

# SpeechDat-Car Fixed Platform

José A. R. Fonollosa and Asunción Moreno

Universitat Politècnica de Catalunya  
c/ Jordi Girona 1-3, Edifici D5  
08034 Barcelona Spain  
e-mail: adrian@gps.tsc.upc.es

## Abstract

SpeechDat-Car aims to develop a set of speech databases to support training and testing of multilingual speech recognition applications in the car environment. Two types of recordings compose the database. The first type consist of wideband audio signals recorded directly in the car while the second type is composed by GSM signals transmitted from the car and recorded simultaneously in a far-end. Therefore, 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. This paper describes the fixed platform software developed by the Universitat Politècnica de Catalunya (ADA-K). This software is able to work with standard inexpensive PC cards for ISDN lines.

## 1. Introduction

The telephone server presented in this paper to automate the recording of speech databases was developed by the authors in the framework of the SpeechDat-Car EC project LE4-8334 [Moreno (2000)].

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). The European project SpeechDat-Car<sup>1</sup> aims at providing a set of uniform, coherent databases for nine European languages and for American English.

SpeechDat-Car continues the success of the SpeechDat project 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, 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.

<sup>1</sup> SpeechDat-Car started in April 1998 in the 4th EC framework under project code LE4-8334 with a 30 months' project duration.

## 2. Recording platforms

Two types of recordings compose the database. The first type consist of wideband audio signals recorded directly in the car and the second type is composed by a GSM signal transmitted from the car and recorded simultaneously in a far-end. 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: at the ceiling of the car near the A-pillar

B: at the ceiling of the car in front of the speaker behind the sunvisor

C: at the ceiling of the car over the mid-console (near the rear mirror)

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

The fixed platform records simultaneously the speech utterances coming from the car through the GSM network. The GSM phone is mounted at the ceiling of the car over the mid-console. The fixed platform is connected to an ISDN line and it records the signals directly in the received format, i.e., A law at 8000 samples per second.

The synchronization mode between the mobile and fixed platforms is based on use of DTMF tones emitted from the GSM terminal placed in car. A synchronization and communication protocol between the two platforms is used to:

- Detecting if the fixed recording platform is still alive during the recordings (and to repair a hang up);
- Allowing synchronization of the recordings on the two platforms;

- Allowing the separation of the items in individual files.

The protocols comprise a series of beeps and DTMF-codes transmitted by both platforms to ensure 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.

### 3. Fixed Platform Software (ADA-K)

The ADA-K call server for ISDN cards is written in C and is based on the COMMON-ISDN-API Version 2.0 (CAPI 2.0) standard (Appendix B). It requires a PC with Windows 95/98/NT/2000 and a CAPI 2.0 compliant ISDN PC card. ADA-K has been designed to support any number of simultaneous lines/controllers.

#### 3.1. Configuration

ADA-K uses 2 different configuration files: the parameter file and the script file.

##### 3.1.1. Parameter file

The parameter file PARAMS.DAT can be edited to modify the default parameters of the application:

- **play\_dir.** The disk directory with the prompts or beeps to be played.
- **rec\_dir.** The recording directory. Each call is recorded in a subdirectory of rec\_dir following the file system hierarchy of the SpeechDat-Car database.
- **rec\_ext.** The extension of the files to be played and recorded.
- **label\_ext.** The extension of the label files to be created
- **label\_master.** The master label file. Label files will be created as a copy of the master label file adding the items corresponding to the mnemonics SES, CCD, RED, RET, DIR, and SRC. These mnemonics must be present in the master label file with a empty item. An example of a master label file is provided with ADA-K.
- **log\_file.** File for logging ADA-K starts and stops. In this file, ADA-K logs the time and date it starts and stops. It also logs incoming calls that cannot be attended because the two B channels are in use, as well as other warnings and errors.
- **task\_file.** Script or protocol file
- **called\_number.** Called number. If you specify a called number, the corresponding ADA-K application only answers the calls directed to this number. If your ISDN BRI interface have two or more numbers assigned to it, you can have two or more ADA applications running in parallel with different configuration and task files.

##### 3.1.2. The protocol file

The protocol file CALL.DAT may be edited to modify or adjust the standard SpeechDat-Car Recording protocol.

The default protocol file shown on figure 1 conforms to the standard SpeechDat Car Synchronization Protocol defined in Appendix A

```
# Welcoming beep
PLAY BEEP
#
# Wait sesion number (4 DTMF)
# 10 is the maximum inter-digit delay in seconds
#
WDTMF      4      10
SETSPK
# Play confirmation beep
PLAY BEEP
#
# Wait prompt number (3 DTMF)
# 10 is the maximum inter-digit delay in seconds
WDTMF      3      10
#
# Record corresponding signal
# *** is the DTMF sequence to stop recording
# 10 is the maximum inter-digit delay in seconds
# V1 are the initial letters of the corresponding file name
# 600 is maximum recording time in seconds
RDTMF      ***    10    V1    600
PLAY BEEP
#
# Returns to the previous WDTMF if the received
# DTMF sequence is not "###"
RET
#
# Play a long beep if the sesion number is not new
OLDSPK      BEEP
PLAY BEEP
PLAY BEEP
PLAY BEEP
PLAY BEEP
PLAY SIL800MS
RET
```

Figure 1. Default Protocol File CALL.DAT

### 4. Utilities

In addition to the main ADA-K program, the following applications for Windows 95/98/NT/2000 are freely available from the authors.

PLAYA.EXE Plays a-law files using a PC sound card

PLAYV.EXE Plays the multiplexed sound files recorded by the mobile platform using a PC sound card. It allows to select the channel to listen.

RECORDA.EXE. Records an a-law file using a PC sound card

ALAWREAD.M. MATLAB script for reading a-law files

### 5. Final Remarks

The URL of the WWW server of the SpeechDat-Car project is <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.

## Appendix A. Synchronization Protocol

The synchronization between the mobile platform (PLTM) and the fixed recording platform (PLTF) follows the protocol described below:

PLTM: Calls to PLTF.

PLTF: Hooks off

PLTM: Sends the session code: 4 DTMFs (0000 to 9999). The first three digits correspond to the speaker code while the last one is the session digit.

PLTF: Plays a signal to acknowledge the reception of the session code.

### For each item to be recorded:

PLTM: Sends the item code. (The first character is coded in ASCII with two DTMFs while the second character is always a digit), For example the item code B7 is indicated by the three DTMF digits '667'.

PLTF: Detects item code, and starts recording

PLTM: Starts recording and sends a synchronization tone (DTMF digit '0') and a beep for time alignment.

PLTM: Detects end record (MMI stop, timeout), stops recording and sends end of item code (DTMF sequence '\*\*\*').

PLTF: Detects end of item code and stops recording.

PLTF: Plays a signal to validate the recording. If the operator of the PLTM does not receive this signal he will not validate it and the same item should be recorded again.

### At the end of the session:

PLTM: Sends the end of session code (DTMF sequence '###' and hooks on.

PLTF: Hook on.

## Appendix B. CAPI

The application was developed using the standard COMMON ISDN Application Programming Interface Version 2.0 (CAPI 2.0)<sup>2</sup>.

CAPI is an application programming interface standard used to access ISDN equipment connected to basic rate interfaces (BRI) and primary rate interfaces (PRI). By adhering to the standard, applications can make use of well-defined mechanisms for communications over ISDN lines, without being forced to adjust to the idiosyncrasies of hardware vendor implementations. ISDN equipment vendors in turn will benefit from a wealth of applications, ready to run with their equipment.

CAPI is a well-established standard. In 1989 manufacturers started to define an application interface which would be accepted in the growing ISDN market. To get an acceptable result, the focus of this standard was the possibility of running the German ISDN protocol, since an ETSI ISDN protocol standard was not available at this

time. Work on this application interface was finished in 1990 by a CAPI working group consisting of application providers, ISDN equipment manufacturers, large customers / user groups and DBP Telekom.

To reflect on the actual situation it can be stated that the international protocol specification is finished and almost every telecommunication provider offers BRI / PRI with protocols based on Q.931 / ETS 300 102. CAPI Version 2.0 was developed to support all Q.931 based protocols.

Experience in ISDN application interface design, knowledge of the market needs and a large installed base of CAPI solutions (hardware controller and applications on top of different operating systems) result in the necessity of developing a new application interface, usable in international ISDNs.

CAPI includes more than 10 years of ISDN business implementation experience in an exploding market. It covers all benefits of CAPI Version 1.1 plus new aspects of ISDN. It is based on Q.931 / ETS 300 102 but not limited to these. It simplifies the development of ISDN applications through many defaults which need not to be programmed. It keeps applications free of ISDN protocol knowledge and thus makes application development easy.

By using CAPI the international market can exploit the available experience and realise a large growth.

CAPI can be evaluated as an invaluable tool to implement powerful communications- applications. It is designed to be the base of a whole range of new protocol-stacks for networking, telephony, file-transfer, application- sharing or any other application you can think of - and even things you cannot imagine yet. The vast amount of CAPI-compliant hardware available right at this moment is probably the most persuading argument for any software-developer to base his applications on this specification. And the broad range of applications available for CAPI is a key factor for every ISDN controller-manufacturer to join in and participate with his own products. One additional argument is of course the fact, that your products will be based on an international agreed standard, set by the European Telecommunications Standards Institute (ETSI). CAPI was released as 'profil B' in prETS 300 325 Ed. 2.

For formal reasons ETSI reorganized this document, so CAPI is now embodied in European standard ETS 300 838 "Integrated Service Digital Network (ISDN); Harmonized Programmable Communication Interface (HPCI) for ISDN". CAPI is also embodied in ITU's recommendation T.200 "Programmable communication interface for terminal equipment connected to ISDN".

CAPI offers many commonly used protocols to applications without deep protocol knowledge. The default protocol is ISO 7776 (X.75 SLP), i.e. framing protocol HDLC, data link protocol ISO 7776 (X.75 SLP), and a transparent network layer.

Other supported framing layer variants are HDLC inverted, PCM (bit-transparent with byte framing) 64/56 kbit, and V.110 sync / async. COMMON-ISDN-API integrates the following data link and network layer protocols: LAPD in accordance with Q.921 for X.25 D-channel implementation, PPP (Point-to-Point protocol),

<sup>2</sup> <http://www.capi.org>

ISO 8208 (X.25 DTE-DTE), X.25 DCE, T.90NL (with compatibility to T.70NL) and T.30 (fax group 3).

CAPI can be used with the following operating systems: MS-DOS, MS-Windows 3.x, MS-Windows 95, MS-Windows NT, OS/2, Novell Netware and UNIX

### B.1 CAPI Features

- Support for basic call features, such as call set-up and clear-down
- Support for several B channels, for data and/or voice connections
- Support for several logical data link connections within a physical connection
- Possibility of selecting different services and protocols during connection set-up and on answering incoming calls
- Transparent interface for protocols above layer 3

- Support for one or more Basic Rate Interfaces as well as Primary Rate Interfaces on one or more ISDN adapters
- Support for multiple applications
- Operating-system independent messages
- Operating-system dependent exchange mechanism for optimum operating system integration
- Asynchronous event-driven mechanism, resulting in high throughput
- Well defined mechanism for manufacturer-specific extensions.

### References

Moreno, A. (2000) SPEECH DAT CAR. A Large Speech Database For Automotive Environments *Proc LREC'00*. Athens. Greece.

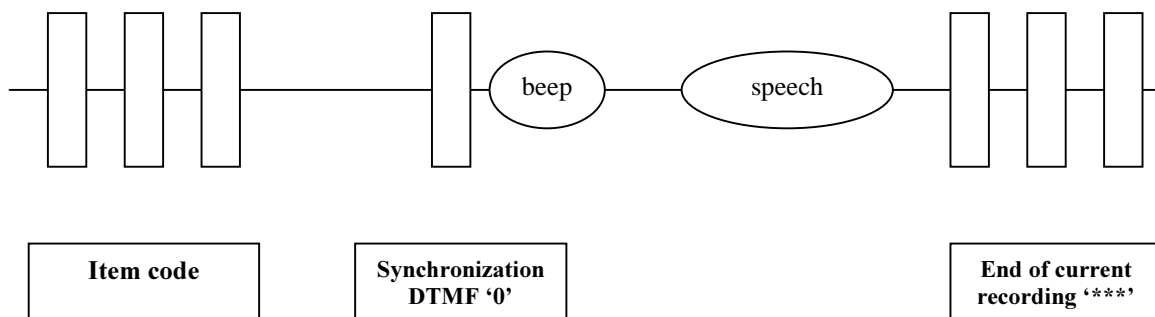


Figure 2. Scheme of the SpeechDat-Car synchronization protocol

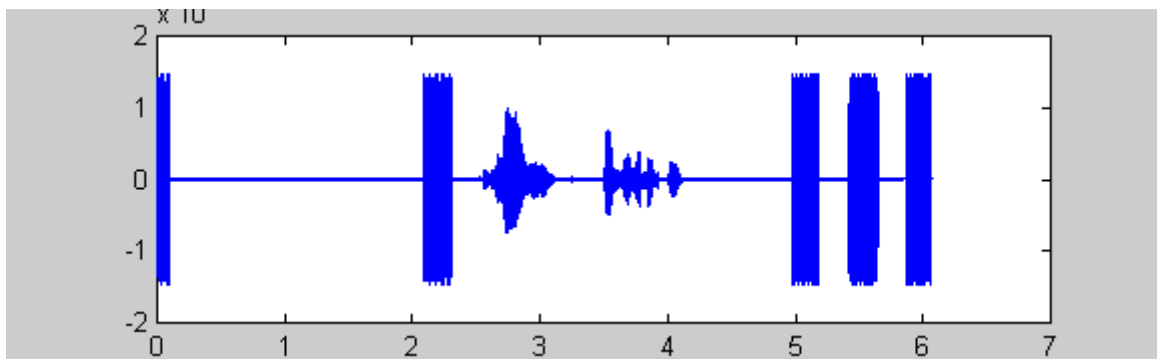


Figure 3. Example of recorded signal at the 'fixed' platform

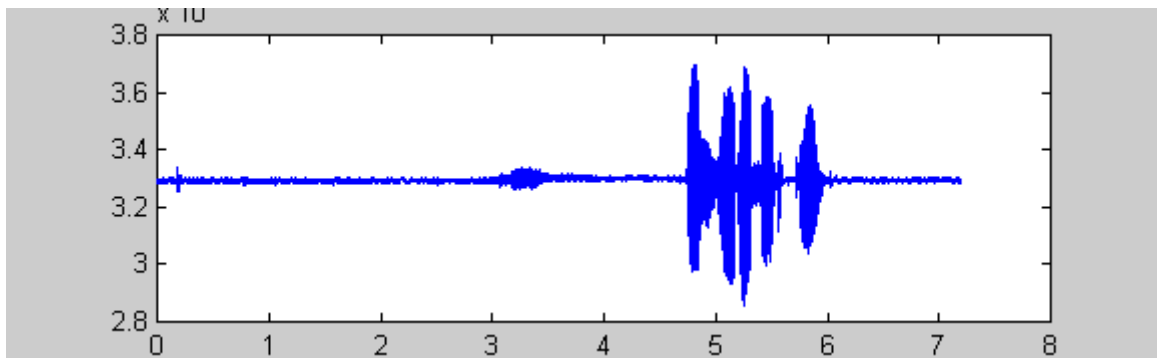


Figure 4. Example of recorded signal at the 'mobile' platform