

Automatic Database Acquisition Software for ISDN PC Cards and Analogue Boards

José A. R. Fonollosa and Asunción Moreno

Dept. of Signal Theory and Communications

Universitat Politècnica de Catalunya

c/ Jordi Girona 1-3. Building D5. E-08034 Barcelona. SPAIN

e-mail: adrian@gps.tsc.upc.es

Abstract

This paper describes an application for automatic speech-databases acquisition (ADA) developed by the authors in the framework of the EC Telematics Project SpeechDat II. The software is able to work with standard inexpensive PC cards for ISDN lines, as well as Dialogic Boards for analogue telephone lines. Both program versions share a common file format and configuration. Other important characteristics of the recording software are its simple set-up, a fast and flexible configuration of the recording session, the real-time monitoring of calls and disk space, and its proven robustness.

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 II (1996) project of the EC Telematics Applications Programme.

The Telematics Applications Programme, one of the European Commission's research programmes, is aimed at stimulating RTD on applications of information and/or communications technologies in areas of general interest.

SpeechDat is producing spoken language resources (speech databases) for use in developing teleservices such as information services (e.g. train timetable information), transaction services (e.g. home shopping and banking) and other call processing services (e.g. voice mail handling and call center systems).

SpeechDat consist of large-scale speech databases for a wide range of voice driven applications and caller environments, covering all official European Union languages and a range of dialect variations. The industrial partners will use the resources to develop a number of applications. Academic partners will also use the resources as a foundation for improving speech processing technology.

The following SpeechDat partners have adopted the ADA call server for the telephone recordings: KTH (Sweden), TELENOR (Norway), Aalborg University (Denmark), Patras University (Greece), Tampere University of Technology (Finland) and Universitat Politècnica de Catalunya (Spain).

Siemens (Germany) has also selected ADA as the acquisition software for new telephone databases as the SpeechDat (E) (East-Europe) and SALA (SpeechDat Across Latin America) (Moreno 1998).

Specifications

When the SpeechDat consortium established the common specifications of the recording platforms (Senia, 1997), it was decided to use a direct connection to an ISDN line. This connection can be 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.

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 SpeechDat consortium additionally recommended to have a recording platform with (among others), the following features:

- The recording tool must be able to manage multiple incoming calls.
- It has to establish a progress report of the recordings.
- For the recordings requiring several repetitions (e.g. for speaker verification) it has to keep track of the various repetitions and conditions.
- Prompt beep tones should be used, as they are commonly used in many automatic telephone applications. Systems without prompts beep should not use any echo-canceller to avoid any alteration in recorded speech quality.
- Speech/silence detection should be used to check the speaker reactions and speed up the overall recording time (this should reduce some fatigue effects).
- Spoken-over-the-beep detection should be used to avoid truncated speech.

- Initial silence timeout should be implemented to avoid missing items.
- The system should start-up automatically at the power-on time.

Recording Hardware

In the initial stage of the SpeechDat project, we compared different hardware configurations to configure the ISDN-based telephone server. In first place, we considered Computer Telephony (CT) boards with an ISDN interface since these boards are designed for voice applications and include an API (Application Programming Interface) for playing and recording files. However, the ISDN CT boards available at that time were expensive PRI boards that also required access to an expensive ISDN PRI.

On the other hand, the inexpensive BRI ISDN boards for PCs are primarily designed for data transmission, and none of them were available with basic voice features as voice recording and playing, voice / silence detection, DTMF detection, beep generation, or play / record synchronization.

Since several partners were interested in a new inexpensive platform for the SpeechDat recording, we finally chose a platform based on standard BRI ISDN PC boards. The additional software cost required to add basic voice functions to these cards was shared among the interested partners.

The application was first developed using the standard COMMON ISDN Application Programming Interface Version 2.0 (CAPI 2.0) described in the following section.

CAPI

COMMON-ISDN-API¹ (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. COMMON-ISDN-API Version 1.1 was a great step towards opening the national ISDN market in Germany. Meanwhile almost every German ISDN solution as well as an increasing count of international ones is based on CAPI; there exists

a well-accepted conformance test laboratory at DBP Telekom.

Currently, the international protocol specification is finished and almost every telecommunication provider offers BRI / PRI with protocols based on Q.931 / ETS 300 102. CAPI 2.0 was developed to support all Q.931 based protocols.

In fact CAPI is now an international telecommunication standard set by the European Telecommunications Standards Institute (ETSI). CAPI is embodied in draft standard prETS 300 838 "Integrated Service Digital Network (ISDN); Harmonized Programmable Communication Interface (HPCI) for ISDN", that entered the national Vote phase V9811 on January 13, 1998.

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), 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

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

Analogue Platform

The recording of telephone speech databases based on the SpeechDat specifications are now being extended to

¹ <http://www.capi.org>

East-Europe, Asia and Latin America. In these new countries digital ISDN lines may not be available and we decided to develop a new compatible version of ADA for analogue lines.

The Analogue version of ADA (ADA-D) was designed for Dialogic® boards and it is based on the Dialogic voice software for Windows 95 and Windows NT. ADA-D supports any combination of voice boards, including the inexpensive ProLine/2V for two analogue lines.

Dialogic Corporation² is the leading manufacturer of high performance, standards-based computer telephony (CT) components.

Dialogic products are used in voice, fax, data, voice recognition, speech synthesis and call centre management CT applications. The company is headquartered in Parsippany, New Jersey, with regional headquarters in Tokyo, Japan, Brussels, Belgium, Argentina and it has sales offices world-wide.

Features

The ADA recording software includes the following features:

- Available for CAPI-based ISDN PC cards and Dialogic® boards.
- Windows 95 and Windows NT Operating Systems.
- Voice/silence detector. For each sentence to be recorded you can specify the maximum duration of the initial silence, the final silence and the maximum recording time. The terminating condition can then be used to request a repetition of the recording to meet the specifications or to cancel the call.
- Automatic restart after power failure.
- A continuously updated display of the status of the lines.
- The names of the files with the prompts and the names of the files with the recorded signals can be easily specified with an ASCII script or task file.
- Two or more ADA applications can run simultaneously with different script files. The script is selected as a function of the called party number (ISDN boards) or active line (Dialogic boards).
- In addition to the above features, the Dialogic version includes a continuously updated display of the status of disk space, channel occupancy, as well as the number of attended and completed calls (Figure 1).

Configuration

We have to edit three files to configure the recording sessions: the ADA configuration file, the script or task file and a file with the number to be assigned to the first or next call. The two versions of ADA share the same file formats and configuration files.

The ADA configuration specifies:

- The directory with the prompts

```
string play_dir = "c:\ada\prompts\";
```

- The recording directory: each call is recorded in a subdirectory (named with the call number) of the recording directory. A file is also created in each subdirectory with the call starting time in seconds (ANSI-C time() function), the channel number, the called number (ISDN), the calling number (ISDN) and the time-date in ASCII (ANSI-C ctime() function)

```
string rec_dir = "c:\ada\calls\";
```

- The extension of the files to be played and recorded

```
string rec_ext = ".esa";
```

- The name of the file with the list of completed calls.

```
string completed_file = "completed.dat";
```

ADA logs the call number of all completed calls to a file. The *completed_file* is by default the file *completed.dat* in the directory specified by *rec_dir*. If it not exists, ADA creates it to include the number of the new completed calls.

- Logging file

```
string log_file = "c:\ada\ada.log";
```

In this file, ADA 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 (ISDN), missing prompt files, and other warnings or errors.

- Script or task file

```
string task_file = "c:\ada\call.dat";
```

The script file lists the names of the files to be played and recorded, as well as the recording terminating conditions. The format of this file is described in the following section

- Called number (ISDN Only): if you specify a called number, the corresponding ADA 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.
- Call number file. This file contains the call number in ASCII and can be edited to set the initial call number. The call number is then automatically updated when a new call is accepted.

The script file

The names of the files to be played and recorded are specified in order in each line of the script file. In the case of recording a file, in addition to its name we have to specify the maximum duration in seconds of the final silence and of the recorded utterance.

The script file can also include an optional *beep* file to be played before recording as well as the file to be played if no voice is detected. If in two consecutive recordings no voice is detected a specific file is played and the call is cancelled.

The entries for the files to be played follow the following format:

```
PLAY          PROMPT01
```

² <http://www.dialogic.com>

while the format to specify the name of the files to be recorded is:

```
RECORD    P01      1.4    3.5
```

where the first number specifies the final silence and the second number the maximum duration in seconds of the recorded utterance.

The optional *beep* file with the beep signal to be played before any recording is specified as:

```
BEEP      BEE
RET
```

and the optional file to be played if no voice was detected as:

```
LOUDER    PROMPT16
RET
```

If no voice is detected two consecutive recordings the message specified with the key PROBLEM is played and the call is cancelled.

```
PROBLEM    PROMPT15
END
```

Utilities

In addition to the main ADA program, the following accompanying applications has been developed and are freely available from the authors.

PLAYA	Plays a-law files using a PC sound card (Win95/NT)
RECORDA	Records an a-law file using a PC sound card (Win95/NT)
ALAWREAD	MATLAB script for reading a-law files
GOFTP	Automatic generation of scripts to backup or transfer completed calls using the FTP internet protocol.

Conclusions and future work

We have described ADA-I, a voice server for telephone recordings using inexpensive PC cards for ISDN lines. ADA-I has been successfully used by several partners of the SpeechDat project for recording more than 20,000 speakers and one million of files.

We have also presented ADA-D a new compatible version for analogue lines of special interest for countries without public digital telephone networks. In the future ADA will be also adapted to support specific recordings as those required in the SpeechDat-Car (1998) project.

ADA-I is also being extended to collect databases with human-human dialogues as well as interactions between a human and a simulated computer system using the Wizard of Oz technique where human subjects are led to believe that they are conversing with a computer

Acknowledgements

This research was supported by the CICYT (Project TIC95-1022-C05-03) and the European Commission (SpeechDat LE2-4001).

References

- Moreno, A., Hoege, H., Koehler, J., Mariño, J., (1998). SpeechDat Across Latin America. *Proc. First LREC Conference, this issue.*
- Senia, F., Chatzi, I. et al. (1997). Installation of the recording device and documentation. SpeechDat Technical Report. LE2-4001.SD2.1. June 1997.
- SpeechDat II (1996). *Speech Databases for the creation of voice driven teleservices.* EC Telematics (LE2-4001). Language Engineering Resources.
<http://www.phonetik.uni-muenchen.de/SpeechDat.html>
- SpeechDat-Car (1998). *Speech Databases for Voice Driven Teleservices and Control in Automotive Environments.* EC Telematics. Language Engineering Resources.
<http://www2.echo.lu/langeng/en/res.html>.

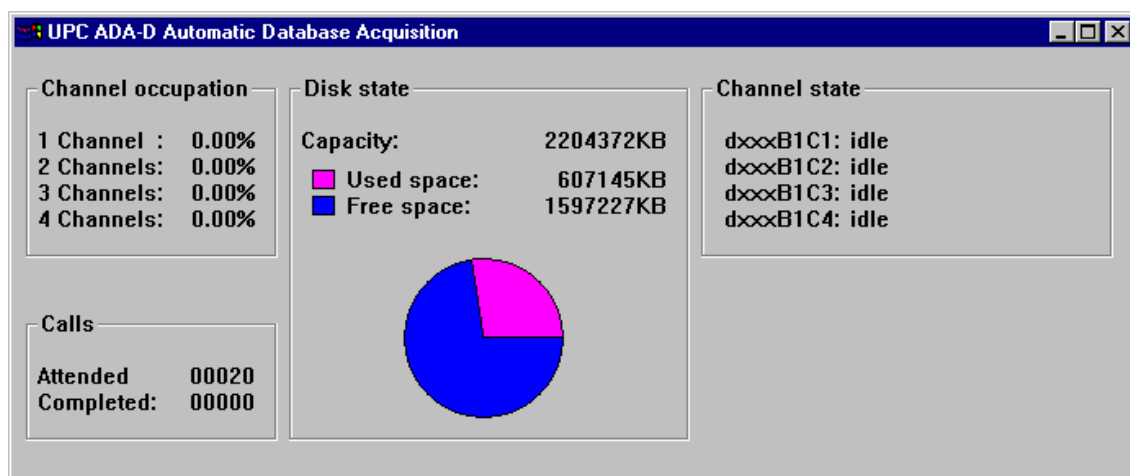


Figure 1. ADA-D Status Windows.