THE ANDOR INTERFACE TO THE AUSTRALIAN NATIONAL DATABASE OF SPOKEN LANGUAGE

J.Bruce Millar and Dave Davies

ANDOSL Project
Computer Sciences Laboratory
Research School of Information Sciences and Engineering
Australian National University

ABSTRACT

This paper introduces the ANDOR user interface, which allows remote access using both special purpose and standard SQL commands, to a descriptive database which describes the current spoken language data holdings of the ANDOSL Project. It also updates the description of the data description format currently used within the ANDOSL project and indicates how it is mounted in an ORACLE system on the ANDOSL node connected to AARNET at the ANU. ANDOR attempts to do several things; give limited access to all-comers on AARNET; allow access to the ORACLE database via e-mail to registered users; provide tutorial help and personal SQL command library support; provide short-cut report generation directly from the standard data description format thus supporting the specific queries expected from the speech science and technology community in an efficient manner.

The facilities provided using these various modes of access are described, examples of the scope of the descriptive data are given, and the status of current data holdings in the ANDOSL system are reported.

INTRODUCTION

The Australian National Database Of Spoken Language (ANDOSL) Project is a cooperative database development project sponsored by the National Spoken Language Database (NSLD) committee of the Australian Speech Science and Technology Association, and implemented by four institutions (Sydney University, the National Acoustic Laboratories, Macquarie University, and the Australian National University) with funding support from the Australian Research Council. Part of the role of Australian National University in this project has been to establish a standard data description format for spoken language data, to implement it for the growing data corpus of the ANDOSL project, and to give access to this a descriptive database to the speech research community over AARNET. This paper focuses especially on the means of giving access to the speech research community to this descriptive scheme and on giving examples of how it may be used to enquire about data holdings.

First of all we review the nature of the descriptive scheme itself, before introducing the structure of the ANDOR (ANDOSL-ORACLE) interface, the access facilities that it offers and the scope of the descriptive data that is stored relative to the holdings of speech signal data.

THE NSLD DESCRIPTIVE SCHEME

The National Spoken Language Description scheme was first described by Millar (1992) prior to its first major application to real data within the ANDOSL project. This description was based on discussions of the components of scheme with a wide range of experts in speech data management who are linked to the Coordinating Committee for Speech Databases and Assessment (COCOSDA) which has a working group on Speech Corpora and Labelling. This consultation is an ongoing process. As a result of this exposure both to real data and to worldwide expertise some small changes have been made and more are anticipated.

The current status of the scheme is summarised in table 1. A spoken language record is described under ten headings which answer specific questions about the data. The descriptive scheme is very rich

providing scope for much greater detail of description than that normally found in headers on data files. It is modelled to some degree on the 'associated description file' of the European ESPRIT SAM project. It exists as a rich archival descriptive scheme which can transmit or receive information to or from any of the major formats in which such information is normally held, as indicated graphically in Figure 1. Of specific interest in this paper is the link, noted in that figure, to Data Base Management System tables.

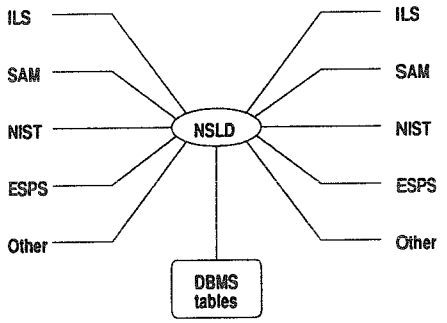

Figure 1: The NSLD scheme links existing formats and DBMS

Significant areas of the NSLD format descriptive data have been mounted as an ORACLE database on the ANDOSL node connected to AARNET at the ANU. This allows queries to be made using standard SQL commands. Some slight changes have been made to the structure of the NSLD scheme to make it fit the relational data structure more efficiently. For the most part the 'descriptive domains' of table 1 translate directly into 'tables' in the relational database. However in the area of the description of speakers there exists a very variable area of description relating to the history of the speaker along a number of dimensions each of which have potential influence on their current speech. Hence a specific speaker history table has been introduced which enables all events which could influence a speaker's speech to be lodged in the database in categories such as medical, educational, lifestyle, occupational, residential, and linguistic.

The ANDOR interface has been provided to give access, that is as universally useable as possible, to this potentially very rich descriptive resource.

THE ANDOR INTERFACE

The ANDOR interface attempts to do several things; it gives access to introductory information from any computer system that is capable of sending and receiving "e-mail" on AARNET; it allows access to the ORACLE database via e-mail to registered users; It provides tutorial help and personal SQL command library support; and it provides short-cut report generation directly from the NSLD formatted data thus supporting the specific queries expected from the speech science and technology community in an efficient manner.

ANDOR provides access to both information about the ANDOSL data corpus and to commands that can interrogate a structured description of the data itself.

The information about ANDOSL covers three areas: (1) documentation giving background on the history, principles, and status of the ANDOSL database; (2) documentation on the structure of the descriptive data held in NSLD or DBMS format; and (3) documentation on the structure of the commands that are available for interrogation.

Interrogation commands are stored in two areas: (1) a public 'system' directory [group of commands] which contains interrogations about the system structure and overall contents of the corpus; and (2) a private 'user' directory [group of commands] which contains interrogations constructed by the user for his/her own purposes.

DESCRIPTIVE DOMAIN	SPECIFIC DESCRIPTORS	QUESTION ANSWERED
SOURCE	Date, time, place of recording Pointer to original data Current segmentation of original	Where did it come from?
ENVIRONMENT	Transducer type / placement Speaker posture / position Local reflective surfaces Ambient acoustics: active / passive Analogue signal conditioning	How was it captured?
TASK	Instructions given to speaker Parametric variables - prompts	Why was it spoken?
SPEAKER	Sex, age, physical attributes Language training and experience Speech quality, fluency, intelligibility Education, Occupation, Residence ENT / thoracic surgery, Hearing status Current medical treatment Lifestyle - smoking, sport, etc Voice training / stress Music training - instrument / voice Parental language, occupation, origin	Who spoke it?
TRANSFORM	Specific transforms since original	What's happened to it?
LOCATION	Corpus name Type of media Physical place of media Logical path to data file	Where can I find it?
FORMAT	Header length / type Channel interleaving Digital data format	How can I read it?
DERIVED	Annotation analysis (multi-level) Feature analysis (acoustic-phonetic) Signal quality checks	What value-added data?
ANOMALIES	Deviation of data from expectations	What happened?
ACKNOWLEDGEMENT	Form of acknowledgement requested Form of any usage constraints	Who is responsible?

Table 1: summary of the NSLD scheme (as at 1 August 1994)

ACCESS FACILITIES

The access facilities to the descriptive database are designed to lead the naive user through a gentle introduction to the system making minimal assumptions about prior knowledge of database systems. The gentle introduction introduces the concepts of descriptive files storing information about the system and how to use it, and a set of simple commands that provide shorthand access to pre-ordained queries deemed to be of wide interest.

The gentle introduction is accessed by sending 'e-mail' to the ANDOSL node "andosl@andosl.anu.edu.au". A message body comprising the words 'send help' generates a short message giving basic information to orient the new user to the ANDOR system, and instructions about where to find more specific information. This information includes instructions for short-hand requests to access general summary information, documentation on the structure of the DBMS system and how to access it with simple SQL commands, and how to store user-specific SQL commands in a personal command library. Any interested person is encouraged to probe the system regularly with a "send help" as updates to the documentation and the data will be notified in the system response.

General summary information is accessed by a number of pre-constructed commands (stored in the system command area) which display general information about the structure of the database and the NSLD data structures that form its basis (see table 2 for examples). The final example in table 2 is a special case of the generic form 'keys/<table_name>' which will list the field names and definitions of all the descriptive fields within the named table. Figure 2 illustrates the operation of one of these. Other short-hand commands which access the NSLD format descriptors directly are exemplified in figures 3 and 4.

Command	Action	Example
"sum speaker"	summarise speaker information according to key measures	figure 4
"sum quality"	summarise the quality of specified data files	figure 3
"keys quality"	list the fields which exist in the quality domain	figure 2

Table 2: Examples of short-cut system commands

General purpose access

General purpose access to the descriptive data held in the DBMS is via SQL (Structured Query Language) commands. There are many SQL commands which may be applied to the data and the ANDOR interface allows access to 'help' information derived from the ORACLE system itself. However there are two basic commands, variants of which will cover the needs of most users of the ANDOR interface. These commands are SELECT and START which respectively select information from the database and start up stored interrogation commands.

ACKNOWLEDGEMENT	ANOMALY	ANOMALY_CODES	DDF
DERIVED	ENVIRONMENT	FORMAT	FIELDS
LOCATION	QUALITY	SOURCE	SPEAKER
SPEAKER_HISTORY	TAPE_INDEX	TEXT	TRANSFORM

Table 3: Names of tables currently implemented in the ANDOSL DBMS

The SELECT command has a simple syntax which enables the user to extract the values of a list of 'fields' from a specified 'table' when certain conditions are met.

```
select <list of fields required> from <table> [where <field> <op> <value>]
```

The 'tables' in ANDOSL are given in table 3. Each table has a number of fields which contain the specific descriptors of each descriptive domain. The "<op>" operator has a wide range of options which are

described in the online documentation. In general the descriptive domains of the NSLD scheme (table 1) map onto the ORACLE 'tables', and the key descriptors within each descriptive domain map onto 'fields' with those tables.

User library

A further function of ANDOR is to assist in managing user libraries of SQL commands that extract data from the database. Such commands, which can become quite verbose, can be saved in a user command group for later use under the control of the START command, or perhaps more usefully as building blocks for more sophisticated commands. A truncated history of all the commands that a user has performed can be reported by ANDOR. The user can select individual commands from this history to re-submit, edit, or to save as a named command in their personal library.

SCOPE OF THE DESCRIPTIVE DATA

At present descriptors of the majority of the data collected up until mid-1994 are included in the system. Some 53,612 signal data files are referenced and 176 speaker records are included. Most of the data files have been subjected to a quality check and the results of those checks are stored in 53,318 quality records. In addition to quality checks there are 209 recorded anomalies that occurred in the recording process. The speaker records vary in their richness but for the 176 speakers there are 2758 speaker history events recorded which give insight into special factors that may have influenced their speech. There are single records describing the recording environment, the format of the digital data, and the location of the digital data. The source, transform, and derived file tables have yet to be implemented.

As an example of the descriptive data we illustrate the signal quality descriptors. Commands to display the NSLD descriptive fields are stored in the 'keys' directory. Figure 2 shows the result of the relevant command.

```
Command: keys quality

Result:  OML <overall_mean_level>            mean displacement value
         NPL <number_of_positive_limits>     measure of positive overload
         NNL <number_of_negative_limits>     measure of negative overload
         MED <mean_segmental_deviation>      mean s.d. in 20ms intervals
         MID <minimum_segmental_deviation>   minimum s.d. in 20ms intervals
         MAD <maximum_segmental_deviation>   maximum s.d. in 20ms intervals
```

Figure 2. Descriptive keys for 'quality' in NSLD format

These NSLD 'quality' keys plus the data file name (QUANUM) and signal peak data from the speech file headers (MAX, MIN) comprise the fields in the Oracle 'quality' table. The quality information can be accessed through the standard SQL commands or using a shorthand command which operates directly on the NSLD format description files. An actual example of the latter (Figure 3) requests a summary of the quality of all the data from speaker No.4, and shows that minor clipping was found in three sentence (type=s) files, and that consistent energy levels and high signal-to-noise ratios were found throughout.

```
Command:        sum quality S004.qua

Result:         Minimum clipping level = 0

                *** Speaker S004 ***
                File: S004s136.qua reports 3 clips
                File: S004s166.qua reports 17 clips
                File: S004s193.qua reports 3 clips

                Number of files processed = 278
```

```
Type    Files   Clips   Clips per file
s       3       23      7.66667
Total:  3       23      7.66667

Average Max Power: 76.9 +- 2.6 db
Average Med Power: 60.8 +- 3.5 db
Average Min Power: 21.0 +- 1.6 db
Average Max - Min: 55.8 +- 3.1 db
Average Med - Min: 39.8 +- 3.9 db
```

Figure 3. Tabulating the quality of a speaker's data files.

STATUS OF CURRENT DATA HOLDINGS

Speech signal data for each component of the ANDOSL core material set (Millar et al., 1994) is stored for the 204 speakers categorised in figure 4. In addition to categorisation into gender and age range, the native Australian speakers are categorised into cultivated, general, and broad, whereas the accented speakers are currently bundled together in the response of this summary command.

```
Command: sum speaker total

Numbers

        Cultivated      General         Broad           Accented
        f       m       f       m       f       m       f       m
18-30   4       4       9       9       4       6       11      9
31-44   6       7       9       8       7       5       9       8
 45+    10      6       9       7       6       7       9       15
          C = 37          G = 51          B = 35          A = 61

              Females = 101       Males = 103       Total = 204
```

Figure 4. Categorisation of speakers on age range, gender, accent.

CONCLUSIONS

The ANDOR interface to descriptive data about a speech corpus is a novel development which should enable both widespread remote access and in-depth access to the data that is available. The discerning speech data user may have many questions answered before becoming involved in accessing the data itself, with all the concerns of storage space and file formats that normally surrounds that operation. It has been developed as an in-house tool within the ANDOSL project but with the definite plan to make it a part of the proceeds of the ANDOSL project for the benefit of the whole research community.

ACKNOWLEDGEMENT

This work has been supported by a grant from the Australian Research Council to the Australian National Database Of Spoken Language (ANDOSL) project.

REFERENCES

Millar,J.B. (1992) The description of spoken language, In "Proceedings of 4th Australian International Conference on Speech Science and Technology", Brisbane, 1-3 December, pp.80-85.

Millar,J.B., Vonwiller,J.P, Harrington,J.M., Dermody,P.J. (1994) The Australian National Database Of Spoken Language, In "Proceedings of ICASSP'94", Adelaide, 19-22 April, Vol.1, pp.97-100.