

# SIEMENS SHANGHAI MANDARIN FDB-1000

## A Mandarin Speech Database for the Fixed Telephone Network

*Version 1.1, December 1998*

*Gabriele Bakenecker*

*Herbert S. Tropsch*

*Siemens AG, München*

*Corporate Technology Department*

*Copyright © 1998 by*

*SIEMENS AG, Berlin and Munich*

### **Content:**

|   |   |
|---|---|
| 1. Introduction.....  | 2 |
| 2. Volume ID, directory structure and file name conventions ..... | 2 |
| 2.1 Volume ID.....  | 2 |
| 2.2 Directory structure .....                                     | 2 |
| 2.3 File name convention .....                                    | 3 |
| 3. Database content .....   | 3 |
| 4. Sample file format.....  | 4 |
| 5. Transcription.....   | 5 |
| 6. Speaker information.....                                       | 7 |
| 7. Speaker statistics.....  | 8 |

## 1. Introduction

SIEMENS SHANGHAI MANDARIN FDB-1000 is a speech database consisting of 1,041 recordings of telephone calls. In total the database comprises 70,189 sample files and corresponding transcriptions.

Each speaker made one call. The residence of each speaker and the place where he made the call was in Shanghai.

Although Wu is the major dialect/language used in the Shanghai area the callers had to answer in Mandarin as it is spoken in Shanghai either as first or as second dialect/language. All speakers were prompted for the same items. The recorded items are 70 expressions suitable for teleservices.

Each speaker followed the prompts on a "prompt sheet" and could read the corresponding answers. In this sense the collected material is not spontaneous.

Sample files containing no spoken material or defective material caused by the speaker were excluded from the speech database. Defective material in this sense are words which were not prompted for and/or unintelligible parts of speech. Sample files containing noise (caused or not caused by the speaker) or filled pauses are included in the speech database.

All sample files are transcribed in Pinyin with tone markers and word boundaries. Noises and filled pauses are marked as well.

The recording platform was installed in Shanghai and directly connected to an EURO-ISDN-switch (DSS1 protocol). Most callers used the fixed telephone network. Only a few callers used cordless or mobile telephones. The speech file format is a-law 8 kHz wave format.

Speaker recruitment, recordings and annotation were carried out by Shanghai Jiao Tong University contracted by Siemens AG, Corporate Technology Department, Munich.

## 2. Volume ID, directory structure and file name conventions

### 2.1 Volume ID

The database is stored on two CDs labeled as SM1000\_1 and SM1000\_2.

### 2.2 Directory structure

The directory structure is as follows:

```
\<partition>\C<channel_ID>S<session_ID>\ANSWER<answer_ID>.WAV
```

```
partition :      G | J | K
channel_ID :     01 .. 13 | 16 .. 27 | 29 | 30
session_ID :     0001 .. 0435 /*not contiguous*/
```

answer\_ID : 04 .. 73 /\*00 .. 03 not included, since personal data\*/

### 2.3 File name convention

Directory names and file names are not longer than 8 characters. File name extensions have 3 characters. Directory names, file names and extensions are in capitals.

## 3. Database content

CD SM1000\_1 contains the documentation directory, the transliteration directory, and partitions G and K.

CD SM1000\_2 contains the documentation directory, the transliteration directory, and partition J.

Partition G comprises 469 calls with 31,933 sample files in total.

Partition J comprises 514 calls with 34,405 sample files in total.

Partition K comprises 58 calls with 3,851 sample files in total.

In total these are 1,041 calls comprising 70,189 sample files, where one speaker made one call.

Following table shows all prompts (identical for each call) in Pinyin with tone markers and word boundaries (next to answer-ID, Chinese characters encoded in GB, and their English equivalent):

| <i>Answer ID</i> | <i>Chinese characters</i> | <i>Pinyin with tone markers and word boundaries</i> | <i>English equivalent</i> |
|------------------|---------------------------|---|---------------------------|
| 04               | ??                        | qu3 xiao1   | cancel                    |
| 05               | ?                         | ba1   | eight                     |
| 06               | ??                        | ji1 huo2  | activate                  |
| 07               | 0(?)                      | dong4   | zero                      |
| 08               | ?                         | ka1   | on                        |
| 09               | ??                        | gai3 bian4  | change                    |
| 10               | ???                       | da3   dian4 hua4                                    | call                      |
| 11               | ??                        | tong1 zhi1  | announcement              |
| 12               | ??                        | lu4 yin1  | record                    |
| 13               | ?                         | guan1   | off                       |
| 14               | 1(?)                      | ya1   | one                       |
| 15               | ??                        | bo1 fang4   | play                      |
| 16               | ??                        | xin4 xil  | information               |
| 17               | ????                      | ying4 da2   she4 bei4                               | answering machine         |
| 18               | ??                        | que4 ren4   | confirmation              |
| 19               | ??                        | fan1 yue4   | skip forward              |
| 20               | ??                        | fu2 wu2   | service                   |
| 21               | ?                         | san1  | three                     |
| 22               | ??                        | shu1 ru4  | enter                     |
| 23               | ?                         | yi1   | one                       |

|    |      |                        |                  |
|----|------|------------------------|------------------|
| 24 | ??   | zhong1 zhi3            | terminate        |
| 25 | ??   | dao4 dai4              | rewind           |
| 26 | ??   | wai4 bu4               | external         |
| 27 | ?    | wu3                    | five             |
| 28 | ??   | bang1 zhu4             | help             |
| 29 | ??   | liu2 xia4              | leave behind     |
| 30 | ???  | xin4 xi1   he2         | information box  |
| 31 | ?    | er4                    | two              |
| 32 | ??   | nei4 bu4               | internal         |
| 33 | ?    | shi4                   | yes              |
| 34 | ???? | hui4 yi4   dian4 hua4  | conference call  |
| 35 | ??   | lie4 biao3             | list             |
| 36 | ??   | shan1 chu2             | delete           |
| 37 | ??   | cai4 dan1              | menu             |
| 38 | ??   | xiao1 xi0              | message          |
| 39 | ???  | xia4   yi0 xiang4      | next             |
| 40 | ?    | bu4                    | no               |
| 41 | ??   | zeng1 jia1             | add              |
| 42 | ?    | jiu3                   | nine             |
| 43 | ???? | jing3 ji2   dian4 hua4 | emergency call   |
| 44 | ?    | ling2                  | zero             |
| 45 | ??   | hao4 ma3               | number           |
| 46 | ??   | bian1 cheng2           | program          |
| 47 | ??   | jing3 hao4             | hash/square      |
| 48 | ??   | hui2 dian4             | call back        |
| 49 | ??   | bao3 hu4               | protection       |
| 50 | ?    | liu4                   | six              |
| 51 | ?    | qi1                    | seven            |
| 52 | ??   | bao3 cun2              | save/store       |
| 53 | ??   | yu3 yan2               | language         |
| 54 | ??   | xing1 hao4             | star             |
| 55 | ??   | ting2 zhi3             | stop             |
| 56 | ??   | dian4 hua4             | telephone        |
| 57 | ??   | zhuan3 fa1             | call forwarding  |
| 58 | ??   | zhuan3 jie1            | call forwarding  |
| 59 | ??   | jie1 tong1             | connect/transfer |

604ie1|tong14ie1 tong1

The speech file format is a-law 8 kHz wave format.

## 5. Transcription

Transcriptions of speech files are kept in directory `TRANSLIT` as one separate file per partition, namely `G.TRL`, `J.TRL` and `K.TRL`.

All transcriptions are encoded in ISO 8859-1.

Format of each line of these sample files:

```
path_name_of_sample_file <BLANK> <background_noise> <BLANK> transliteration <CR> <LF>
```

Example:

```
\K\C02S0002\ANSWER11.WAV - +in tong1 -in zhi1 <CR> <LF>
```

```
background_noise : - /* no stationary background noise */ |
                  l /* low stationary background noise */ |
                  m /* medium stationary background noise */ |
                  h /* high stationary background noise */
```

The `transliteration` contains the answer given to the corresponding prompt. The transliteration is done manually after listening to the sample file. The elements of `transliteration` are:

- *Canonical Pinyin syllable with tone marker* for tone 1, tone 2, tone 3, tone 4 and tone 0 (neutral tone). If there are more than one syllables they are separated by blank, e.g. `qu3 xiao1`
- *Word boundary*: Besides at the begin and end of `transliteration` word boundaries are marked as `|` and separated by blank, e.g.: `ying4 da2 | she4 bei4`
- *Filled pause*: Marked as `+fp` and separated by blank (hesitation sounds like "uh", "um", "er", "ah")
- *Non-speech speaker noise*: Marked as `+sn` and separated by blank. It comprises all kinds of sounds and noises by the calling speaker that are not part of the prompted text, e.g. lip smack, cough, grunt, throat clear, tongue click, loud breath, laugh, loud sigh, etc.) E.g. `jie1 xian4 +sn yuan2`
- *Very short noise* that is not caused by the speaker: Marked as `.in` and separated by blank. That kind of noise may occur only once or isolated and is very short (like door slam), e.g. `jie1 xian4 .in yuan2`
- *Intermittent noise* that is not caused by the speaker: Marked pair-wise as `+in` (begin) and `-in` (end) and separated by blank: That kind of noise is of intermittent nature, i.e. it has pauses (like phone ringing), or suddenly changes its color over the time (like background speech, baby crying, door bell, paper rustle, cross talk), e.g. `jie1 +in xian4 yuan2 -in`

File `ELCOUNTS.TXT` contains the list of all transliterated elements (syllables, word boundary, non-stationary background noise, filled pause) as they occur in files `{G|J|K}.TRL` together with their frequency of occurrence.

Frequency of stationary background noises:

| <i>Type of stationary background noise</i> | <i># sample files</i> |
|--|-----------------------|
| - /* no */                                 | 50,176                |
| l /* low */                                | 12,829                |
| m /* medium */                             | 5,383                 |
| h /* high */                               | 1,801                 |
| <i>Total</i>                               | <i>70,189</i>         |

Frequency of non-stationary noises:

| <i>Type of non-stationary noise</i>  | <i># sample files</i> |
|--------------------------------------|-----------------------|
| .in /* very short noise */           | 8,322                 |
| +sn /* non-speech speaker noise */   | 6,048                 |
| +in /* intermittent noise (begin) */ | 3,652                 |
| -in /* intermittent noise (end) */   | 3,652                 |
| +fp /* filled pause */               | 263                   |

## 6. Speaker information

Speaker information is provided in file `SPEAKER.DAT`

Format of each line of these files:

```
speaker-id <TAB> gender <TAB> age <TAB> mother_tongue type_of_telephone province_of_growing_up
<CR> <LF>
```

where

```
speaker-id :           <partition>_<channel-number>_<session-number>
partition :           J | G | K
channel-number :      1 .. 30
session-number :      1 .. 435
gender :              m /*male*/ | f /*female*/ | ? /*unknown*/
age :                 8 .. 71 | ? /*unknown*/
mother_tongue :       14 | 42 | 111 | 112 | 113 | 121 | 122 | 123 | 124 | 131 | ?
/* see table "Distribution of mother tongues" below */
```

```

type_of_telephone :   pu3 tong1 dian4 hua4 /*plain old telephone*/ |
                      wu2 sheng2 dian4 hua4 /*cordless telephone*/ |
                      yi2 dong4 dian4 hua4 /*mobile telephone*/

province_of_growing_up : an1 huil | bei3 jing1 | chong2 qing4 | fu2 jian4 |
                          guang3 xil | gui4 zhoul | he2 bei3 | he2 nan2 |
                          heil long2 jiang1 | hu2 bei3 | hu2 nan2 | ji2 lin2 |
                          jiang1 sul | jiang1 xil | liao3 ning2 | nei4 meng2 gu3 |
                          ning2 xia4 | qing1 hai3 | shan1 dong1 | shan3 xil |
                          shang4 hai3 | si4 chuan1 | tian1 jing1 | xin1 jiang1 |
                          yun2 nan2 | zhe4 jiang1 | ? /*Unknown*/

```

Extract of file SPEAKER.DAT :

| <i>Speaker-ID</i> | <i>Gender</i> | <i>Age</i> | <i>Mother tongue</i> | <i>Type of telephone</i> | <i>Province of growing up</i> |
|-------------------|---------------|------------|----------------------|--------------------------|-------------------------------|
| G_29_100          | f             | 11         | 121                  | pu3 tong1 dian4 hua4     | shang4 hai3                   |
| G_29_102          | m             | 40         | 111                  | pu3 tong1 dian4 hua4     | shan3 xil                     |
| G_29_103          | f             | 48         | 121                  | pu3 tong1 dian4 hua4     | shang4 hai3                   |
| G_29_109          | m             | 41         | 121                  | wu2 sheng2 dian4 hua4    | shang4 hai3                   |
| G_29_111          | f             | 43         | 112                  | pu3 tong1 dian4 hua4     | an1 huil                      |
| G_29_115          | m             | 24         | 121                  | yi2 dong4 dian4 hua4     | shang4 hai3                   |
| G_29_116          | f             | 21         | 121                  | pu3 tong1 dian4 hua4     | shang4 hai3                   |

## 7. Speaker statistics

In total there are 1041 calls recorded and annotated. Each speaker made one call. And for all speakers Shanghai was the province of current residence and province where the call was made.

Distribution of gender:

| <i>Gender</i> | <i># Speakers</i> |
|---------------|-------------------|
| Male          | 412               |
| Female        | 526               |
| Unknown       | 103               |
| <i>Total</i>  | <i>1041</i>       |

Distribution of age:

| <i>Age group</i> | <i># Speakers</i> |
|------------------|-------------------|
| < 16             | 80                |
| 16 - 30          | 281               |

|              |             |
|--------------|-------------|
| 31 - 45      | 329         |
| 46 - 60      | 193         |
| > 60         | 51          |
| Unknown      | 107         |
| <i>Total</i> | <i>1041</i> |

Distribution of mother tongues (as (sub-)branches of major linguistic groups):

| <i>Language Code</i> | <i>Mother tongue</i>                      | <i># Speakers</i> |
|----------------------|---|-------------------|
| 121                  | Sino-Tibetan : Southern : Wu              | 830               |
| ?                    | Unknown                                   | 104               |
| 112                  | Sino-Tibetan : Mandarin : Eastern         | 40                |
| 111                  | Sino-Tibetan : Mandarin : Northern        | 33                |
| 113                  | Sino-Tibetan : Mandarin :<br>Southwestern | 14                |
| 122                  | Sino-Tibetan : Southern : Gan             | 7                 |
| 123                  | Sino-Tibetan : Southern : Xian            | 4                 |
| 14                   | Sino-Tibetan : Kam-Tai                    | 3                 |
| 42                   | Altai : Mongolian                         | 2                 |
| 124                  | Sino-Tibetan : Southern : Min             | 2                 |
| 131                  | Sino-Tibetan : Tibetan : Amdo             | 2                 |
| <i>Total</i>         |   | <i>1041</i>       |

Distribution of provinces where grown up:

| <i>Province of growing up</i> | <i># Speakers</i> |
|-------------------------------|-------------------|
| ?? shang4 hai3                | 817               |
| ? (Unknown)                   | 104               |
| ?? jiang1 sul                 | 23                |
| ?? zhe4 jiang1                | 13                |
| ?? an1 hui1                   | 13                |
| ?? si4 chuan1                 | 7                 |
| ?? jiang1 xi1                 | 7                 |
| ?? he2 nan2                   | 6                 |
| ?? tian1 jing1                | 5                 |
| ?? liao3 ning2                | 5                 |
| ?? bei3 jing1                 | 5                 |
| ?? yun2 nan2                  | 4                 |
| ?? hu2 nan2                   | 4                 |
| ?? hu2 bei3                   | 4                 |
| ?? shan3 xi1                  | 3                 |
| ?? shan1 dong1                | 3                 |
| ??? he1 long2 jiang1          | 3                 |
| ?? guang3 xi1                 | 3                 |
| ?? qing1 hai3                 | 2                 |
| ?? gui4 zhou1                 | 2                 |
| ?? fu2 jian4                  | 2                 |
| ?? xin1 jiang1                | 1                 |

|                    |             |
|--------------------|-------------|
| ?? ning2 xia4      | 1           |
| ??? nei4 meng2 gu3 | 1           |
| ?? ji2 lin2        | 1           |
| ?? he2 bei3        | 1           |
| ?? chong2 qing4    | 1           |
| <i>Total</i>       | <i>1041</i> |

Distribution of type of telephone set:

| <i>Type of telephone set</i>                        | <i># Speakers</i> |
|---|-------------------|
| ???? pu3 tong1 dian4 hua4 /* plain old telephone */ | 874               |
| ???? wu2 sheng2 dian4 hua4 /* cordless telephone */ | 56                |
| ???? yi2 dong4 dian4 hua4 /* mobile telephone */    | 7                 |
| ? /* unknown */                                     | 104               |
| <i>Total</i>  | <i>1041</i>       |