SPEECHDAT-CAR. A Large Speech Database For Automotive Environments

Asunción Moreno (1), BørgeLindberg (2), Christoph Draxler (3), Gaël Richard (4),

Khalid Choukri (5), Stephan Euler (6), Jeffrey Allen (5)

(1) Universidad Politécnica de Cataluña, Spain; (2) CPK, Denmark (3) IPSK of the University of Munich.
(4) Lernout & Hauspie , France (5), ELRA, France, (6), Robert Bosch GmbH Germany.

Address: (1) UPC, Jordi Girona 1-3, 08034 Barcelona Spain
e-mail asuncion@gps.tsc.upc.es

Abstract

The aims of the SpeechDat-Car project are to develop a set of speech databases to support training and testing of multilingual speech recognition applications in the car environment. As a result, a total of ten (10) equivalent and similar resources will be created. The 10 languages are Danish, British English, Finnish, Flemish/Dutch, French, German, Greek, Italian, Spanish and American English. For each language 600 sessions will be recorded (from at least 300 speakers) in seven characteristic environments (low speed, high speed with audio equipment on, etc.). This paper gives an overview of the project with a focus on the production phases (recording platforms, speaker recruitment, annotation and distribution).

1. Introduction

Automatic speech recognition (ASR) appears to be a particularly well adapted technology for providing voice-based interfaces (based on hands-free mode) that will enable new in-car applications to develop while taking care of safety aspects. However, the car environment is known to be particularly noisy (street noise, car engine noise, vibration noises, bubble noise, etc...). To obtain an optimal performance for speech recognition, it is necessary to train the system on large corpora of speech data recorded in context (i.e. directly in the car). For this reason, language-specific initiatives for database collections have been developed since about 1990 [Langmann (1998)]. The European project SpeechDat-Car[1] aims at providing a set of uniform, coherent databases for nine European languages and for American English.

The SpeechDat-Car aims at continuing the success of the SpeechDat project [Draxler (1998); Höge (1998); Höge (1999)] in developing large-scale speech resources for a wide range of languages and for in-car applications (voice dialling, car accessories control, etc.). It will produce resources for ten languages: Danish, British English, Finnish, Flemish/Dutch, French, German, Greek, Italian, Spanish, and American English. The consortium of the project comprises 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 participation of external partners to the original consortium is also possible. Siemens is an 'external' partner.

It is also important to note that SpeechDat-Car commits itself to a strict validation protocol to ensure optimal quality and exchangeability of the databases [Van den Heuvel (1999)].

This paper gives an overview of the project with a focus on production phases. It is organised as follows: the next section describes the database specifications (database content, recording platforms and validation procedures). Then, Section 3 provides additional information on speaker recruitment and an extensive description of the annotation procedure and tools is given in section 4. The paper then concludes with a short section about database availability and dissemination.

2. Database specifications

Each database produced in the SpeechDat-Car project is intended to provide enough data to adapt speaker independent recognition systems to the automotive environment. Database contents were designed to cope with different applications. The design includes a phonetically balanced corpus to train basic speech recognition systems and an application corpus. The application corpus is aimed at two applications: Telecommunication systems (IVR, dialling, remote access to teleservices and servers) and Car equipment (radiotelephones, car radio, car accessories, navigation).

Each database contains recordings with 600 different sessions from at least 300 speakers. A session consists of either 119 or 129 read and spontaneous items recorded using five microphones installed in a car. Signals from four of these microphones are recorded and stored in a mobile platform installed in the car. The signal from the fifth microphone is transmitted simultaneously by GSM to a fixed platform connected to an ISDN telephone interface. The GSM signal is recorded in A-law format.

Recordings are conducted in different recording conditions. There are defined 7 environment conditions. Every environment is equally represented in the final database:
- car stopped with motor running
- car in town traffic
- car in town traffic, with noisy conditions
- car moving at a low speed with rough road conditions

Items	Corpus contents
Digits and strings of digits	
1	sequence of 10 isolated digits
1	sheet number (4+ digits)
1	spontaneous telephone number
3	read telephone numbers
1	credit card number (16 digits)
1	PIN code (6 digits)
4	isolated digits
Dates	
1	spontaneous date, e.g. birthday
1	prompted date, word style
1	relative and general date exp.
Spellings	
1	spontaneous, e.g. own forename
1	spelling of direct. city name
4	real word/name
1	artificial name for coverage
Money amount/ natural number	
1	money amount
1	natural number
Names	
1	spontaneous, e.g. own forename
1	city of growing up (spontaneous)
2	most frequent cities
2	company/agency /street names
1	forename/surname
Times	
1	time of day (spontaneous)
1	time phrase (word style)
Application words	
13	Mobile phone Application words
22	IVR functions keywords
32	car products keywords
2	voice activation keywords
2	language dependent keywords
Phonetically rich words	
4	phonetically rich words
Sentences	
2	phrases using an application word
9	phonetically rich sentences
10	Prompts for spontaneous speech

TABLE 1. SpeechDat Car Database Contents

- car moving at a low speed with rough road conditions, with noisy conditions
- car moving at a high speed with good road conditions
- car moving at a high speed with good road conditions with audio equipment on

Each session is manually annotated. Only speech recorded by the close talk microphone is annotated. The transcription included in this database is an orthographic, lexical transcription with a few details that represent audible acoustic events (speech and non speech) present in the corresponding waveform files. The extra marks contained in the transcription aid in interpreting acoustic events of the utterance. Extra marks point to mispronunciation, truncations, unintelligible words and extra noises.

A rigorous validation procedure is applied to each database produced in the project to assure a quality level.

The validation is carried out by the Validation Centre SPEX. Only those databases whose validation has been positive are considered as acceptable by the consortium. Unacceptable databases should be updated until a final quality status is achieved.

This following section describes the contents of the database and gives a short description of the recording platforms, recording environments and validation procedure.

2.1. Database contents

The content of the database includes speech utterances to train recognition systems designed for different purposes and applications. The contents of the database can be grouped as follows:

- Digits, numbers and strings of numbers
- Directory assistance names: Cities, Company names, Forenames, Surnames
- Dates and Times
- Spellings
- Phonetically rich words and sentences
- Application words

Each database contains recordings with 600 different sessions from at least 300 speakers. 400 sessions consist of the recording of 119 items and the remaining 200 sessions contain 129 items. These last 200 sessions includes 10 spontaneous sentences spoken in the car. The sentences are the answer to a specific situation that is explained to the speaker. A total of 22 different situations were described including the management of car accessories, access to databases, etc. Table 1 shows the contents of each of the 129 items each speaker has to utter. Information about spontaneous items is indicated in Table 1.

2.2. Recording platforms

Two types of recordings are found in the database. The first type consists of wideband audio signals recorded directly in the car and the second type is a GSM signal transmitted from the car and recorded simultaneously in a far-end central location. Two recording platforms were used, a 'mobile' recording platform installed inside the car and a 'fixed' recording platform located at the far-end fixed side of the GSM communications system.

The mobile platform records the signals from four high quality audio channels. For this purpose, four microphones were used: a close-talk microphone, and 3 far-talk microphones placed at different locations in the car. The positions for the far-talk microphones are:
A: on the ceiling of the car near the A-pillar
B: on the ceiling of the car behind the sunvisor that is in front of the speaker
C: on the ceiling of the car over the mid-console (near the rear-view mirror)

The mobile platform stores the recorded signals as sequences of 16 bit, 16 kHz uncompressed and multiplexed. Channels are sequentially multiplexed and the format of samples is short unsigned.

The fixed platform [Fonollosa (2000)] records simultaneously the speech utterances coming from the car through the GSM network (8 kHz sample frequency, A

law encoding). The GSM hands-free microphone is mounted on the ceiling of the car over the mid-console.

The synchronisation mode between the mobile and fixed platforms is based on the use of DTMF tones emitted from the GSM terminal placed in the car. A synchronisation and communication protocol between the two platforms is used for:
- 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 for the separation of the items into individual files.

The protocols include a series of beeps and DTMF codes transmitted by both platforms. This ensures that each recorded item is preceded by a simultaneous beep on all recording channels to allow rapid off-line synchronization of the recordings on both platforms.

Each prompted utterance is stored within a separate file. Each speech file has an accompanying ASCII SAM label file generated both by the fixed and the mobile platforms.

2.3. US SpeechDat-Car

The major difference between the American SpeechDat-Car database and the other databases is the issue of mobile/cell telephony. In order to keep consistent with other databases, it was decided to collect the data via GSM networks instead of the widespread local standards TDMA and CDMA, although the US GSM operates at 1900 MHz while European GSM operates at 900 or 1800 MHz. The GSM terminal is also different since the one selected for the European collections (the Nokia 5110) is not available in the USA. The Nokia 5190 was chosen as a "compatible one". An American "family" car (Ford Taurus) was selected for the recordings.

The textual material for the speaker prompts was adapted from British English to American English and US conditions: For instance the word "EURO" was excluded as most Americans are not aware of what it refers to and how it is pronounced. This was also done for "Beaujolais" and other typical European words. The European cities were also replaced with American cities as it is not expected that many Americans will visit Europe in their cars! So the prompt texts are not a simple adaptation of the British part of the SpeechDat-Car but constitute the design of an extra language to be incorporated within SpeechDat-Car collections.

2.4. Validation

In order to maintain a high quality level on the databases generated in the project, a validation procedure has been established [Van der Heuvel (2000)] The validation centre SPEX performs some exhaustive checks on the databases:
- Design and completeness
- Formats and Structure
- Annotation quality
- Signal quality

Validation is carried out in two steps. The first step is conducted at a very early stage of the project and is intended to avoid irremediable errors. In this moment, all of the European databases have passed this check.

The second step consists of the validation of each completed database and is carried out when each respective complete database is finished.

3. Speaker Recruitment

Because of the considerable efforts spent in instructing and installing speakers in the car only 300 speakers per language could be accomplished within the present project. Also, for efficiency reasons, each speaker records two sessions, though in two different driving environments.

Demographic constraints were defined in order to maximize the descriptive power of the population being recorded. These were as follows [Dufour (1999)]:

- a minimum of 50 speakers per dialect/ accent region
- a maximum of 10% variation from the 50-50% balance of male and female speakers
- age-distribution required is indicated in the chart below:

Age	<16	16-30	31-45	46-60	>60
	0	≥20%	≥20%	≥15%	

The constraints were to be considered as independent.

3.1. Accent regions

Because of the constraints of a minimum of 50 speakers per accent region, a maximum of six different accent regions are allowed per language.

These accent regions were chosen as merged regions of the regions defined within the SpeechDat(II) project, when possible.

The Table 2 below shows the accent regions chosen per language.

3.2. Recruitment methods

Given the demographic and recording environment constraints, speaker recruitment and recording planning requires some logistic consideration, though speaker recruitment was found less problematic than observed in the SpeechDat(II) project [Lindberg (1998)]. Although each speaker was more difficult to recruit because of the efforts and attention needed by the speaker, a lower number of speakers was required than within SpeechDat(II). Also in the present project incentives such as gifts, telephone cards, lottery tickets etc. were offered, typically around the equivalent of 25 Euro.

Speakers were recruited from various sources: recruitment through companies participating in the SpeechDat-Car project which was the main recruitment source within the project; general speaker databases containing volunteers from the SpeechDat(II) project; friends, family members and their relatives; university students and professors and their relatives; driving schools; sports clubs; public calls for participation in local newspapers or via the Internet; companies specialized in recruiting [Lindberg, 1999)].

In some cases incentives were also offered to the recruiters (i.e. the persons giving names of potential speakers). Companies kindly accepted their employees' participation in the recordings during working hours.

Language	Area/description
American English	The North Central
	Inland North
	The Midland
	The South
	The West
British English	Scotland
	Nothern England
	Wales
	The Midlands
	South West England
	East Anglia and the South East
Danish	Northern Zealand
	Southern Zealand, Lolland Falster, Bornholm
	Northern Funen, southern Funen and islands
	Northern, -western and eastern Jutland
	South, -central Jutland and southern Jutland
Finnish	Central and North Pohjanmaa, Peräpohja, South Pohjanmaa
	Savolax, South Eastern
	Häme, South Western
French	Paris and its suburban area
	Haute-Normandie, Basse-Normandie, Bretagne, Centre, Pays de la Loire
	Nord-Pas-de-Calais, Picardie, Champagne-Ardenne, Alsace, Lorraine, Bourgogne, Franche-Comté
	Auvergne, Rhône-Alpes, Languedoc-Rousilllon, Provence-Côte d'Azur, Corse
	Poitou-Charente, Limousin, Aquitaine, Midi-Pyrénées

Language	Area/description
Flemish/ Dutch	West-Vlaanderen, Zeeland
	Antwerpen, Vlaams-Brabant, Oost-Vlaanderen
	Belgisch Limburg, Nederlands Limburg, Noord-Brabant
	Groningen, Friesland, Overijssel, Drenthe
	Utrecht, Gelderland
	Noord-Holland, Zuid-Holland
German	Bremen, Hamburg, Mecklenburg-Vorpommern, Niedersachsen, Schleswig-Holstein
	Hessen, Nordrhein-Westfalen
	Baden-Württemberg, Bayern, Rheinland-Pfalz, Saarland
	Brandenburg, Berlin, Sachsen, Sachsen-Anhalt, Thüringen
Greek	Athens, South Euboea, Thessalonike, Peloponnese, Kythera, Ionian islands
	From the northern shore of the Corinthian gulf up to the northen Greek frontiers, Lefkas, Sterea Hellas, Hepeiros, Macedonia, Thessay, Thrace, North Sporades, Thasos, Lemnos, Imvros, Lesbos, Samos, Tinos
	Crete
	The Dodecanese, the Cyclades, South Sporades, Chios
Italian	Piemonte, Valdaosta, Liguria, Lombardia
	Trentino-AltoAdige, Veneto, Friuli-VeneziaGiulia, Emilia-Romagna
	Toscana, Umbria, Abruzzo, Molise, Lazio
	Basilicata, Campania, Calabria, Puglia, Sardegna, Sicilia
Spanish	Aragon, Cantabria, Castilla La Mancha, Castilla León, La Rioja, País Vasco, Madrid, Navarra, Extremadura (North)
	Andalucía, Canarias, Extremadura (South), Murcia
	Cataluña, Valencia, Baleares
	Galicia, Asturias

TABLE 2 Accent regions for the 10 languages being recorded within the SpeechDat-Car project

Since only a single car was used in most languages this also added to the logistics in particular for accent regions covering larger geographical areas. Another factor was whether the speaker were driver or co-driver. The speaker should preferably be the driver, but in two countries (Spain and Finland) this was prohibited. In Spain, both a driver and an instructor/experimenter were present.

The duration of the recording session should also be taken into account. The typical duration of a speaker recording session (two different environments) was between one and a half and two hours with no spontaneous utterances included; between two and two and a half hour if included.

4. Annotation

SpeechDatCar is characterized by having five replicas of each utterance (a close-talk microphone, three in-car microphones, and a microphone connected to GSM). Although the specifications require only annotation of the close-talk channel it is preferable to have annotation tools which allow visualizing and transcribing at the same time of all the given channels. In this way events that occur only in some of the channels can be found. In this section we will describe as an example the annotation procedure used for the German database in detail. Then we will give a short summary for other languages.

4.1. Annotation of the German database

The annotation software used is WWWSigTranscribe, an extension of the SpeechDat annotation software developed at IPSK [Draxler (1998); Draxler (1999)]. It is based on the WWW, and it features auditory output of the multi-channel in-car signals as well as the mobile phone channel, and an optional waveform display of all speech signals. To facilitate annotation, editing buttons implement often-needed tasks such as conversion from digits to the appropriate string of digit words, e.g. from "1 2 3" to "eins zwei drei", or to number words, e.g. "ein hundert dreiundzwanzig".

For the annotation, the SAM label file created by the in-car recording platform is read in and then renamed, and the prompt text is displayed in the editing field. The annotator listens to the signal output and modifies the prompt text accordingly. Then he or she enters a validation (one of "ok", "bad signal", "wrong text" or "garbage") and optionally adds a comment. After a lexical check, which ensures that the annotation is syntactically correct, it is saved to a new SAM label file under the name of the original SAM label file.

In principle, the annotation was an auditory annotation of the close talk microphone. However, in order to detect recording problems reliably, annotators were instructed to listen to all four channels for the 10 read sentence items – these items are distributed randomly across recording sessions and thus checking them would reveal technical problems. These 10 items were the first to be annotated so that when a problem was found the annotation session could be aborted early. Furthermore, they were asked to listen to the mobile phone channel for all recordings, and, in case of doubt, check the oscillogram curve.

For the annotation, some of the students who had already worked for the SpeechDat project could be recruited for SpeechDat-Car. Further students were added to the team later in the project. In total, up to 10 part time annotators with between 8 and 19 hours per week were employed. The annotation of a full recording session with the 121 mandatory items plus 10 short spontaneous items plus 9 additional language-specific items in the German data collection takes between 90 minutes for skilled to 150 minutes for inexperienced annotators.

By February 2000, 40.816 signal files of a total of 84.000 signal files had been transcribed (~ 47.7 %). Table 3 shows the number of events found in the transcriptions.

Type	Count	Percent
signal truncation	183	0.44 %
noise markers	38747	46.1 %
mispronunciations	1302	3.2%
incomprehensible speech	707	1.7%

TABLE 3 Events found in the transcriptions of the German Database

There are noise markers for dial tone, speaker noise (breathing, laughing, coughing), intermediate or stationary noise, or filled pauses. The high number of noise markers can be explained by the prompt beep that precedes every recording which was audible even in many of the close-talk microphone recordings.

4.1.1. Problems

The annotation of the SpeechDat-Car recordings faced three main problems: adaptation of the SpeechDat annotation guidelines to the SpeechDat-Car requirements, achieving annotation consistency, and software problems.

The original annotation guidelines were designed for quick annotations of telephone speech. The signal quality of the in-car recordings was on the one hand substantially better than telephone speech, but on the other hand much more variable due to the traffic situation, external noises, etc. Hence, the annotation guidelines were updated by the project consortium even after annotations had begun.

Annotation consistency was achieved by a) recruiting a number of annotators that had worked for SpeechDat, and b) asking experienced annotators to function as tutors for new annotators. As such, the veteran annotators would go through one or two recording sessions with the new annotators and monitor their annotations.

4.2. Annotation of the French database

For the French database annotation, the software JavaSgram is used. JavaSgram was developed at IRST with the objective of being flexible for independent annotation of all of the input channels, which can be visualized, zoomed, and unzoomed together in a compact graphic representation. Currently, 79 sessions have been annotated. On a general basis, there is a fairly low amount of mispronunciations and or truncations (less than 1%) while speaker noises are much more frequent (about 20% to 25 % on the average but it is strongly speaker dependent).

4.3. Annotation of the Spanish Database

Only signals from the close talk microphone have been transcribed. Very few errors due to mispronunciation, unintelligible words and truncations have been detected. This is probably due to the direct supervision of recordings in the car and to the fact that the speaker is co-driver and can be concentrated in the recording task. Annotation has been done using the software tool UPCRevBD.v1, developed at UPC [Nogueiras (1998)]. The software allows a quick and robust annotation. At the moment of writing this paper, 360 sessions have been annotated. These sessions do not include spontaneous sentences and the annotation time per session is around 30 minutes. Each person is not allowed to annotate more than two hours without rest and no more than four hours per day.

5. Distribution

The European Language Resources Association (ELRA) was established as a non-profit association in Luxembourg in February 1995, to provide a European-wide, open platform for the selection and distribution of speech, text and terminology resources to be embedded in language enabled systems, and to promote the use of Language Resources (LRs) within the Human Language Technologies sector. In order to effectively provide such services to research and development groups in academic, commercial and industrial environments, it is necessary for ELRA to address legal, logistic and other practical issues. ELRA has been granted the rights to distribute most of the speech data bases collected within the European funded projects, in particular Speechdat(M), Speechdat(II), SpeechDat-East. ELRA will be also trusted with the distribution of the databases being collected within SpeechDat-Car. As one may imagine, developing such resources is prohibitive, even for large organizations, regardless of the projected market size. Developing a SpeechDat-Car database is very expensive and time consuming. Linguistic

Resources (LR) are universally acknowledged to be critical for the development of robust, broad-coverage, and cost-effective applications for all sectors of HLT, in particular those addressing multilingual issues. It has been decided to make the databases that are produced within the SpeechDat-Car project commercially available via ELRA to third parties that agree to enter into an agreement with ELRA. ELRA will negotiate a distribution agreement with each and every data producer/owner. Third parties will have to enter into only one agreement with ELRA (unless the customer chooses to go to each individual provider and sign as many licenses as languages available playing on several different judicial systems).

The availability rule is clearly stated in the contract that the speechdat-consortium partners signed with the European Commission. It states that, as soon as the speech databases of the different languages are recorded and validated by an independent organization (SPEX, the ELRA validation Unit, see [Van den Heuvel (2000)] in these proceedings) they will be available for exploitation to all the other partners of the consortium after they have completed their own databases. All the databases will be distributed to third parties via ELRA no later than 18 months after the official end of the project. In this manner, all the collected speech data will be made available to research institutes and companies all over the world for further exploitation in research and commercial operations.

6. Final Remarks

The WWW server of the SpeechDat-Car project is hosted by the IPSK: http://www.speechdat.org/. It contains public and internal deliverables – including all public specifications –, sample recordings, images, videos and country-specific information on the SpeechDat-Car data collections in Europe.

7. References

Draxler, C., (1998); WWWSigTranscribe – A Java Extension of the WWWTranscribe Toolbox, *Proceedings LREC 1998*, Granada, Spain.

Draxler, C., (1999); WWWSigTranscribe - Annotation via the WWW, *Proceedings of MATISSE Workshop,* London.

Draxler C., R. Grudszus, S. Euler, K. Bengler, (1999) First Experiences of the German SpeechDat-Car Database Collection in Mobile Environments, *Proceedings Eurospeech 99*, Budapest.

Draxler, C., H. Van den Heuvel, H. Tropf, (1998) SpeechDat Experiences in creating Large Multilingual Speech Databases for Teleservices. *Proceedings LREC 19 98, pp 361-366.* Granada, Spain

Dufour, S. (1999) Specification of the car speech database (definition of corpus, scripts and standard), Car environments and speaker coverage, *SpeechDat-Car Technical Report, D.1.12, V.3.0.*

Fonollosa, J.A.R., A. Moreno, (2000) SpeechDat-Car Fixed Platform. *Proceedings LREC 2000.* Athens. Greece.

Lindberg, B., R. Comeyne, C. Draxler, F. Senia, (1998) Speaker Recruitment Methods and Speaker Coverage. Experiences from a Large Multilingual Speech Database Collection, *Proceedings ICSLP 1998,* Sydney, Australia.

Lindberg, B., D. Andersen, A. Bach, J. Boudy, O. Jameton, K. Laurila, R. Comeyne, A. Bonafonte, S. Michos, C. Draxler, L. Di Pasquale, K. Haslam, (1999) Speaker Recruitment and Recording Plans, *SpeechDat-Car Technical Report, D.2.4.1, V. 1.1.*

Höge, H., C. Draxler, H. Van den Heuvel, F.T. Johansen, E. Sanders, H. Tropf, (1999) SpeechDat multilingual databases for teleservices: across the finish line. *Proceedings Eurospeech 99,* Budapest.

Höge, H., H. Tropf, R. Winski, H. Van den Heuvel, R. Haeb-Umbach, K. Choukri, (1997) European speech databases for telephone applications. *Proceedings ICASSP 97*, pp. *1771-1774.* Munich, Germany

Langmann, D., H. Pfitzinger, T. Schneider, R. Grudszus, A. Fischer, M. Westphal, T. Crull, U. Jekosch, (1998) CSDC – the MoTiV car speech data collection. *Proceedings LREC 98, pp. 1107-1110.* Granada, Spain.

Nogueiras, A., A. Moreno, (1998) ; NaniBD: A set of Tools for Transcribing and Validating Speech Databases", *Proceedings of the 1st. Intl. LREC,* Granada, Spain.

Van den Heuvel , H., L. Boves, K. Choukri, S. Goddijn, E. Sanders, (2000) SLR Validation: Present State Of Affairs And Prospects *Proceedings LREC 2000.* Athens. Greece.

Van den Heuvel, H., J. Boudy, R. Comeyne, S. Euler, A. Moreno, G. Richard, (1999), The SpeechDat-Car multilingual speech databases for in-car applications: some first validation results. *Proceedings Eurospeech'99.* Budapest.