, 100 language dependent
Most important companies	2	8	Option: street names instead (Vodis compatibility) Out of 150
Forename (spontaneous)	1		
Forename + surname	1	4	Option: surname (Vodis compatibility) Out of 150 names
Spelling			
Forename/surname	1		Spontaneous, linked with the spontaneous forename
Word/name	4		(Map to street and person names)
Artificial word	1	600	For coverage spoken in isolation
City name	1		Linked to one of the city names
Phonetically rich words	4		200 samples per phoneme
Phonetically rich sentences	9		250 samples per phoneme; max. 5 repetitions/sentence
Application words	67	200	Out of 201
Spontaneous phrases using an keyword	2		
Keywords for voice activation	2	240	Out of 5
Additional language dependant keywords	2	120	Out of 10 language dependent
TOTAL	119		

Table 1. SpeechDat-Car corpus contents

* : 50 month names, 85 weekdays, 20 ordinal numbers

** : 25 (50) for hours, 10 for minutes and seconds

*** : For example: quarter past, half, quarter to, sharp, morning, just before…

4. RECORDING PLATFORMS

Each of the nine recording groups for the nine European languages uses common recording platforms. Two types of recordings were considered necessary. First, wideband recordings (60-7000 Hz) for systems which are installed and operate in the car itself; second, narrow band recordings (300-3400 Hz) for systems that operate centrally outside the car and obtain their spoken input from the driver over the cellular telephone network. For this reason, each group will use two recording platforms:

- A ‘mobile’ recording platform (PltM) installed inside the car, recording multi-channel speech utterances in a high bandwidth mode (16kHz sample frequency)
- A ‘fixed’ recording platform (PltF) located at the far-end fixed side of the GSM communications simultaneously recording the speech utterances coming from the car (8 kHz sample frequency)

Multi-channel recordings are performed simultaneously in the car and through the GSM network.

The recordings are made through an Acoustic front-end (AFE) installed inside the car and connected to the recording platform PltM. Three kinds of AFEs are used simultaneously during the recordings: a close-talking microphone, a remote noise cancelling microphone with 3 Handsfree microphones placed at different locations in the car and a commercial Handsfree car-kit equipment for GSM radiotelephones in cars.

The synchronisation mode between the PltM and PltF is based on use of DTMF tones emitted from the GSM terminal placed in car.

4.1 Car Platform

The mobile recording platform in the car (PltM) is the Master platform, it uses a PC to drive the recording process and to control the remote PltF platform. Data Acquisition is performed by a dedicated hardware in the PC and the storage is made directly on hard disk. The recordings are always made on four channels (1 close-talk signal as reference and 3 far-talk signals). The speaker is usually the driver and the co-driver operates the system. The car is equipped in such a way that the speaker can also be the co-driver (for safety or country-specific legal reasons). The PItM recording platform is shown in **Figure 2**.

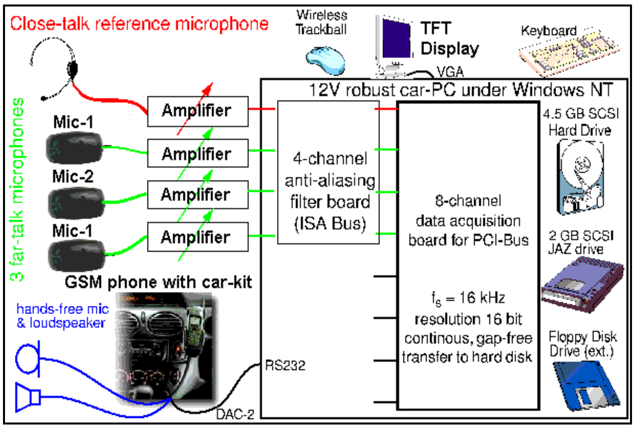

Figure 2: Overview of the PItM recording platform

4.1.1 Acoustic front-end

4.1.1.1 Close-talk microphone

A cardioid dynamic, head-worn microphone is used as close-talk reference microphone for the recordings. It offers hands-free operation without user-fatigue and features unidirectional polar pattern for effective noise reduction with a smooth natural voice frequency response from 50 to 15000 Hz. It can therefore be used in noisy conditions without loss or masking of voice signals.

4.1.1.2 Far-talk microphones

While recording speech in the car there is always a high-power low-frequency background noise (f<70Hz) which leads to poor utilisation of the dynamic range of the A/D converters. As there is no extra high-pass filtering of the speech signals before A/D conversion, microphones designed for automotive applications are used featuring a slight „high-pass“ effect in their frequency response for lower frequencies.

<u>Microphone type:</u>

A two-microphone scenario has been selected for all the recordings with two far-talk microphone types with the following characteristics:

Mic-1: pre-amplified condenser microphone with hypercardioid directivity.

Mic-2: pre-amplified condenser microphone with super cardioid directivity.

Both have mouse housing, conform VDA automotive standards and fulfil EMC and GSM constraints.

<u>Microphone arrangement:</u>

Suitable positions for the far-talk microphones have been selected on the basis of SNR values measured in real car environment. The three best positions in terms of the SNR were selected:

- A: at the ceiling of the car near the A-pillar
- B: at the ceiling of the car in front of the driver behind the sunvisor
- C: at the ceiling of the car over the mid-console (near the rear mirror)

To permit also to the co-driver to play the role of speaker, these three positions have been duplicated in some cases also for the front passenger position.

This position C is optimal suited for in vehicle integration although suboptimal in comparison to position A with regard to wind and ventilation effects. As a compromise, the following microphone arrangement is used for the two far-talk microphones as specified in 0 (see **Table 2**).

Speakers	A	B	C
150	Mic-1	Mic-2	Mic-1
150	Mic-2	Mic-1	Mic-2

Table 2: Far-talk microphone arrangement

4.1.2 Microphone amplifiers

Speech has to be recorded in different environmental conditions ranging from low noise conditions, e.g. stopped car with engine running, to noisy conditions, e.g. car moving in high speed on the highway. Additionally there is speaker variability from speakers with a weak voice to speakers with a quite loud one. Due to these reasons adjustable microphone amplifiers are used that allow the selection of one suitable fixed input signal dynamic range for all the specified scenarios of the recordings.

4.1.3 Handsfree car kit

The GSM phone with hands-free car-kit and special hands-free microphone is installed in the car. The hands-free microphone is mounted at the ceiling of the car over the mid-console, i.e. beside the far-talk microphone position C. As some speech data are collected in the presence of car audio, it is possible to activate and deactivate the mute functionality.

The remote control of the GSM phone comprises the following basic functions:

- Make a call
- Generate DTMF tones
- Hang up
- Detect dropped call

The remote control of these functions is performed by a simple Windows NT program using the standard AT command set without using any special API. The GSM mobile mounted in the car-kit is connected to the serial port of the PC.

4.1.4 Car-PC

The computer for in-car use powered by the car battery (12V DC), is mounted in the boot of the car.

The PC features a Pentium II 266MHz processor, 64MB RAM, 4.5 GB SCSI hard disk drive, CPU board with sound chip. (Operating system is Windows NT4). It hosts the data acquisition board and anti-aliasing filter board.

As backup medium an internal 2GB SCSI JAZ drive or a removable SCSI Hard Disk Drive is used.

A flat panel TFT colour-display for in-vehicle use is attached to the windscreen or the dashboard of the car.

4.1.5 Data Acquisition Hardware

The data acquisition board installed in the Car-PC is a combination of two plug-in boards:

- Multifunction data acquisition board
- Anti-aliasing filter board

The data-acquisition board is a PCI bus-mastering multifunction data acquisition board with either 16 single-ended or 8 differential analogic inputs, 23 digital I/O lines and 4 counter/timer channels. It provides multiple input ranges by using software-programmable gain amplifiers.

The anti-aliasing filter board consists of an ISA bus carrier board for holding up to 4 filter modules and 4 gain modules each offering two channels. For this application, 2 Cauer filter modules with software selectable cut-off frequency from 0.1Hz to 50kHz have been selected. The extremely sharp cut-off-frequency roll-off makes Cauer filters ideally suited for most anti-alias applications. They also have a good passband flatness and low wideband noise.

4.2 PltM Recording Software

The software has been developed under Windows NT 4 using Microsoft Visual C++ 5.0.

All modules (AT commands API, data management, recording API,...) are independent from each others. So it is possible to update a module without influencing the others. The output of the module is an organised set of files and a data structure. The output is configured allowing interfacing with any annotation tool.

The PltM recording software can be decomposed into following independent tasks:

Telephone control API

The API for telephone does:

- Write a command string on serial port
- Read response on serial port

The main AT commands are:

- Hook on/off
- Dial phone number
- Send DTMF string
- Check the status of the GSM link

Multi-channel board recording API

The API to control the multi-channel data acquisition board control the recording process:

- Start / End recording
- Four channels are recorded and stored multiplexed. De-multiplexing is embedded in the annotation tool.

A software layer specifies the interface between the multichannel device and the API.

User Interface (MMI)

The experimenter can access the software either via keyboard or mouse.

The PC-based software that controls the entire mobile recording platform interacts with both users (i.e. experimenter and test person) by two different means:

- **Visually** through the colour TFT screen on the dashboard of the vehicle
- **Acoustically** using pre-recorded messages played back through the car-loudspeakers

From the start-up screen the following three actions can be initiated:

- Entering session data
- Performing a system test
- Starting a recording session

Prompt file management

A tool that creates all the sheets (600) language independent (with item numbers) has been developed. For each language, the appropriate label is associated to each item number. Each language has an identical number of items.

4.3 Fixed Recording Platform (PltF)

The fixed recording platform, located at the far-end fixed side of the GSM communication, record simultaneously the speech utterances coming from the car.

The main characteristics of PltF are:

1. Direct connection to an ISDN line (either a Basic Rate Interface (BRI) with a maximum capacity of two channels or a Primary Rate Interface (PRI) with a maximum capacity of 30 channels).
2. With this kind of ISDN connection, the samples are directly written to the disk in the incoming digital format, i.e., with a sampling rate of 8KHz and the A-law coding. This speech coding standard was defined by the CCITT (now ITU) to compress telephone speech signal (usually with a bandwidth 300

Hz - 3000 Hz) by using 8 bits samples with a logarithmic law, (Recommendation G.707).

The fixed platform has the following functionality:

- DTMF detection (simultaneously with recording of speech)
- Full duplex operation (record while playing).

4.4 Synchronisation & Communication between Fixed and Mobile Platforms

The objectives of the communication between the two platforms are:

- Detecting if PltF is still alive during the recordings (and to repair a hang up);
- Allowing synchronisation of the recordings on the two platforms;
- Allowing the separation of the items in individual files.

In the project several protocols have been proposed to meet these goals. These protocols comprise a series of beeps and DTMF-codes transmitted by PltM to PltF to ensure that each recorded item is preceded by a simultaneous beep on all recording channels to allow rapid off-line synchronisation of the recordings on both platforms.

5. POST-PROCESSING OF THE SIGNALS: ANNOTATION AND VALIDATION

Apart from the validation in the car itself, there is a quality evaluation of each database produced, in order to ascertain that the high quality requirements of the project are met. This quality check is an integral part of each project in the SpeechDat family and is also termed "validation".

An independent validation centre, SPEX carries out this validation, which is also in charge of the database validation in the other SpeechDat projects. "Independent" is to say that the validation centre itself does not produce a database in the project, neither as consortium member nor as third party.

The following aspects of a database are checked and compared to the validation criteria as agreed by the consortium:

- Completeness and correctness of documentation;
- Compliance to the database format specifications;
- Completeness of recordings;
- Correctness of the distributions of individual items;
- Quality of the speech signals; balances of speaker and environmental distributions;
- Completeness of the lexicon;
- Quality of orthographic transcriptions.

The validation procedure comprises three steps:

1. Pre-validation. Each partner sends a complete minidatabase of 6 speakers to the validation centre after the platform is installed and before the main recordings start. The main goal of the pre-validation is to detect errors in the database design before the main series of recordings start.
2. Full validation of the complete database. After the check a validation report is edited. This report lists all the requirements that a database does and does not meet. In case of relevant deficiencies, the consortium decides upon the approval of the database.
3. Re-validation. This phase is entered only if the consortium and/or the producer considers it necessary that (part of) a database be rectified.

6. PROJECT LINKS

This project puts together the efforts of individual partners to generate multilingual databases for the in-car environment. Each partner responsible for a language will have access to the eight other languages represented in the project. In order to open the consortium to other languages, the consortium has defined a procedure which allows third parties that are interested, to become "external partners" of the project by producing car databases with all the SpeechDat-Car specifications, and thus to exchange their database with consortium members on an equal-by-equal basis [3].

Furthermore the results of the project will be presented in the European work group of the ETSI and within the AURORA Group defining Distributed Speech Recognition which particularly deals with the mobile environment, and thus also with the car environment. The SpeechDat-Car results will also be presented to the International work group ISO TC22WG8. In particular, it could be interesting to make recommendations about the different type of speech utterances to be used for activation of the car-equipment (including telephone).

In order to produce reusable resources, the consortium has procured to define pre-standard specifications with respect to expected functionality and command words. A first draw-up of the specifications was widely distributed and the feedback was considered in order to fix the specifications.

When completed and fully validated the SpeechDat-Car databases will be made available via the European Language Resources Association (ELRA).

7. TEST FOR USE OF DATABASES

All nine databases developed in the SpeechDat-Car project can train dedicated in-car speech recognisers. Thus the databases will help to improve the operation of in-car voice systems for nine European languages. In the frame of the project we want to support the later statement and demonstrate clearly the use of the SpeechDat-Car databases:

It is planned for several languages to determine two recognition rates of speech recognisers; one which will have been trained with a SpeechDat-Car database and the other not. The difference between each two recognition rates will be a direct and obvious measure for the value of the SpeechDat-Car databases.

Since the databases recorded in each language will be recorded in different car-types and from different car manufacturers, each car having different acoustic levels and distances from the speaker to the microphones, this influence will also be checked. The car speech recognisers trained with the databases obtained in the SpeechDat-Car will also be tested in car types different from the original ones where the database is recorded, in order to verify that the final product is robust in different environments.

8. WHERE ARE WE NOW

After having defined the specifications of the databases and the recording platforms, once tasks of mounting, testing and synchronisation of platforms are almost finished, the first recordings in cars are now ready to be started in all languages (beginnings of March 1999). In a short period of time it is expected to send the first recordings to SPEX in order to initiate the pre-validation of the databases. Next phase is the recording sessions in each language in the different regions specified. This task is foreseen to last until April 2000, when the databases recruited will be sent to validate.

After the conclusion of the pre-validation, reports containing the specification of the databases [4-8] will be made available.

The Consortium has recently approved a baseline document [3] for new partners in order to determine the conditions of acceptance for new databases or possible extensions of the existing ones. Please visit our WWW-site for additional information [9].

9. SUMMARY

To our knowledge, the effort developed during the first phase of this project enables the comprehensive investigations of the impact of real circumstances on speaker-independent speech recognition in the car. Although other initiatives have already worked in this field, SpeechDat-Car databases will provide a great amount of valuable information, being the first time that such a big amount of speech data is recorded for automotive purposes. Experience accumulated during the recordings of more than 5000 sessions will be of great use when designing ergonomic and robust speech recognisers in car environments.

In addition, final obtained databases in nine different European languages could also be used as an excellent basis when training voice-operated applications for vehicles.

10. REFERENCES

[1] Jean-Claude Junqua, Jean-Paul Haton. *Robustness in automatic speech recognition: Fundamentals and Applications*. Kluwer Academic Publishers

[2] Höge, H., Tropf, H., Winski, R., Van den Heuvel, H., Haeb-Umbach, R. & K. Choukri. *European speech databases for telephone applications*. Proceedings ICASSP 1997, München, Vol. III, pp.1771.1774, 1997

[3] Richard, G. & Boudy, J. *Baseline document for interested candidates for recording additional languages*. SpeechDat-Car Technical Report D0.1, 1999

[4] Dufour, S. *Specification of the car speech databases*. SpeechDat-Car Technical Report D1.12, 1998

[5] Van den Heuvel, H. *Validation criteria*. SpeechDat-Car Technical Report D1.3.1, 1998

[6] Van den Heuvel, H. *Orthographic transcription conventions*. SpeechDat-Car Technical Report D1.3.2, 1998

[7] Draxler, Christoph *Specification of database interchange format*. SpeechDat-Car Technical Report D1.3.3, 1998

[8] Grudzus, R. *Specification of the car recording and remote fixed platforms* SpeechDat-Car Technical Report D1.4, 1998

[9] URL:*http://speechdat.phonetik.uni-muenchen.de/SP-CAR/*

[10] URL:*http://www2.echo.lu/langeng/en/le1/vodis/vodis.html*

[11] Langmann et al. *CSDC- The MOtiv Car Speech Data Collection*

[12] Laurent Barbier, Chafic Mokbel, Gérard Chollet. *Trainable noise substraction filters for speech enhancement in the car*. L.Torres, E. Masgrau and M.A. Lagunas (eds.), *Signal Processing V Theories and Applications Vol II*. Elsevier Science Publishers B.V., 1990