

# SPEECHDAT ENGLISH DATABASE

FOR THE MOBILE TELEPHONE NETWORK

Author(s):	Louise Helliker / Marek Pawlewski
Institute:	BT Labs
Address	Martlesham Heath, Ipswich, Suffolk, IP7 3RE, UK
email:	louise@saltfarm.bt.co.uk / mark@saltfarm.bt.co.uk
Date:	May 1997
Version:	pre-validation DRAFT

# CONTENTS

<b>1. INTRODUCTION</b>	<b>4</b>
1.1 Speech file format	4
1.2 File nomenclature	4
1.3 Directory structure	5
1.4 Label files	6
<b>2. DATABASE DESIGN AND COLLECTION</b>	<b>9</b>
2.1 Recording site and platform	9
2.2 Speaker recruitment	9
2.3 Design of prompting and prompt-sheet	9
2.4 Transcription	9
<b>3. DATABASE CONTENTS DEFINITION</b>	<b>9</b>
3.1 Application words	10
3.2 Isolated digits	10
3.2.1 Single digit	11
3.3 Connected digits	11
3.3.1 Telephone number	11
3.3.2 Credit card number	11
3.3.3 PIN code	11
3.4 Dates	11
3.5 Embedded application word	12
3.6 Spelled names/words	12
3.7 Money amount	12
3.8 Natural number	12
3.9 Directory assistance names	12
3.9.1 Spontaneous forename	12
3.9.2 Spontaneous city name	12
3.9.3 City name (set of 500)	12
3.9.4 Company/agency name (set of 500)	13
3.9.5 Forename & surname (set of 150)	13
3.10 Yes/No questions	13
3.11 Phonetically rich sentences	13
3.12 Times	13

<b>3.13 Phonetically rich words</b>	<b>13</b>
<b>3.14 Any other additional material</b>	<b>13</b>
<b>3.15 Links to other databases</b>	<b>13</b>
<b>4. SPEAKER DEMOGRAPHIC INFORMATION</b>	<b>13</b>
<b>4.1 Accent/Regions</b>	<b>13</b>
<b>4.2 Speaker characteristics</b>	<b>15</b>
<b>5. THE LEXICON</b>	<b>15</b>
<b>6. RECORDING CONDITIONS</b>	<b>15</b>
<b>6.1 Environments</b>	<b>15</b>
<b>6.2 Network called from</b>	<b>15</b>
<b>7. DEVIATIONS FROM SPEECHDAT SPECIFICATIONS</b>	<b>15</b>
<b>8. SAMPLE PROMPT SHEETS</b>	<b>15</b>

## 1. Introduction

The SpeechDat(II) project is sponsored by the EC under contract number LE2-4001.

The British English database comprises telephone recordings from 1000 speakers recorded over the GSM digital mobile network, using an E-1 ISDN interface at the recording site. It was produced by BT Labs in Suffolk, England.

The database is available on \*\*\* CD-ROM discs in ISO 9660 format. The CD-ROM volumes are structured as follows:

\*\*\*

The precise details of the distribution discs and directories are contained in the README.TXT file stored on each CD-ROM. Further details regarding the database contents, files and directories are provided in the documentation files in the DOC, TABLE, INDEX and PROMPT directories.

### 1.1 Speech file format

It has been agreed to follow the ESPRIT Project SAM standards for speech file storage. Speech files are stored simply as sequences of 8-bit 8-KHz A-law speech samples. Each prompted utterance is stored in a separate file. Speech signal files have no header; each signal file is accompanied by an ASCII SAM label file which contains the relevant descriptive information.

### 1.2 File nomenclature

File names follow the ISO 9660 file name conventions (8 plus 3 characters) according to the main CD-ROM standard. The following template is used:

DD NNNN CC. LL F

where:

DD	Database identification code (00-ZZ) For SpeechDat: A1 = fixed network recordings; B1 = mobile network recordings; C1 = speaker verification database
NNNN	Recording session progressive number (0000-9999)
CC	Corpus code (A1-Z9) obtained by collating the corpus and the item identifiers
LL	Two letter ISO 639 language code: EN for English
F	File type code O = Orthographic label file, A = A-law coded speech file

**Table <\$ code> - SpeechDat filename convention**

As it is useful for users to clearly identify the speech file contents by looking at the filename, we have specified the corpus code to support a one letter corpus identifier and a one letter item identifier, by the following table. All items are read, unless marked as spontaneous.

Corpus identifier	Item identifier		Corpus contents
A	1-3	1-6	3 application words (6 for MDB and FDB collections with a small number of speakers - corpus codes from A1 to A6)
B	1		1 sequence of 10 isolated digits
C	1		1 sheet number (5+ digits)
C	2		1 telephone number (9-11 digits)
C	3		1 credit card number (14-16 digits)
C	4		1 PIN code (6 digits) (set of 150 SDB codes)
D	1		1 spontaneous date, e.g. birthday
D	2		1 prompted date, word style
D	3		1 relative and general date exp.
E	1		1 word spotting phrase using an application word (embedded)
I	1		1 isolated digit (2 for MDB with codes I1 - I2)
L	1		1 spontaneous, e.g. own forename
L	2		1 spelling of direct. city name
L	3		1 real/artificial for coverage
M	1		1 currency money amount
N	1		1 natural number
O	1		1 spontaneous, e.g. own forename
O	2		1 city of birth / growing up (spont)
O	3	3-4	1 most frequent cities (set of 500) (2 for 250 speakers MDB with codes O3 and O4 - set of 25)
O	5	5-6	1 most frequent company/agency (set of 500) (2 for 250 speakers MDB with codes O5 and O6 - set of 25)
O	7		1 "forename surname" (set of 150 SDB "full" names) (set of 25 for 1000 speakers MDB)
Q	1		1 predominantly "yes" question
Q	2		1 predominantly "no" question
S	1-9		9 phonetically rich sentences
T	1		1 time of day (spontaneous)
T	2		1 time phrase (word style)
W	1-4		4 phonetically rich words

**Table 1 - FDB/MDB corpus codes**

The proposed format uses mnemonic values. It permits selection of all files belonging to one of the twelve corpora by using one command (e.g. in DOS "dir /s/b ??????C\*", in UNIX "find . -name "??????C\*" -print").

### 1.3 Directory structure

The directory structure uses a shallow directory nesting with contiguous numbers to identify the individual sub-directories and call directories. The following three-levels directory structure is defined:

\<database>\<block>\<session>

Where:

<database>	Defined as: <name><#><language code> i.e. MOBIL1EN Where: <name> is MOBIL indicating a mobile network database <#> is 1 for SpeechDat <LL> is the ISO 2-letters code EN for English
<block >	Defined as: BLOCK<nn> where <nn> is a progressive number from 00 to max. 99 These numbers are the same as the first 2 digits used in <nnnn> described below.
<session >	Defined as: SES<nnnn> Where <nnnn> is a progressive number in the range 0000 to max. 9999, being the numeric call identification number also encoded in each filename. As there are no more than 50 utterances per call, the total number of speech files and associated transcription files does not exceed the CD-ROM recommended limit of approximately 100 files in a directory.

**Table <\$ code> - SpeechDat directory structure**

Both signal files and label files are put in the same directory.

All sessions have complete recordings for all prompted items with the following exceptions:

[#list of missing files; sorted by item and/or session]

In addition to the previous structure the following directories are used to store some other files:

\<database>\DOC	documentation files, including subword occurrence files
\<database>\TABLE	speaker and lexicon tables
\<database>\INDEX	index files - contents file
\<database>\PROMPT	prompt sheet tables
\<database>\SOURCE	source code for SAMLIB DOS/Unix file access routines

**Table <\$ code> - Non-speech related directory structure**

Finally the root directory contains three files:

- a “README.TXT” ASCII file describing all files in the CD-ROM; signal and label files are reported by specifying their templates;
- a “DISK.ID” ASCII file containing the volume name (11 characters long); it supplies the volume label to UNIX systems that are unable to read the physical volume label, e.g. “MOBIL1EN\_01”
- a “COPYRIGH.TXT” ASCII file to protect the authors’ rights.

All these support files are duplicated on each CD-ROM.

## 1.4 Label files

SAM format ASCII label files are used.

See the file SD131v40.doc in the \DOC directory for a description of the label files. An example label file is given below:

LHD: SAM,5.10  
DBN: SpeechDat\_English\_Mobile\_Network  
VOL: MOBIL1EN\_01  
SES: 1011  
CMT: \*\*\* Speech file information \*\*\*  
DIR: \MOBIL1EN\BLOCK10\SES1011  
SRC: B11011A1.ENA  
CCD: A1  
CRP:  
BEG: 0  
END: 39999  
REP: BT MARTLESHAM IPSWICH UK  
RED: 31/Jan/1997  
RET: 11:44:17  
CMT: \*\*\* Speech data coding \*\*\*  
SAM: 8000  
SNB: 1  
SBF:  
SSB: 8  
QNT: A-LAW  
CMT: \*\*\* Speaker information \*\*\*  
SCD: 0951823476  
SEX: M  
AGE: 38  
ACC: NORTH\_EAST\_ENGLAND  
CMT: \*\*\* Recording conditions \*\*\*  
REG: NORTH\_EAST\_ENGLAND  
ENV: STREET  
NET: GSM  
PHM: MOBILE  
LBD:  
CMT: \*\*\* Label file body \*\*\*  
CMT: LBR, begin, end, input gain, min, max, prompt  
CMT: LBO, begin, centre, end, transcription  
CMT: \*\*\*\*\*  
LBR: 0, 39999,,,, English  
EXT:  
LBO: 0,, 39999, English  
EXT:  
ELF:

## **2. Database design and collection**

### **2.1 Recording site and platform**

The recordings were made at the BT Labs via an ISDN-30 circuit with DASS-2 signalling.

The recording platforms were Pentium PCs running Consensus UNIX. An Aculab E1/PRI 30 channel ISDN card and Aculab speech card were used. Maximum recording durations were set for each item according to expected durations. Four recording channels were provided for the English MDB and 4 for the Welsh fixed database, which was recorded during the same period.

### **2.2 Speaker recruitment**

A market research company recruited speakers from 15 dialect areas. The number of speakers recruited from each area was determined according to density of population, with London and the South East of England (taken as one area) contributing 25% of the calls (see section 4.1 for more details).

### **2.3 Design of prompting and prompt-sheet**

The prompt sheets were designed to be easy to read and understand. A table was used to make it more comfortable for speakers making a call from a moving vehicle.

A set of 7 spontaneous questions was included in the database: date, time, place name, proper name, spelt proper name and 2 yes/no questions. These questions were spread across the sheet.

### **2.4 Transcription**

Annotation was performed by a native English speaker, with a degree in linguistics and a background in speech recognition. Annotation followed the SpeechDat project conventions.

The VOX! annotation software was used, under Windows 95 on a Pentium PC. VOX! is marketed by Eikon for CSELT in Italy.

The ISO-8859-1 character set was used. See file ISO8859.ps in the \DOC directory.

Transcription conventions comply with SD1.3.2, which is found in the \DOC directory.

## **3. Database contents definition**

The final specification for the English recordings is as follows:

(Items are read unless otherwise stated.)

The speaker's full name (spontaneous).  
1 telephone number.  
1 credit card number.  
1 6-digit string.  
1 money amount.  
1 question expecting 'yes' (spontaneous).  
1 question expecting 'no' (spontaneous).  
1 date (spontaneous).

- 1 date (read).
- 1 time (spontaneous).
- 1 time (read).
- 1 natural number.
- 6 command words.
- 1 extra application word (optional).
- 1 sentence included embedded application word.
- 1 place name (spontaneous).
- 1 place name (read).
- 1 company/organisation name.
- 1 forename and surname combination.
- 1 name (spontaneous).
- 1 spelt name (spontaneous).
- 1 spelt word.
- 1 spelt place name.
- 4 phonetically rich words.
- 9 phonetically rich sentences.
- 1 10-digit string (paused).
- 2 isolated digits.

### **3.1 Application words**

A set of 30 application words has been selected. Each speaker says 6 different SpeechDat application words, plus one extra from the set "short/medium/long". The applications words are listed below:

English  
 Terminate  
 Menu  
 Help  
 Cancel  
 Stop  
 Continue  
 Repeat  
 Operator  
 Call  
 Dial  
 Redial  
 Directory  
 List  
 Previous  
 Next  
 End  
 Add  
 Change  
 Delete  
 Save  
 Play  
 Record  
 Send  
 Program  
 Remove  
 Forward  
 File  
 Read  
 Reply

### **3.2 Isolated digits**

Each speaker records one string of 10 “isolated” digits.

### **3.2.1 Single digit**

"0" may be pronounced "zero", "oh" or "nought" in English (and, very rarely, "nothing"). Therefore "0" occurs three times in the file which is used to generate the digits for the prompt sheets, so that there are sufficient recordings of the main pronunciations. Each speaker records two single digits.

## **3.3 Connected digits**

### **3.3.1 Telephone number**

The telephone number is printed in the standard format, with brackets around the code, to encourage natural pronunciation as a phone number. However, no division has been made of the 6-digit numbers into 2 groups of 3 or 3 groups of 2. This is so that speakers can say the number in the way that feels most natural for them. A range of different number types are included.

### **3.3.2 Credit card number**

A 16 digit credit card number: 4 groups of 4 digits. These were provided by the English SDB project (speaker verification database).

### **3.3.3 PIN code**

The set of 6-digit PIN codes provided by the English SDB project was used.

## **3.4 Dates**

The caller is prompted for three dates:

- i) Date of birth or other familiar date (spontaneous)
- ii) 1 date (read, word style)

Dates have two standard formats:

Thursday 21st April 1994 or Thursday April 21st 1994.

Most speakers say "Thursday the 21st of April 1994" or "Thursday April the 21st 1994" (although “the” and “of” are never written in dates in English).

- iii) Relative and general date expressions

Each speaker records one of the following:

Today  
Yesterday  
Tomorrow  
The day before yesterday  
Last Monday  
Last Tuesday  
Last Wednesday  
Last Thursday  
Last Friday  
Last Saturday  
Last Sunday

Next Monday  
Next Tuesday  
Next Wednesday  
Next Thursday  
Next Friday  
Next Saturday  
Next Sunday  
Next week  
Last week  
Christmas Day  
Christmas Eve  
Boxing Day  
New Year's Eve  
New Year's Day  
Ash Wednesday  
Maundy Thursday  
Good Friday  
Easter Sunday  
Easter Monday  
May Day

### **3.5 Embedded application word**

Five phrases were constructed for each application word. They reflect likely usage in real applications and reflect experience within the BT Labs Dialogues Team.

### **3.6 Spelled names/words**

Each speaker spells a name which may be their own or another familiar name (which they have just said), a word (or random alpha-numeric sequence) from the sheet and a place name from the sheet (the place name which they said a few items earlier).

[# include table with full frequency counts of for each letter ]

### **3.7 Money amount**

Each speaker says one money amount from a set which includes a variety of pence-only, pounds-only and pounds-and-pence amounts, varying from 1 p to about £940.

### **3.8 Natural number**

Each speaker records 1 natural number from a randomly generated set ranging from 39 to 10,434.

### **3.9 Directory assistance names**

#### **3.9.1 Spontaneous forename**

Each speaker says their own name, or another name familiar to them. They are given the option of not saying their own name, as some people may feel sensitive about this.

#### **3.9.2 Spontaneous city name**

The speaker's place of birth.

#### **3.9.3 City name (set of 500)**

A list of 500 British city, town and village names was used.

#### **3.9.4 Company/agency name (set of 500)**

The recommended list of 500 was used, although it was modified to exclude those difficult for British people to pronounce and to include companies and organisations well known in Britain.

#### **3.9.5 Forename & surname (set of 150)**

The set provided by the English SDB was used.

#### **3.10 Yes/No questions**

Each speaker answers 1 fuzzy 'yes' and 1 fuzzy 'no' question. There is a set of 5 questions which expect fuzzy 'yes' and 5 which expect fuzzy 'no'.

#### **3.11 Phonetically rich sentences**

A large database containing 2565 different sentences has been designed. Each speaker records a set of nine sentences.

[# include table of phone frequencies; also statistics on diphones and triphones if available]

[# did every caller utter each phone at least once? (this is recommended)]

#### **3.12 Times**

The speaker is prompted for two times:

- i) The current time of day (spontaneous).
- ii) 1 analogue time phrase (read), from a set which includes a variety of different styles.

#### **3.13 Phonetically rich words**

Each speaker also records 4 phonetically rich words from a set of 1224 words.

[# include table of phone frequencies; also statistics on diphones and triphones if available]

#### **3.14 Any other additional material**

#### **3.15 Links to other databases**

[# state explicitly links to other databases (FDB, MDB and SDB) here ]

### **4. Speaker demographic information**

#### **4.1 Accent/Regions**

The regions represent the major accent areas of the United Kingdom.

There is much pronunciation variation within each area, particularly within the large Scottish areas, but a reasonable compromise has been made.

The East Anglian region has its own vowel sounds and intonation, distinct from the South Eastern accent, with which it is often grouped.

The London and the South East area covers the Home Counties and several other neighbouring areas on which the London accent has had a significant effect. Its large population and high concentration of mobile phones are reflected in the high proportion of calls from this area.

The major population centres of the Midlands and North of England, with their individual accents, have been separated along geographical grounds.

The three Welsh areas have been allocated according to phonological differences in the Welsh language dialects, which affect the way in which English is spoken in these areas. (The smaller numbers reflect the lower population of Wales; about 1 million in total.)

Scotland has been divided into three areas; the population centres containing Glasgow and Edinburgh, and the more sparsely populated area of Northern Scotland.

Northern Ireland has been allocated one area, with its tiny population and most of the data collection being carried out in Belfast, which has residents from all over the province.

All areas may contain speakers from Britain's ethnic minorities, providing they have lived in the relevant area since birth or early childhood.

Number	Name of accent/region	Number of speakers	Number of speakers (%)
1	EAST ANGLIA	50	5
2	EAST MIDLANDS	50	5
3	LONDON AND SOUTH EAST	250	25
4	NORTH EAST ENGLAND	60	6
5	NORTH WEST ENGLAND	80	8
6	NORTH EAST WALES	30	3
7	NORTH WEST WALES	20	2
8	NORTHERN ENGLAND	50	5
9	NORTHERN IRELAND	60	6
10	NORTHERN SCOTLAND	40	4
11	SOUTH WALES	60	6
12	SOUTH WEST ENGLAND	60	6
13	SOUTH WEST SCOTLAND	40	4
14	SOUTH EAST SCOTLAND	50	5
15	WEST MIDLANDS	100	10

	TOTAL	1000	100
--	-------	------	-----

Table <\$ code> - Distribution of speakers over regions

## 4.2 Speaker characteristics

Age groups	Number of male speakers	Number of female speakers	Percentage of total
under 16			0
16-30			30
31-45			30
46-60			30
over 60			10

Table <\$ code> - Distribution of speakers over age groups and sexes

See file SPEAKER.TBL in the /TABLE directory.

## 5. The lexicon

See the SAMPALEX.ps file in the \DOC directory for the phonetic alphabet used. Orthographic entries are all in lower case in the lexicon. Entries are split only at spaces, not at apostrophes or hyphens. Entries are in simple alphabetic order.

## 6. Recording conditions

### 6.1 Environments

25% of calls are made from each of the following environments: home-office; public place; street; moving vehicle.

### 6.2 Network called from

All calls are from the GSM network.

## 7. Deviations from SpeechDat specifications

## 8. Sample Prompt sheets

Sample prompt and instruction sheets are given in the /PROMPT directory.