

# CZECH SPEECON ADULT DATABASE



Author(s):	Petr Pollák, Czech Technical University in Prague (1) Jan Černocký, Brno University of Technology (2)
Institute:	(1) Dept. of Circuit Theory (2) Inst. of Computer Graphics and Multimedia
Address:	(1) ČVUT FEL K331 Technická 2 16627 Praha 6 – Dejvice  (2) VUT FIT Božetěchova 2 61200 Brno
email:	<a href="mailto:pollak@feld.cvut.cz">pollak@feld.cvut.cz</a> <a href="mailto:cernocky@fit.vutbr.cz">cernocky@fit.vutbr.cz</a>
Date:	6 Apr 2004
Database owner:	Castel GmbH Bergstr. 115, 71579 Spiegelberg, Germany  Contact persons: Kalman Szechenyi - <a href="mailto:kalman.szechenyi@gmx.de">kalman.szechenyi@gmx.de</a>
Version:	1.0 – 5 May 2003 – Petr Pollák - the first draft – Common version for Adult and Child
	1.1 – 15 May 2003 – Petr Pollák – the second draft - Adult only – before prevalidation
	1.2 - 10 Nov 2003 – Petr Pollák –before validation of complete database (draft without final statistics)
	1.3 (Final) - 14 Nov 2003 – Petr Pollák, remarks of Jan Černocký – final – before validation of complete database,
	1.4 (PostFinal) - 6 Apr 2004– Petr Pollák,– remarks and correction after validation and acceptance of the database,

<b>1. Introduction</b> .....	<b>4</b>
<b>2. Distribution media</b> .....	<b>4</b>
2.1 STORAGE MEDIA .....	4
2.2 FILE AND DIRECTORY STRUCTURE.....	4
2.2.1 Root directory .....	4
2.2.2 Signal and label file structure.....	5
2.2.3 Signal and label file conventions .....	5
2.2.4 Corpus Codes.....	5
2.3 DOCUMENTATION DIRECTORIES AND FILES .....	7
2.3.1 Documentation directories .....	7
2.3.2 Files in DOC directory.....	7
2.3.3 Files in TABLE directory.....	8
2.3.4 Files in INDEX directory.....	10
2.4 NUMBER OF MEDIA AND THEIR CONTENTS .....	10
<b>3. Format of speech and label files</b> .....	<b>11</b>
3.1 FORMAT OF SPEECH SIGNAL FILES.....	11
3.2 FORMAT OF LABEL FILES .....	11
<b>4. Prompting and recording procedure</b> .....	<b>15</b>
4.1 PROMPTS.....	15
4.2 SAMPLE OF PROMPT SHEET .....	15
4.3 RECORDING PROCEDURE .....	20
4.4 SPEAKER RECRUITMENT.....	21
<b>5. Database design</b> .....	<b>21</b>
5.1 FREE SPONTANEOUS SPEECH (F01-F30).....	21
5.2 ELICITED SPEECH.....	26
5.2.1 Birth date, current date, relative date (ED1-3) .....	26
5.2.2 Time of day (ET1-2) .....	27
5.2.3 City names (EC1-2).....	27
5.2.4 Proper names (EP1-3).....	27
5.2.5 Spelling of proper name (EL1) .....	27
5.2.6 Yes/No answers (EQ1-2) .....	27
5.2.7 Languages (EO1).....	27
5.2.8 Telephone numbers (EN1-3).....	28
5.3 READ SPEECH .....	28
5.3.1 Phonetically rich sentences (S01-S30).....	28
5.3.2 Phonetically rich words (W01-05).....	28
5.3.3 Isolated digits (CI1-4) .....	29
5.3.4 Isolated digit sequences (CB1, CB2) .....	29
5.3.5 Digit strings (CC1-4) .....	30
5.3.6 Telephone number (CE1) .....	30
5.3.7 Natural numbers (CNI-3).....	31
5.3.8 Money amounts (CM1).....	31
5.3.9 Times (CT1-2).....	31
5.3.10 Dates (CD1-3).....	32
5.3.11 Spelled words (CL1-3).....	35
5.3.12 Proper Name (CPI) .....	35
5.3.13 Cities (CO1)/street names (CO2).....	36
5.3.14 Yes/No (CQ1-2).....	39
5.3.15 Email & Web addresses (CWI-2).....	40

5.3.16	<i>Special keyboard characters (CK1-2)</i> .....	47
5.3.17	<i>Application specific commands</i> .....	48
<b>6.</b>	<b>Recording platforms</b> .....	<b>69</b>
<b>7.</b>	<b>Recording environments</b> .....	<b>69</b>
7.1	OFFICE ENVIRONMENT .....	70
7.2	ENTERTAINMENT ENVIRONMENT .....	72
7.3	PUBLIC ENVIRONMENT .....	75
7.4	CAR ENVIRONMENT .....	79
<b>8.</b>	<b>Speaker distribution</b> .....	<b>81</b>
8.1	AGE AND GENDER DISTRIBUTION .....	81
8.2	DIALECT DISTRIBUTION.....	81
<b>9.</b>	<b>Annotations</b> .....	<b>83</b>
9.1	GENERAL DESCRIPTION OF ANNOTATIONS .....	83
9.2	ANNOTATIONS OF SPONTANEOUS SPEECH .....	84
9.3	ANNOTATIONS OF SPELLED ITEMS.....	84
<b>10.</b>	<b>Lexicon information</b> .....	<b>88</b>
10.1	CZECH SAMPA TABLE .....	88
10.2	PHONETIC ANNOTATIONS – SOURCE FOR LEXICON GENERATION.....	90
10.3	LIST OF FOREIGN WORDS.....	92
10.4	LEXICON OVER-COMPLETENESS.....	98
<b>11.</b>	<b>Double checks</b> .....	<b>98</b>
<b>12.</b>	<b>Other information</b> .....	<b>99</b>
12.1	IMPULSE RESPONSE MEASUREMENT .....	99
12.2	DESCRIPTION OF ADDITIONAL SESSIONS 550-589 JOINED TO THE FINAL DATABASE... 99	99
12.3	COMMENTS TO PROMPT & LEXICON AND MINI-DATABASE PRE-VALIDATIONS .....	100
12.4	COMMENTS TO THE ERRORS FOUND DURING FINAL DATABASE VALIDATION .....	102
12.5	MODIFICATION AFTER VALIDATION .....	104

# 1. Introduction

Czech SpeeCon is the speech database collected in the frame of the project No. IST-1999-10003, “Speech Driven Interfaces for Consumer Applications“. This document describes the 550 speaker adult database.

Another 40 recorded sessions are included in the database. These sessions contain additional recordings realized as compensation for found problems. Re-recorded bad sessions were not thrown out but they are also included in the database because from some points of view, these sessions are as good as the others. The detailed description of all these additional sessions can be found in section 12.2 of this document.

All statistics and numbers required by database specification are computed on the set of main 550 adult speakers.

The Czech SpeeCon database collection is carried out by the Department of Circuit Theory of Czech Technical University in Prague (contact person Petr Pollák – pollak@feld.cvut.cz) and by the Institute of Computer Graphics and Multimedia of the Brno University of Technology (contact person Jan Černocký – cernocky@fit.vutbr.cz) with financial support of Castel GmbH (contact person Kalman Szechenyi – kalman.szechenyi@gmx.de) and EC (project officer Domenico Perrotta – perrotta@cec.eu.int).

The final DB is owned by *Castel*, *CTU* and *VUT* have rights to use the database for their own purposes. Final database is distributed by ELRA.

## 2. Distribution media

### 2.1 Storage media

The storage media for validation and distribution will be named according to the following scheme

`<database><p><oo>`

where `<database>` is “ADULT1CS” as defined in D214, `<p>` is “\_” for data media or “D” for documentation media. `<oo>` is a two digit code. For Czech DBs it is

`ADULT1CS_##`

`ADULT1CSD01`

The medium name is always stored in file DISK.ID in the root directory.

### 2.2 File and Directory structure

#### 2.2.1 Root directory

The following files are present in the root directory of each DB's DVD:

DISK.ID	disk identification – as written above in 2.1.
README.TXT	plain text database description file (only on the documentation disc)
COPYRIGH.TXT	plain text copyright file

**Table 1 – Contents of root directory**

### 2.2.2 Signal and label file structure

The general structure for all signal and the label files is

/<database>/<block>/<session>

with “/” a generic file system separator symbol. Session numbers are provided by the recording platform; they are generated automatically. The following directories are used for Czech database

<database>	ADULT1CS
<block>	BLOCK<NN> <NN> is a number from 00 to 99
<session>	SES<NN><M> <NN> is the block number <M> is session number from 0 to 9

**Table 2 – Signal and label files directories**

### 2.2.3 Signal and label file conventions

File names adhere to the common subset of the ISO 9660 standard, i.e. file names with 8 characters followed by a 3 character file extension. Generally, the name is composed as

<dbID><NNM><CCC>.<LL><F>

<dbID>	Database Identification Code, For SpeeCon it is: “SA” = adult,
<NNM>	Progressive recording session number (000 to 999), where NN is the block number and M is the session number; see Section 2.2.2
<CCC>	Corpus code
<LL>	“CS” - ISO 639 language code
<F>	File type code: “O”: orthographic label file “0,1,2,3”: speech signal files for channels 0 - 3

**Table 3 – Signal and label files names**

### 2.2.4 Corpus Codes

According to the corpus description in D213, the following corpus codes have been defined for SpeeCon (D214):

Corpus id.	Item id.	Corpus contents	
<b>Calibration data</b>			
– (underscore)	01 – 06	noise recording(s): medium distance: 01=mid position, 02=left position, 03=right position far distance: 04=mid position, 05=left position, 06=right position	
N	01	The “silence word” recording	
<b>Free spontaneous items</b>			
F	01 – 30	Free spontaneous, rich context items (story telling)	an open number of spontaneous topics out of a set of 30 topics
<b>Elicited spontaneous items</b>			
E	D1 – D3	3 elicited dates	17 elicited spontaneous items
E	T1 – T2	2 elicited times	
E	P1– P3	3 elicited proper names	
E	C1 – C2	2 elicited city name	
E	L1	1 elicited letter sequence	
E	Q1 – Q2	2 elicited answers to questions	
E	N1 – N3	3 elicited telephone numbers	
E	O1	1 elicited language	
<b>Read speech</b>			
S	01 – 30	30 phonetically rich sentences	
W	01 – 05	5 phonetically rich words	
<b>Core words (read), 31 general words and phrases, 208 applic. specific words and phrases</b>			
C	I1 – I4	4 isolated digits	32 general words and phrases,
C	B1 – B2	2 isolated digit sequences <sup>1</sup>	
C	C1 – C4	4 connected digit sequences	
C	E1	1 telephone number	
C	N1 – N3	3 natural numbers	
C	M1	1 money amount	
C	T1 – T2	2 time phrases T1 : analogue, T2 : digital	
C	D1 – D3	3 dates :D1 : analogue, D2 : relat. and gen. date, D3 : digital	
C	L1 – L3	3 letter sequences	
C	P1	1 proper name	
C	O1 – O2	2 city or street names	
C	Q1 – Q2	2 questions	
C	K1 – K2	2 special keyboard characters	
C	W1	1 Web address	
C	W2	1 email address	
Y	01 – 95	core word synonyms	
1	01 – 85	Basic IVR commands	Total of 208 words and phrases per session out of 450.
2	01 – 40	Directory navigation	
3	01 – 22	Editing	
4	01 – 57	Output control	
5	01 – 70	Messaging & Internet browsing	
6	01 – 33	Organiser functions	
7	01 – 39	Routing	
8	01 – 12	Automotive	
9	01 – 95	Audio & Video	

## 2.3 Documentation directories and files

### 2.3.1 Documentation directories

The documentation will be held in a file system with the following structure

/ADULT1CS/DOC	Documentation
/ADULT1CS/TABLE	Speaker, recording condition, environment conditions, and lexicon tables
/ADULT1CS/INDEX	Index files

Table 4 – Documentation directories

### 2.3.2 Files in DOC directory

This directory contains documentation files, including a description of the database design in one of these formats:

DOC		Microsoft Word text processor file
	DESIGN.DOC	
	VALREP.DOC	
TXT		ISO 8859-1 DOS-formatted text file
	SUMMAR0.TXT	
PDF		Adobe Portable Document Format
	DESIGN.PDF	
	ISO88592.PDF	
	SAMPALEX.PDF	
PS		Adobe PostScript format
	ISO88592.PS	
	SAMPALEX.PS	
HTM		XML or HTML 4.0 format

Table 5 – Content of the DOC directory

#### **DESIGN.DOC | DESIGN.PDF**

This file in English contains the information about design, collection, annotations, and completion of Czech SpeeCon database.

#### **VALREP.DOC**

Validation report provided by SPEX.

#### **ISO88592.PS | ISO88592.PDF**

Sample character table corresponding to the Czech database in PostScript format.

#### **SAMPALEX.PS | SAMPALEX.PDF**

The SAMPA table used in PostScript format.

#### **SUMMAR0.TXT**

The mandatory summary file describing all recording sessions. It has the following fields:

DIR	Full directory path of the session
SES	Session number
CCD <sup>N</sup>	A string of typically N codes, where N is the number of total items. The N codes must be given in the same order as they were prompted. The N codes must be concatenated and separated by N-1 COMMAS (no other character must be used).  This first string is followed by a second one (separated from the first by BLANK) which is a copy of the first string but with missing items indicated by the string '---'.
RED	Recording date of first item
RET	Recording time of first item

**Table 6 – Summary file contents**

All fields are separated by **spaces**. It is a documentation file and not a table one (this means that it should not be used as input to any software tool).

The summary file name is SUMMAR0.TXT. It contains positive entries for every item that was recorded and whose close talk channel (channel 0) was annotated.

### 2.3.3 Files in TABLE directory

The table files are

- SPEAKER.TBL
- REC\_COND.TBL
- SESSION.TBL
- LEXICON.TBL

The speaker, session, and recording condition tables are related to each other. All data is stored in a DBMS-like structure, i.e. without redundancy and unique key values in each table. The relationship between tables is established by using a common SAM label in the related tables (in DBMS terminology the SAM labels are “attributes“. A SAM label is a “primary key“ attribute in one table, and in all related tables it is a “secondary“ or “foreign key“ attribute).

#### **File - SPEAKER.TBL**

This file contains mandatory information about the speaker. To guarantee a unique identification key, speakers are given a speaker code SCD. The speaker code is generally independent of the current recording session number, however majority of speaker codes are same as session number. Due to found problems, several sessions had to be re-recorded with new speakers and in these cases we have used new SCDs. Logically, when we re-recorded some sessions with same speakers and both sessions are present in the database, speaker codes are the same.

SPEAKER.TBL contains the following fields:

SCD	Unique speaker code
SEX	Speaker gender



AGE	Speaker age
ACC	Speaker accent

**Table 7 – Contents of SPEAKER.TBL**

**File – REC\_COND.TBL**

The recording condition table stores all information relevant to a recording session. It contains the following fields:

SES	session number
MIP	Microphone positions
MIT	Microphone types
SCC	scenario code

**Table 8 – Contents of REC\_COND.TBL**

The SES field relates this table to SESSION.TBL.

**File – SESSION.TBL**

SESSION.TBL stores information about the recording session. A session is identified by a unique session number.

SES	Session number
SCD	Unique speaker code
REP	Recording place
RED	Recording date
RET	Recording time

**Table 9 – Contents of SESSION.TBL**

All fields are mandatory.

**File – LEXICON.TBL**

The lexicon file is an alphabetically ordered table of distinct lexical items which occur in the corpus with the corresponding pronunciation information. Each distinct word should have a separate entry. As the lexicon is derived from the database it must use the same alphabetic encoding for special and accented characters as used in the transcriptions

The SpeeCon lexicon file consists of three mandatory fields:

- Orthography (related to conventions used for annotations, as described in the Section 9),
- Frequency of the occurrence count of a given orthographic form,
- SAMPA pronunciation (Czech SAMPA is described in D216CZE and WEB-  
<http://noel.feld.cvut.cz/sampa> or at official SAMPA WEB-page  
<http://www.phon.ucl.ac.uk/home/sampa/home.htm>),
- and additional optional tab-delimited pronunciation variants fields.

The first line contains the names of the fields:

- Orthography    Frequency    Pronunciation Variants

The lexicon is lowercase (unless spelling items), and it contains at least all word forms found in the orthographic transcriptions, see Section 9. If more pronunciation variants exist for a given orthographic form, then the most common form is entered into the pronunciation field, and the others are placed in the pronunciation variants fields in decreasing order of occurrences. High number of pronunciation variants appears especially in words related to e-mails and WEB addresses.

As in the predecessor projects, the SpeeCon representation of the SAMPA pronunciations differs from the format originally proposed by SAM:

- each SAMPA symbol is delimited by an extra space,
- multiple pronunciations are now in tab-delimited distinct fields,
- the use of the underscore to separate words in idiomatic phrases is not allowed.

More detailed description of lexicon generation is in chapter 10.

#### 2.3.4 Files in INDEX directory

The index files allow quick access to speech and transcription data. Only the **CONTENT0.LST** file is defined. It stores the transcription of the close talk microphone as given in the LBO field and relates it to properties of the signal file and speaker. It is a TAB delimited ASCII file.

DIR	Directory
SRC	speech signal file names
CCD	corpus code
SCD	speaker code
SEX	speaker gender
AGE	speaker age
ACC	speaker accent
SCC	Scenario code
LBO	speech transcription without the numerical data

Table 10 – Contents of CONTENT0.LST

## 2.4 Number of media and their contents

The database is distributed on a set of 19 DVDs (18 with signals + 1 with label-files and documentation). Each DVD contains the following blocks of recording sessions.

DVD label	BLOCKs	SESSIONs
ADULT1CS_00	00-03	000-032
ADULT1CS_01	03-06	033-066
ADULT1CS_02	06-09	067-099

ADULT1CS_03	10-13	100-132
ADULT1CS_04	13-16	133-166
ADULT1CS_05	16-19	167-197
ADULT1CS_06	19-22	198-228
ADULT1CS_07	22-25	229-259
ADULT1CS_08	26-29	260-290
ADULT1CS_09	29-32	291-322
ADULT1CS_10	32-35	323-352
ADULT1CS_11	35-38	353-384
ADULT1CS_12	38-41	385-414
ADULT1CS_13	41-44	415-444
ADULT1CS_14	44-47	445-475
ADULT1CS_15	47-50	476-507
ADULT1CS_16	50-54	508-540
ADULT1CS_17	54-57	541-573
ADULT1CS_18	57-58	574-589
ADULT1CSD00	Documentation + Label files	

**Table 11 – Final media list**

Finally, the adult database contains 550 + 40 speakers.

### **3. Format of speech and label files**

#### **3.1 Format of speech signal files**

The signals are stored in a raw file format, i.e. without headers in the signal file. Each of the four speech channels is recorded at 16 kHz with 16 bit quantization with the least significant byte first (“lohi” or Intel format) as (signed) integers. A description of the sample rate, the quantization, and byte order used is held in the SAM label file.

#### **3.2 Format of label files**

Given the need for some small modifications to the label formats, it was decided to introduce a new version number (version 6.1) for the modified SAM label files. Label files adhere to a modified SAM label format:

```
ABC: item_1, item_2, ..., item_n
```

where

- ABC is a three letter mnemonic followed by a colon; the mnemonic must contain only 7-bit US-ASCII character and may not contain spaces or colons
- items after the mnemonic are separated by commas, i.e. they cannot contain commas themselves

- items can be empty
- spaces after the colon or in between items are recommended to improve readability
- a label line is delimited by <CR><LF>, the line end sequence according to the DOS operating system.

"A label file begins with the mnemonic "LHD:" and ends with "ELF:". The mnemonic "LBD:" splits a label file into two sections: the LABEL FILE HEADER and the LABEL FILE BODY. After LBD: only LBR:, LBO: and ELF: may follow.

SAM label files in Czech Adult database contain also additional EPI field which contains phonetic transcription of utterance in SAMPA, i.e. transcription containing used pronunciation variant in this utterance. In the case of mispronunciation, "\*" is used same way also for EPI field as it is used in LBO field.

**There is one SAM label file assigned to each utterance (i.e. one for all 4 recording channels).**

A detailed description of the used SAM labels can be found in the following table:

SAM Label	Description	Format	Format string
LHD	Label header	Fixed vocabulary item	%s
ELF	End of label file		
CMT	Comment	Free-form text	%s
DBN	Database name	SPEECON_ <LL>	%s
SES	Session number	3-digit number	%03d
SCD	Speaker code	a 3-digit number	%03d
SEX	Speaker gender	Fixed vocabulary item: [M F]	%s
AGE	Speaker age	Integer	%d
ACC	Speaker accent	Fixed vocabulary item from dialect list	%s
DIR	Speech file directory	Fixed vocabulary item from file system \ <dbname&gt;\block&lt;nn&gt;\ses&lt;nn&gt;&lt;m&gt;\< td=""> <td>%s</td> </dbname&gt;\block&lt;nn&gt;\ses&lt;nn&gt;&lt;m&gt;\<>	%s
SRC	Speech file names	A comma separated list of 8.3 file names	%8c.%3c,%8c.%3c,%8c.%3c,%8c.%3c
CCD	Corpus code	3 character code	%3c
REP	Recording place	The value of the PLC attribute in the SCC label	%s
RED	Recording date	DD/Mon/YYYY	%02d/%3c/%4d
RET	Recording time	HH:MM:SS	%02d:%02d:%02d
BEG	Labelled sequence begin position	Integer	%d
END	Labelled sequence end position	Integer: number of sample points in recording	%d
SAM	Sampling frequency	Integer: 16000	%d

SNB	Number of (8-bit) bytes per sample	Integer: 2, signed	%1d,%s
SBF	Sample byte order	Integer: [0 lohi]	%s
SSB	Number of significant bits per sample	Integer: [16]	%d
QNT	Quantization	Fixed vocabulary item, e.g.: PCM	%s
NCH	Num.of channels	Integer: 4	%d
LBD	Label file body		
LBR	Prompt text	BEG,END,<gain>,<min>,<max>,<prompt text> with <gain>, <min>,<max> optional signal values; if they are not known, the values may be left empty, but the correct number of commas must remain. <prompt text> is the text that appears on the screen.	%d, %d, %d, %d, %d, %s
SCC	Scenario code	An attribute-value pair list, ENV = %s, PLC = %s , POS = %s, SIZ = %s, AUD = %s, DRV = %s, that indicates the acoustic environment and the factors that define the actual recording scenario as explained in D212. A value for the ENV attribute is mandatory, but if the sub-scenario is unknown, the values may be left empty. However, the correct number of commas must remain.	%s,%s,%s,%s,%s,%s
MIP	Microphone positions	An attribute-value pair list, CHN0= %s, CHN1= %s, CHN2= %s, CHN3= %s, where the attribute values are specified in D214. The value may left empty, if signal is given channel is not recorded (e.g. CHN2 and CHN3 for impulse response measurement).	%s,%s,%s,%s
MIT	Micophone types	An attribute-value pair list, CHN0= %s, CHN1= %s, CHN2= %s, CHN3= %s, where the attribute values are specified in D214. The value may left empty, if signal is given channel is not recorded (e.g. CHN2 and CHN3 for impulse response measurement).	%s,%s,%s,%s
DBA	Noise level	The noise level, as measured by the noise-level meter during recording of the silence word	%f

SNQ	Signal/Noise Quality	Attribute value pair list, CHN0 = %f, CHN1 = %f , CHN2= %f, CHN3 = %f The SNR values estimated by the recording platform	%f
LBO	Orthographic transcription	<transcription text>	%d, %d, %d, %s

**Table 12 – Mandatory labels for SpeeCon databases**

EXP	Labelling expert	Name Surname, Organization	%s
SYS	Labelling system	Software description	%s
DAT	Date of completion of labelling	DD/Mon/YYYY	%s

**Table 13 – Optional labels used in Czech SpeeCon database**

EPI	Phonetic transcription	Phonetic transcription of the utterance pronunciation (in SAMPA)	%s
-----	------------------------	--	----

**Table 14 – Additional label in Czech SpeeCon database**

An example label file is given below.

```
LHD: SAM 6.1
DBN: SPEECON_CS
SES: 004
CMT: *** Speech Label Information
SRC: SA004S15.CS0, SA004S15.CS1, SA004S15.CS2, SA004S15.CS3
DIR: \ADULT1CS\BLOCK00\SES004
CCD: S15
BEG: 0
END: 100999
REP: PUBHALL_02
RED: 24/Apr/2003
RET: 11:44:10
CMT: *** Speech Data Coding ***
SAM: 16000
SNB: 2 signed
SBF: lohi
SSB: 16
QNT: PCM
NCH: 4
CMT: *** Speaker Information ***
SCD: 004
SEX: M
AGE: 27
ACC: Central Moravia
CMT: *** Recording Conditions ***
SNQ: CHN0=34.015, CHN1=17.528, CHN2=10.906, CHN3=8.745
MIP: CHN0=CLOSE_HEADSET, CHN1=CLOSE_LAVALIER, CHN2=MEDIUM, CHN3=MEDIUM
MIT: CHN0=SENNHEISER_ME104, CHN1=NOKIA, CHN2=SENNHEISER_ME64, CHN3=MBF_HAUN
SCC: ENV=PUBLIC_PLACE, PLC=PUBHALL_02, POS=CLOSE_WALL_01, SIZ=SQM_100_200, AUD=, DRV=
DBA: 57.0
CMT: *** Labelling information ***
SYS: FTP Transcriber 3.0
```

```

EXP: Jan Cernocky
DAT: 01/May/2003, 23:22:42
EPI: [int] [sta] teNkra:t gdiS Slo o zemi *zasli:beno_u jsme *povolili povolili
ale [int] tec se to nestane
CMT: *** Label File Body ***
LBD:
LBR: 0,100999,,,, Tenkrát, když šlo o Zemi zaslíbenou, jsme povolili, ale teď se
to nestane.
LBO: 0,50500,101000,[int] [sta] tenkrát když šlo o zemi *zaslíbenou jsme *povolili
povolili ale [int] teď se to nestane
ELF:

```

**Table 15 – Sample of a label-file**

## 4. Prompting and recording procedure

### 4.1 Prompts

All recording sessions were prompted same way using the recording studio developed for SpeeCon collection. Speaker had always his/her own display where just prompted utterance was displayed commonly with the mark signalling running recording.

After the recording of noises for room impulse response measurement and the recording of silence word item, other recorded items were prompted in the following way.

ADULT1CS database:

- Firstly, free spontaneous speech items F01 – F30 were prompted. Corpus codes were kept for the theme specified in D213, however, all 30 themes were randomly mixed for each speaker. Each speaker has recorded up to 10 items, but due to mixing all themes have sufficient coverage in the final database.
- After the block of free speech, all other items (read and elicited) were recorded in randomly generated sequence to avoid list effect.

### 4.2 Sample of prompt sheet

```

_01: --- noise_rec_1 ---
_02: --- noise_rec_2 ---
_03: --- noise_rec_3 ---
_04: --- noise_rec_4 ---
_05: --- noise_rec_5 ---
_06: --- noise_rec_6 ---
N01: --- silence_word ---
F11: ?? Zadejte navigačnímu systému v automobilu informace o Vaší oblíbené
restauraci, tj. název, adresu, město.

```

F16: ?? Představte si, že jste v restauraci. Co byste si rád objednal ke snídani, k obědu nebo k večeři?  
F14: ?? Jaké zařízení byste rád ovládal v automobilu hlasem?  
F13: ?? Jmenujte zleva doprava nástroje resp. zařízení na palubní desce Vašeho automobilu.  
F29: ?? Jaký je doposud Váš pocit z nahrávání?  
F24: ?? Řekněte Vaší televizi požadavek na přepnutí na určitý program.  
F25: ?? Popište Váš oblíbený televizní program.  
F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si dovolenou, zájezd nebo letenku. Zeptejte se na druh dopravy a dobu trvání.  
F12: ?? Zadejte navigačnímu systému v automobilu požadavek na cestu k nejbližšímu letišti, nádraží, parkovišti, garážím, apod.  
F17: ?? V jakém městě byste rád žil a proč?  
F06: ?? Nadiktujte krátký obchodní dopis s pozváním na pracovní schůzku.  
F19: ?? Do jaké země byste se rád podíval a proč?  
F09: ?? Popište aktuální nebo předchozí dopravní situaci na silnici.  
F05: ?? Představte si, že voláte do knihkupectví nebo hudebnin a zeptejte se na dostupnost a cenu určité knihy nebo CD.  
F01: ?? Pošlete hlasovou zprávu Vašemu lékaři se žádostí o zpětné zavolání.  
F21: ?? Popište, jak se dostat na vlakové nádraží.  
F08: ?? Řekněte Vašemu mobilnímu telefonu s hlasovým rozhraním požadavek na konferenční hovor s určitou osobou, Vaším úřadem a klientem.  
F10: ?? Popište policii resp. záchranářům Vaší aktuální pozici na silnici.  
F20: ?? Prosím, zarezervujte si po telefonu hotelový pokoj.  
F28: ?? Popište Váš oblíbený sport a sportovce.  
F04: ?? Představte si, že voláte do divadla, kina, koncertní síně, restaurace či nočního klubu a zeptejte se na volná místa na určitý program.  
F30: ?? Jaká zařízení byste rád ovládal hlasem?  
F27: ?? Zeptejte se na obsah daného programu v televizi.  
F15: ?? Zeptejte se na určité technické detaily automobilu.  
F18: ?? Popište některá zajímavá místa ve Vašem městě.  
F03: ?? Představte si, že voláte do Vaší banky. Zadejte požadavek na prodej či nákup akcií, případně se zeptejte na ceny na burze.  
F22: ?? Popište stručně Váš oblíbený film.  
F07: ?? Představte si, že voláte do hotelu. Požádejte o změnu Vámi rezervovaného pokoje.  
F23: ?? Řekněte Vašemu Hi-Fi požadavek na puštění určitého CD.  
F26: ?? Vyjmenujte všechna audio-video zařízení, která vlastníte.  
166: později  
S03: Nenosil nic jiného, než vytahaná trička a džínsy.  
540: zákaznické centrum  
S25: Autobusy a sanitky už byly připravené.  
Y67: vytočit  
111: zamknout  
917: dlouhý záznam  
441: hloubky  
612: kontakty  
Y62: stáhnout  
543: auto  
S06: Je to hrozné, co se Frankovi přihodilo, že?  
425: zesvětlit  
236: sloupec  
403: maximum  
113: potvrdit  
ET1: ?? Kolik je právě teď hodin?  
431: hlasitost  
180: infraport  
CC4: osum dva čtyři pět čtyři  
S23: Za dnešek došlo objednávek za třiadvacet miliónů.  
630: stát  
803: topení



934: obraz  
Y45: nízká  
616: textový editor  
617: poznámky  
S10: Jinou náhradní pneumatiku už nemáme.  
932: audio-video  
CI2: osm  
457: číst  
721: konec cesty  
Y39: editovat  
S11: Dali jsme se tomu do smíchu a pak jsme se na ně vrhli oba.  
CW2: usenet@news.vutbr.cz  
948: jazz  
979: motoristické sporty  
410: ubrat  
407: spodní  
984: hudba  
207: začátek  
CK1: trubka  
419: stupnice  
ET2: ?? V kolik hodin přicházíte domů ze zaměstnání?  
Y15: každý měsíc  
CC1: čtyři sedm nula čtyři čtyry  
S12: Uprchlíci nastoupili do vytopených vagónů, dostali jídlo a byli ošetřeni.  
182: hostitel  
310: vložit  
CE1: 971 801 436  
EN1: ?? Uveďte nějaké telefonní číslo do pevné sítě.  
939: sport  
404: minimum  
CL1: Š K U D L Y  
S14: Když potom ztroskotaly mongolské nájezdy, byl zklamán.  
S17: Abys věděl, nenapsal to Paul, ale někdo jiný.  
237: stránka  
156: nahrát  
414: minimalizovat  
CB1: sedm jedna čtyři šest osm dvě čtyři pět  
136: restartovat  
533: rychlé vytáčení  
726: křížovka  
143: nápověda  
216: dolů  
Y32: stránku  
CW1: www.mpsv.cz/scripts/default.asp  
904: titulky  
Y61: mailbox  
705: navigace  
625: firma  
EC2: ?? Ve kterém městě mimo Vaše současné bydliště žijí Vaši přátelé či příbuzní.  
Y44: vysoká  
315: seřadit  
907: satelit  
608: schůzka  
529: zavolat  
509: zavolat zpět  
121: CD vypalovačka  
183: síť  
546: kancelář  
118: fotoaparát

S07: Poněkud nervózně jsem jej otevřel.  
518: SMS  
CQ2: ne  
103: deaktivovat  
Y79: směrovací číslo  
CC3: jedna šest nula nula čtyři  
S18: Je to ekonomická záležitost farmaceutických firem a lékáren.  
201: menu  
CO1: Choceň  
Y55: zvýraznit basy  
730: čerpací stanice  
712: vzdálenost  
S28: Neustále byly vyžadovány nové a nové dokumenty.  
965: komedie  
736: východ  
132: video  
159: čas  
175: stav  
552: mezinárodní  
808: okno  
554: domácí  
CN3: devět set čtyřicet tisíc čtyři sta  
Y84: výstaviště  
238: vybrat  
CN2: deset milionů  
320: upravit  
561: podržet  
W02: odzemek  
211: dole  
Y13: každý den  
307: kopírovat  
Y65: link  
Y75: jednání  
219: doprava  
CI3: osm  
Y77: nástroje  
CL3: Ď Ú Ň Ý Ó Q Y Ď W  
Y31: skočit na  
225: červený  
CD3: osmáctého sedmý, tisíc devět set devadesát pět  
566: délka  
919: disketa  
941: drama  
S24: Plánička vybíhá a Podrazil vsítil první branku Viktorky.  
733: restaurace  
810: rozmrazování  
CI4: šest  
319: OK  
452: rychleji  
Y83: veletrh  
CM1: čtyři tisíce sto dvacet korun, nula haléřů  
235: řádek  
Y27: zelená  
S19: Do značné míry byl útlum způsoben špatným odbytem džinsů.  
EP1: ?? Jak se jmenuje Váš otec nebo Vaše matka? (Je možné uvést jméno křestní, příjmení, či obojí.)  
909: AUX  
163: týdně  
Y47: skrýt  
943: věda  
Y46: program

983: kreslené filmy  
946: oddech  
703: GPS  
606: rok  
CN1: osmdesát pět tisíc tři sta  
724: autopůjčovna  
Y20: nabídka  
532: volit číslo  
S01: A to nemluvím o mnohem vyšší ceně podzemního uložení.  
Y72: přílet  
EL1: ?? Hláskujte jméno vašeho rodného města.  
959: dobrodružný film  
EO1: ?? Uveďte jméno jednoho jazyka, kterým nemluvíte.  
440: barva zvuku  
427: barva  
EQ1: ?? Snídal jste dnes ráno?  
436: utlumit  
525: následovat  
W05: neodzrnila  
991: vaření  
Y92: seriál  
ED1: ?? Uveďte datum narození jedné osoby z Vaší rodiny (rodičů, sourozenců, dětí, prarodičů).  
153: trénovat  
CD1: čtvrtek, dvacátého devátého února, dva tisíce patnáct  
539: tísňové volání  
512: e-mail  
537: předplacené služby  
ED3: ?? Kdy odjíždíte na Vaši příští dovolenou? (Je možné použít relativní datum)  
501: zpráva  
S22: Odrazem světla hned na kartónu všechny ty drahokamy dvojnásob vynikly.  
935: digitální rádio  
S21: Bolest poražených Ázerbájdžánců ovšem nezmizí.  
981: zimní sporty  
930: kapacita  
992: reklama  
Y26: žlutá  
S26: Otiskla se tam s přesností vskutku podivuhodnou.  
945: rock  
EN3: ?? Uveďte nějaké telefonní číslo do mobilní sítě.  
149: dobrý den  
717: prohlídka  
CT1: za tři minuty tři čtvrtě na jedenáct  
318: nahradit  
976: tenis  
106: heslo  
420: rozlišení  
626: ulice  
Y42: normální  
101: zapnout  
EP2: ?? Uveďte jedno jméno Vašeho oblíbeného herce, zpěváka, spisovatele či jiné významné osobnosti.  
S04: Mně nikdo gaunerů nadávat nebude, prohlásil znovu kapitán.  
185: opakovat  
Y80: lokalizace  
916: běžný záznam  
321: číslo  
EN2: ?? Uveďte nějaké telefonní číslo do pevné sítě.  
722: letiště  
807: teplota

424: jas  
 165: dříve  
 405: úroveň  
 147: zkontrolovat  
 605: měsíc  
 545: mobil  
 570: rezervovat  
 CO2: Zelená  
 437: zvuky  
 812: údržba  
 401: profil  
 739: jih  
 CC2: osm čtyři čtyři sedm čtyry  
 CB2: nula sedum čtyry čtyry dva devět osum tři  
 950: domácí scéna  
 964: horor  
 S05: Neustále jim nohy pode klesávaly a chtěli stále ještě jít někam.  
 Y82: exit  
 Y49: měřítko  
 213: poslední  
 968: romantický film  
 172: makro  
 412: zobrazit  
 S16: Já už jsem člověk z formy, se mnou byste neužila!  
 ED2: ?? Kolikátého je dnes?  
 732: nádraží  
 EP3: ?? Uveďte jméno nějaké společnosti či firmy.  
 449: hands-free  
 621: pracovní  
 Y19: počítač  
 203: volby  
 CI1: nula  
 Y63: příchozí pošta  
 W01: lymfu  
 S29: Kdyby tak neučinila, neohrozila by pouze je, ale i nás ostatní.  
 442: výšky  
 952: folk  
 929: nabíjení baterie  
 S15: Pokud mohl myslit, vynořila se mu jen myšlenka, aby ukryl pušku.  
 725: centrum  
 CT2: pět hodin dvacet sedum minut  
 921: disk  
 906: anténa  
 112: odemknout  
 513: schránka  
 151: demo  
 229: bílý  
 536: konference  
 138: pauza

**Table 16 – Sample prompt sheet**

### 4.3 Recording procedure

First, the speaker is interviewed to ensure the correct speaker profile (suitable age, gender, accent). Then the speaker is instructed about the recordings: only to speak after the red light of the recording software shows that the recording has begun, to repeat the word/sentence if he or she has mispronounced it. All the time a supervisor was present at the recordings to

answer possible questions. A *Perl* script was used to track the distribution of speakers and environment to allow progress monitoring and compatibility with the specifications. The signal quality was regularly checked and a daily backup of the data was performed.

#### **4.4 Speaker recruitment**

Speakers were recruited following ways:

- by distributing invitation sheets to family members, friends, colleagues,
- by asking other employees and students of Czech Technical University in Prague and Brno University of Technology for their participation together with their relatives or friends,
- by co-operation with external institutions with more employees (VZP – Health Insurance Company)
- by advertising in the Internet (<http://www.job.cz>), this way was found as very efficient especially for recruitment in Prague.

All speakers were paid for their participation.

### **5. Database design**

The SpeeCon adult database should consist of 550 adults speaker. The corpus and item identifiers of the various items are specified in the section 2.2.4. The design of particularly recorded items in the database is described with more details in following sections.

#### **5.1 Free spontaneous speech (F01-F30)**

A set of **30** topic classes was defined for each of the SpeeCon application areas. The items were circulated among speakers to achieve a uniform occurrence frequency of each topic at the end.

The questions were prompted by the supervisor to help the speaker to clearly understand the situation or task, but the question is not recorded. Only the answer is recorded.

Following contexts were distinguished for the free spontaneous prompts:

- Mobile phone and PDA (8 prompts, F01-F08)
- Automotive and information kiosk (13 prompts, F09-F21)
- Audio/Video and toys (7 prompts, F22-F28)
- Miscellaneous (2 prompts, F29-F30)

One question from all groups are prompted in each prompt sheet. This set is then randomly mixed and up to 10 items are recorded, see section 4.1. (Several sessions contain the recordings of all these spontaneous items as they were prompted.)

All questions as they are prompted for particular groups (i.e. recording items) are summarized in the following list.

### **F01 - Voice mail messages**

F01: ?? Pošlete hlasovou zprávu Vašemu příteli s pozváním na večírek. (?? *Send a voice mail message to your friend and invite him to a party.*)

F01: ?? Pošlete hlasovou zprávu Vašemu příteli s pozváním do divadla, kina, muzea, restaurace, kasina, apod. (?? *Send a voice mail message to your friend and invite him to theatre, cinema, museum, restaurant, casino, etc.*)

F01: ?? Pošlete hlasovou zprávu Vašemu příteli s pozváním na procházku nebo na výlet. (?? *Send a voice mail message to your friend and invite him to a trip.*)

F01: ?? Pošlete hlasovou zprávu Vašemu příteli se zprávou o významné události ve Vašem životě. (?? *Send a voice mail message to your friend about an important event in your life.*)

F01: ?? Pošlete hlasovou zprávu na určité číslo s dotazem na nějakou informaci a žádost o zpětné zavolání. (?? *Send a voice mail message to given number, ask an information, and ask for the call-back.*)

F01: ?? Pošlete hlasovou zprávu Vašemu lékaři se žádostí o zpětné zavolání. (?? *Send a voice mail message to your family doctor and ask him for the call-back.*)

F01: ?? Pošlete hlasovou zprávu Vašemu lékaři se žádostí o potvrzení Vaší návštěvy. (?? *Send a voice mail message to your family doctor and confirm him your visit.*)

### **F02 - Call a travel agency**

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si dovolenou, zájezd nebo letenku. Zeptejte se na začátek či konec cesty nebo na dobu trvání. (?? *Make believe that you are calling a travel agency and booking tour/trip/flight. Ask about date of start/end of your journey as well as the duration.* )

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si dovolenou, zájezd nebo letenku. Zeptejte se na datum odjezdu a příjezdu resp. odletu a přiletu. (?? *Make believe that you are calling a travel agency and booking tour/trip/flight. Ask about time of departure and arrival.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si dovolenou, zájezd nebo letenku. Zeptejte se na druh dopravy a dobu trvání. (?? *Make believe that you are calling a travel agency and booking tour/trip/flight. Ask about kind of transportation and the duration.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si dovolenou, zájezd nebo letenku. Zeptejte se na cenu dopravy nebo zda je doprava zahrnuta v ceně Vámi plánovaného zájezdu. (?? *Make believe that you are calling a travel agency and booking tour/trip/flight. Ask about transportation costs, whether they are included in the costs of the tour.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si dovolenou, zájezd nebo letenku. Zeptejte se na informace o hotelu resp. služby, které poskytuje. (?? *Make believe that you are calling a travel agency and booking tour/trip/flight. Ask about hotel information, offered services, etc.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si dovolenou, zájezd nebo letenku. Zeptejte se na program zájezdu, případné výlety v ceně, apod. (?? *Make believe that you are calling a travel agency and booking tour/trip/flight. Ask about tour program, whether possible excursions are included in the costs.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si dovolenou, zájezd nebo letenku. Zeptejte se na zajištěné stravování. (?? *Make believe that you are calling a travel agency and booking tour/trip/flight. Ask about catering.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si dovolenou, zájezd nebo letenku. Zeptejte se na slevy pro děti. (?? *Make believe that you are calling a travel agency and booking tour/trip/flight. Ask about children discounts.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si dovolenou, zájezd nebo letenku. Zeptejte se na zdravotní pojištění, zda je zahrnuto v ceně zájezdu. (?? *Make believe that you are calling a travel agency and booking tour/trip/flight. Ask about medical insurance, whether it is included in tour costs.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si dovolenou, zájezd nebo letenku. Zeptejte se kdy a jak je možné si zájezd zakoupit. (?? *Make believe that you are calling a travel agency and booking tour/trip/flight. Ask when and how it is possible to buy the tour*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si dovolenou, zájezd nebo letenku. Zeptejte se na možnosti platby (hotově, platební kartou, šekem, apod.). (?? *Make believe that you are calling a travel agency and booking tour/trip/flight. Ask about payment conditions – in cash, by credit card, by cheque, etc.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si jízdenku. Zeptejte se na spojení do cíle Vaší cesty, specifikujte druh spojení (přímé či s přestupy nebo expresní). (?? *Make believe that you are calling a travel agency and booking a ticket. Ask about the transportation to the point you need (direct, with/without change, express).*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si jízdenku. Zeptejte určité spojení v požadovaném datu, použijte přesné nebo relativní datum. (?? *Make believe that you are calling a travel agency and booking a ticket. Ask about the connection in given date, you may use absolute or relative date.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si jízdenku. Zeptejte se na čas odjezdu a zastávky na požadované trase. (?? *Make believe that you are calling a travel agency and booking a ticket. Ask about departure time and stop along the way.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si jízdenku. Zeptejte se na cenu jízdenky případně na slevu pro děti. (?? *Make believe that you are calling a travel agency and booking a ticket. Ask about ticket price, maybe also for children discount.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si jízdenku. Zeptejte se, zda není Vámi požadované spojení již obsazené. (?? *Make believe that you are calling a travel agency and booking a ticket. Ask if the place for selected connection is available.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si jízdenku. Zeptejte se dostupnost konkrétního místa ve vlaku či autobuse podle Vašich požadavků. (?? *Make believe that you are calling a travel agency and booking a ticket. Ask about exact place in the train/bus.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si jízdenku. Zeptejte se, kdy je možné si jízdenku zakoupit. (?? *Make believe that you are calling a travel agency and booking a ticket. Ask when it is possible by the ticket.*)

F02: ?? Představte si, že voláte do cestovní kanceláře a zajišťujete si jízdenku. Zeptejte se na možnosti platby (hotově, platební kartou, šekem, apod.). (?? *Make believe that you are calling a travel agency and booking a ticket. Ask about payment conditions – in cash, by credit card, by cheque, etc.*)

### **F03 - Call a bank**

F03: ?? Představte si, že voláte do Vaší banky. Požadujte informace o zůstatku na Vašem účtu. (?? *Make believe that you are calling your bank. Ask about account information.*)

F03: ?? Představte si, že voláte do Vaší banky. Zadejte jednorázový platební příkaz pro převod na jiný účet. (?? *Make believe that you are calling your bank. Order money transfer to an account.*)

F03: ?? Představte si, že voláte do Vaší banky. Zadejte požadavek na prodej či nákup akcií, případně se zeptejte na ceny na burze. (?? *Make believe that you are calling your bank. Order to buy/sell of stocks, ask about stock prices.*)

F03: ?? Představte si, že voláte do Vaší banky. Požadujte informaci, zda konkrétní platba již přišla na Váš účet. (?? *Make believe that you are calling your bank. Ask if you have already received given money transfer.*)

F03: ?? Představte si, že voláte na poštu a zeptejte se, zda jste nedostali doporučený dopis. (?? *Make believe that you are calling the post-office. Ask if you have already received registered letter.*)

#### **F04 - Call a theatre**

F04: ?? Představte si, že voláte do divadla, kina, koncertní síně, restaurace či nočního klubu a zeptejte se na volná místa na určitý program. (?? *Make believe that you are calling theatre, cinema, concert hall, restaurant, or night club. Ask about available place for given program.*)

#### **F05 - Call a bookstore/music shop**

F05: ?? Představte si, že voláte do knihkupectví nebo hudebnin a zeptejte se na dostupnost a cenu určité knihy nebo CD. (?? *Make believe that you are calling book/music store. Ask about availability and price of a book/CD.*)

#### **F06 - Dictate a business letter**

F06: ?? Nadiktujte krátký obchodní dopis s informacemi o služební cestě Vašeho kolegy či zaměstnance. (?? *Dictate a short business letter with an information about business trip of your colleague or employee.*)

F06: ?? Nadiktujte krátký obchodní dopis s pozváním na pracovní schůzku. (?? *Dictate a short business letter with an invitation for negotiation.*)

F06: ?? Nadiktujte krátký obchodní dopis s pozváním na konferenci. (?? *Dictate a short business letter with an invitation for an conference.*)

F06: ?? Nadiktujte krátký obchodní dopis s objednávkou určitého zboží či služeb. (?? *Dictate a short business letter with an order of a goods or services.*)

#### **F07 - Call a hotel to make a change in the reservation**

F07: ?? Představte si, že voláte do hotelu. Potvrďte Vaši rezervaci a upozorněte na Váš pozdní příjezd. (?? *Make believe that you are calling into the hotel. Confirm your reservation and inform about your late arrival.*)

F07: ?? Představte si, že voláte do hotelu. Požádejte o změnu Vámi rezervovaného pokoje. (?? *Make believe that you are calling into the hotel. Ask for the change of the room in your reservation.*)

F07: ?? Představte si, že voláte do hotelu. Požádejte o změnu termínu Vaší rezervace. (?? *Make believe that you are calling into the hotel. Ask for the change of the date of your reservation.*)

#### **F08 - Give commands to speech-savvy mobile phone**

F08: ?? Řekněte Vašemu mobilnímu telefonu s hlasovým rozhraním požadavek na přečtení doručeného mailu resp. faxu. (?? *Ask your voiced-driven mobile-phone for reading of received mail or fax.*)

F08: ?? Řekněte Vašemu mobilnímu telefonu s hlasovým rozhraním požadavek na zkontrolování Vaší poštovní schránky (mailboxu). (?? *Ask your voiced-driven mobile-phone to check your mailbox..*)

F08: ?? Řekněte Vašemu mobilnímu telefonu s hlasovým rozhraním požadavek na zadání naplánované schůzky do diáře. (?? *Ask your voiced-driven mobile-phone to write an appointment into your schedule.*)

F08: ?? Řekněte Vašemu mobilnímu telefonu s hlasovým rozhraním požadavek na konferenční hovor s určitou osobou, Vaším úřadem a klientem. (?? *Ask your voiced-driven mobile-phone to organize conference call with your client.*)

#### **F09 - Description of the traffic**

F09: ?? Popište aktuální nebo předchozí dopravní situaci na silnici. (?? *Describe current or recent traffic situation.*)

#### **F10 - Description of a direction**

F10: ?? Popište policii resp. záchranářům Vaši aktuální pozici na silnici. (?? *Give to the police or rescue the description of your current position.*)

F10: ?? Popište Vašemu známému, jak se dostane autem k Vám domů. (?? *Describe to your friend how to go to your home.*)

F10: ?? Popište Vašemu známému, jak se dostane autem do vybrané restaurace. (?? *Describe to your friend how to go to given restaurant.*)



**F11 - Coordinates of a restaurant**

F11: ?? Zadejte navigačnímu systému v automobilu informace o Vaší oblíbené restauraci, tj. název, adresu, město. (?? Give to the car navigation system the coordinates of your favorite restaurant. )

**F12 - Questions for the route**

F12: ?? Zadejte navigačnímu systému v automobilu požadavek na cestu k nejbližšímu letišti, nádraží, parkovišti, garážím, apod. (?? Ask to the car navigation system for the route to the nearest airport, railway station, parking, garage, etc. )

**F13 - Instruments of the dashboard**

F13: ?? Jmenujte zleva doprava nástroje resp. zařízení na palubní desce Vašeho automobilu. (?? Name the instruments of the dashboard from left to right. )

**F14 - Voice controlled car equipment**

F14: ?? Jaké zařízení byste rád ovládal v automobilu hlasem? (?? What equipment would you like to use in the car by voice. )

**F15 - Technical details of the car**

F15: ?? Zeptejte se na určité technické detaily automobilu. (?? Ask about any technical details of the car. )

**F16 - Description of a meal**

F16: ?? Představte si, že jste v restauraci. Co byste si rád objednal ke snídani, k obědu nebo k večeři? (?? Make believe that you are in restaurant. What do you like to order for breakfast, lunch, or dinner?.)

**F17 - City to live in**

F17: ?? V jakém městě byste rád žil a proč? (?? Which city would you like to live in and why? )

**F18 - Places of interest**

F18: ?? Popište některá zajímavá místa ve Vašem městě. (?? Describe some interest place in your city.)

**F19 - Country to travel to**

F19: ?? Do jaké země byste se rád podíval a proč? (?? Which country would you like to travel and why? )

**F20 - Hotel reservation call**

F20: ?? Prosím, zarezervujte si po telefonu hotelový pokoj. (?? Please, make a hotel reservation call. )

**F21 - Route to a train station**

F21: ?? Popište, jak se dostat na vlakové nádraží. (?? Describe, how to go to the railway station.. )

**F22 - Favourite movie**

F22: ?? Popište stručně Váš oblíbený film. (?? Describe briefly your favorite movie. )

**F23 - Commands for speech driven radio**

F23: ?? Řekněte Vašemu Hi-Fi požadavek na přepnutí na určitou rozhlasovou stanici. (?? Ask your Hi-Fi to select given radio station. )

F23: ?? Řekněte Vašemu Hi-Fi požadavek na puštění kazety. (?? Ask your Hi-Fi to play cassette. )

F23: ?? Řekněte Vašemu Hi-Fi požadavek na puštění určitého CD. (?? Ask your Hi-Fi to play given CD. )

F23: ?? Řekněte Vašemu Hi-Fi požadavek na puštění konkrétní skladby z CD nebo kazety. (?? *Ask your Hi-Fi to play specific track from CD or cassette.* )

#### **F24 - Favourite program**

F24: ?? Řekněte Vaší televizi požadavek na přepnutí na určitý program. (?? *Ask your TV to select specific program.* )

#### **F25 - Favourite TV show**

F25: ?? Popište Váš oblíbený televizní program. (?? *Describe your favorite TV show.* )

#### **F26 - Audio/video devices**

F26: ?? Vyjmenujte všechna audio-video zařízení, která vlastníte. (?? *Name all audio and video devices you own.* )

#### **F27 - Content of TV show**

F27: ?? Zeptejte se na obsah daného programu v televizi. (?? *Ask about content of a TV show.* )

#### **F28 - Favourite sport**

F28: ?? Popište Váš oblíbený sport a sportovce. (?? *Describe your favorite sport and favorite sportsman.* )

#### **F29 - Others - What do you think about the recordings till this time?**

F29: ?? Jaký je doposud Váš pocit z nahrávání? (?? *What is your impression of the recordings so far?* )

#### **F30 - Others - Which device would you like to control by voice?**

F30: ?? Jaká zařízení byste rád ovládal hlasem? (?? *What kind of devices would you like to control by voice?* )

## **5.2 Elicited speech**

Following questions were used for prompting elicited speech. Total number of 17 elicited spontaneous items were recorded per adult speakers.

### **5.2.1 Birth date, current date, relative date (ED1-3)**

#### **- birth dates**

ED1: ?? Uveďte datum narození jedné osoby z Vaší rodiny (rodičů, sourozenců, dětí, prarodičů). (?? *Give any birthday of a person from your family (parents, brothers, sisters, children, grandparents)?* )

#### **- current date**

ED2: ?? Kolikátého je dnes? (?? *What date is today?* )

ED2: ?? Kdy má nebo měl v nejbližší době svátek či narozeniny někdo z Vašich blízkých? (Je možné použít relativní datum) (?? *When is any close birthday or name-day of somebody from your relatives? (You may use relative date.)* )

#### **- relative day**

ED3: ?? Kdy odjízíte na Vaši příští dovolenou? (Je možné použít relativní datum) (?? *When do you leave for your next vacation? (You may use relative date.)* )

### 5.2.2 Time of day (ET1-2)

ET1: ?? Kolik je právě teď hodin? (?? *What time is it now?* )

ET2: ?? V kolik hodin posloucháte Vaše oblíbené zprávy? (?? *At what time do you listen your favorite news?* )

ET2: ?? V kolik hodin přicházíte domů ze zaměstnání? (?? *At what time do you come home from your job?* )

### 5.2.3 City names (EC1-2)

EC1: ?? Ve kterém městě jste strávil většinu Vašeho života. (?? *In which city did you live longest time in your life?* )

EC2: ?? Ve kterém městě mimo Vaše současné bydliště žijí Vaši přátelé či příbuzní. (?? *Excepting the city where you live now, in which city do your friends or relatives live?* )

### 5.2.4 Proper names (EP1-3)

EP1: ?? Jak se jmenuje Váš otec nebo Vaše matka? (Je možné uvést jméno křestní, příjmení, či obojí.) (?? *What is the name of your father or your mother? (You may give name, surname, or both.)* )

EP2: ?? Uveďte jedno jméno Vašeho oblíbeného herce, zpěváka, spisovatele či jiné významné osobnosti. (?? *Give the name of your favorite actor, singer, writer, or other known person.* )

EP3: ?? Uveďte jméno nějaké společnosti či firmy. (?? *Give the name of any company or institution.* )

### 5.2.5 Spelling of proper name (EL1)

EL1: ?? Hláskujte jméno vašeho rodného města. (?? *Spell the name of the city where you were born.* )

EL1: ?? Hláskujte vaše křestní jméno. (?? *Spell your given name.* )

### 5.2.6 Yes/No answers (EQ1-2)

EQ1: ?? Snídal jste dnes ráno? (?? *Have you a breakfast this morning?* ) – preliminary “Yes”

EQ2: ?? Žil jste delší dobu v zahraničí? (?? *Did you live longer time abroad?* ) – preliminary “No”

### 5.2.7 Languages (EO1)

EO1: ?? Uveďte jméno jednoho jazyka, kterým mluvíte. (?? *Give the name of language which you speak.* )

EO1: ?? Uveďte jméno jednoho jazyka, kterým nemluvíte. (?? *Give the name of language which you do not speak.* )

## 5.2.8 Telephone numbers (EN1-3)

EN1-2: ?? Uved'te nějaké telefonní číslo do pevné sítě. (?? Give any telephone number within fixed network.)

EN3: ?? Uved'te nějaké telefonní číslo do mobilní sítě. (?? Give any telephone number within mobile (GSM) network.)

## 5.3 Read speech

Total number of items in the read prompting material is 275 (adult DB). The contents of the particular items is specified in the following sections.

### 5.3.1 Phonetically rich sentences (S01-S30)

#### Adult corpus (S01-S30)

The phonetically rich sentences were chosen out of 14095 sentences collected from Czech newspapers and several books from classical Czech writers available on the Internet. Above mentioned number of sentences is the result of the first pre-filtering which removed very long sentences, orthographically or grammatically strange sentences, etc. It was realized by a *Perl* script.

Further selection was made by *CorpusCrt*, which was downloaded from

<http://gps-tsc.upc.es/veu/personal/sesma/index.html>.

The sentences were read and corrected in case there was a grammatical or orthographic error or the content was not suitable. The corpus contains 3300 different sentences and each sentence is repeated 5 times.

The following phonemes are considered as rare: “o:, e\_u, d\_z, d\_Z, F“. However, we guaranteed that on the prompt level, each phoneme should be present in the set of phonetically rich sentences for each speaker.

Final statistics of phoneme appearance is presented in the chapter 10.

### 5.3.2 Phonetically rich words (W01-05)

The phonetically rich words were chosen to add into the corpus the words with phonemes which are rare in the phonetically rich sentences (o:, e\_u, d\_z, d\_Z, F), to guarantee their appearance also at the transcription level in the each session.

300 different words were chosen from a original corpus of 2798072 words from lexicon used in Linux *ispell* for Czech, which was reduced to 9484 words containing rare phonemes. The selection was made again by *CorpusCrt* as for phonetically rich sentences. All words should be correct Czech words, however, some words are used very rarely, some of them are more specialized (e.g. to some scientific area), some of them are usually not-used forms of verbs, etc. In case problems, the meaning of the word is explained to speaker.

The maximum of the appearance of one word in the whole database is ten.

Final statistics of phoneme appearance is presented in the chapter 10.

### 5.3.3 Isolated digits (CI1-4)

In each session four isolated digits are read. The frequency of occurrence of all tokens is (approximately) uniform. Language specific peculiarities are specified in following Table 17. The frequency of every spoken digit at prompt level is set uniformly to 137 per each digit (for total 16 variants of basic digits). The digits were presented to the speakers as words.

The minimum number of tokens per digit at transcription level is 117. These minimal representation was achieved for final 550 adult DB, excepting the variant “**čtyry**” which has just 113 representations – see common explanation note for CI, CB, and CC items.

0	nula
1	jedna
2	dva, dvě
3	tři
4	čtyři, čtyry, štyři, štyry
5	pět
6	šest
7	sedm, sedum
8	osm, osum
9	devět

Table 17 – List of basic digit variants

### 5.3.4 Isolated digit sequences (CB1, CB2)

The isolated digit sequence provides isolated digit examples of all digits in different positions. Two isolated digit sequences are recorded for each speaker. This is different from general specification where only one isolated digit sequence is prompted. Since the coverage of each variant is required, the number of digits in the sequence had to be increased. To avoid too long utterances with 16 digits per each, we have decided to split this item into two items CB1 and CB2.

The sequence of the digits is random. For each speaker, items CB1 and CB2 contain all variants mentioned above in Table 17. The digits are presented to the speaker as words. The minimum number of tokens (repetitions of prompted words) per digit at prompt level is set to 550.

The minimum required number of tokens per digit at transcription level was 487.

These required coverage of each token at transcription level was not reached for following digit variants with achieved representation: “**čtyry**” (448 appearances), “**štyři**” (449), “**sedm**” (413), “**osm**” (382) – see common explanation note for CI, CB, and CC items.

### 5.3.5 Digit strings (CC1-4)

Four continuous digit strings with together 20 digits are read. The digits are represented as words. The prompts ensure that each speaker provides at least one example of each digit. The digit strings were generated randomly. The number of tokens (repetitions of words) per digit at prompt level is set to 687.

The minimum number of tokens per digit at transcription level is 585.

These required coverage of each token at transcription level for CC items was not reached for following digit variants (with achieved representation): “čtyry” (549 appearances), “sedm” (517), “osm”(466) – see common explanation note for CI, CB, and CC items.

**NOTE – EXPLANATION: Problem – not achieved minimal representation of some digit variants at transcription level.**

However all digit variants are very well known and commonly used, sometimes they may be region dependent and for some people the pronunciation of exact variant may be very strange, especially when they see this variant *written* for the first time in life. Concerning this, many speakers have some problems with these items so to avoid very strange and un-natural pronunciation, they stick with the closest one, that they use. Further, the pronunciation of different variants of digits in one digit string (as it will always appear within CB1 and CB2) was found sometimes very difficult and people pronounced such string very hardly. With the requirement for pronunciation without pauses (CC items) it is even harder. Typically, in string pronunciation, people tend to pronounce variants “sedum” and “osum” instead of “sedm” and “osm”, etc.

**After several attempts with re-recordings of bad pronunciation of digit variant, we preferred finally to let such speaker pronounce it in his/her way.**

Finally obtained coverage in worst case is close to 80% of required representation, so digit variants are not extremely unbalanced. Moreover, other tokens at transcription level are available in additional sessions, mainly SES550 – SES567, i.e. the sessions with good signal quality, however, whole amount of missed tokens cannot be covered within these 18 sessions.

### 5.3.6 Telephone number (CE1)

This is a **read** 9-13 digit telephone number. The number of digits, spacing and presentation reflects the typical telephone numbers in the language for national numbers including area codes, international numbers, and cellular network numbers. This item was prompted in digit form.

Country code for Czech Republic is +420. International calls are invoked by dialing „00“.

Internally, all calls within Czech Republic, including GSM calls, are dialed by 9-digit number, excepting numbers with the first digit equal 1. Numbers may be divided into following groups:

- fixed network calls begin by 2,3,4,5,
- mobile network calls begin by 6,7,
- green and blue lines begin by 8,
- special services (internet connections, etc.) begin by 9.

No Czech number may begin by 0. However, several numbers beginning by „0“ or „00“ are used in the database because the zero often represents outside code for calls from company networks and 00 is a prefix for international calls.

The spacing and the place for the hyphen can vary, though the standard for numbers presented by Czech-Telecom in Yellow Pages is the sequence of three 3-digit numbers. This is used in DB for Czech national numbers. International numbers were collected from several WEB-pages and they are presented in different forms.

### 5.3.7 Natural numbers (CN1-3)

Three natural number phrases were read from the prompt sheet. The numbers chosen aimed to maximize the natural number word equivalents of: *10-19, 20...90, hundred, thousand*, and all their inflected variants.

The prompts covered numbers  $X \leq 10,000,000$ . Finally, 550 different natural numbers (including 60 decimal numbers with 4-5 significant digits) are prompted with enough examples of words to permit training of all lexical items.

Although the natural numbers higher than 1,000,000 are not recommended in the specification, we use them in minor amount, because in Czech such numbers are quite common (the price of houses in cities starts at approximately 1 mil CZK).

### 5.3.8 Money amounts (CM1)

The prompts elicit typical phrases used with money amounts, including the currency words ‘koruna’ and ‘halér’ with their inflected variants. Also abbreviation ‘Kč’ is used. Items are read and provided in orthographic form.

The items are a mixture of small (i.e. including decimal currency units) and larger money amounts (not including decimal currency units).

‘Euro’ and ‘Cent’ are used as foreign currency.

### 5.3.9 Times (CT1-2)

Two time phrases were read:

**T1: One phrase in analogue form** to provide adequate lexical coverage of all necessary words for training, including equivalents of (in English), see Table 18:

AM, PM, half, quarter, past, to, noon, midnight, morning, afternoon, evening, night, minutes, hours, o'clock, nearly, exactly, etc.

v noci	AM (period from 0:00 till 3:00)
ráno	AM (period from 3:00 till 9:00)
dopoledne	AM (period from 9:00 till 12:00)
odpoledne	PM (period from 12:00 till 6:00pm)
večer	PM(period from 6:00pm till 12:00pm)
asi	approximately   about
přesně	exactly
skoro	nearly
hodina	o'clock ( 1,21 o'clock ) (ATTENTION - DECLINED FORM)
hodiny	o'clock ( 2,3,4,22,23,24 o'clock ) (ATTENTION - DECLINED FORM)
hodin	o'clock ( 5-20 o'clock ) (ATTENTION - DECLINED FORM)
minuta	minute   minutes ( 1,21,31,41,51 minutes ) (ATTENTION - DECLINED FORM)
minuty	minutes ( 2,3,4,22,23,24,32,33,34,42,43,44, 52,53,54 minutes ) (ATTENTION - DECLINED FORM)
minut	minutes ( X minutes, others than above) (ATTENTION - DECLINED FORM)
poledne	midday   noon
půlnoc	midnight
čtvrt_na	a_quarter_past [hour-1] ( Followed by BASIC numeral in 4-th case )
půl	half_past [hour-1] ( Followed by ORDINAL numeral in 2-th case and female form )
tři_čtvrtě_na	a_quarter_to [hour] ( Followed by BASIC numeral in 4-th case )
a	and (conjunction)

Table 18 – List of most common Czech analogue time expressions

**T2: One phrase in digital form.**

The times were prompted as words, not as digits (e.g. five past ten).

Example: “pět hodin dvacet sedm minut”.

**5.3.10 Dates (CD1-3)**

**D1: analogue form.**

The analogue dates cover all weekday names and month names (uniformly distributed). An example of a prompt is the following:



čtvrtek, dvacátého devátého února, dva tisíce patnáct

Monday	pondělí
Tuesday	úterý
Wednesday	středa
Thursday	čtvrtek
Friday	pátek
Saturday	sobota
Sunday	neděle

Table 19 – List of Czech weekdays expressions

January	leden	ledna
February	únor	února
March	březen	března
April	duben	dubna
May	květen	května
June	červen	června
July	červenec	července
August	srpen	srpna
September	září	září
October	říjen	října
November	listopad	listopadu
December	prosinec	prosince

Table 20 – List of Czech month names

Note: Month names are used in declined form (2nd case of singular). These forms are summarized in the third column of Table 20 (basic form in the second column is not used).

## **D2: relative and general date expressions**

General and relative date expressions used in recording are summarized in the following Table 21.

Dnes	today (+0d from now)
Zítří	tomorrow (+1d from now)
Pozítří	the day after tomorrow (+2d from now)
Včera	yesterday (-1d from now)
Předevčím	the day before yesterday (-2d from now)
Víkend	weekend
pracovní_den	workday
tento_týden	this week (+0w from now)
příští_týden	next week (+1w from now)
minulý_týden	last week (-1w from now)
tento_měsíc	this month (+0 m from now)
příští_měsíc	next month (+1m from now)
minulý_měsíc	last month (-1m from now)
letos   tento_rok	this year (+0y from now)
vloni   minulý_rok	last year (-1y from now)

příští_rok	next_year (+1y from now)
Velký_pátek	Good_Friday (the Friday before Easter Sunday)
Velikonoční_neděle   Boží_hod_velikonoční	Easter_Sunday
Velikonoční_pondělí	Monday after Easter Sunday
Štědrý_den	Christmas_Eve (24 <sup>th</sup> of December)
První_svátek_vánoční   Boží_hod_vánoční	(First) Christmas_Day   (25 <sup>th</sup> of December)
Silvestr	New_Year's_Eve   31st of December
Nový_rok	New_Year   1st of January
První_máj   Svátek_práce	Labour_Day   1st of May
Den_matek	Mother's_Day
Den_děti	Children's day
Narozeniny	Birthday
Svátek	Name day
Osvobození_republiky	Liberty of the Republic (end of the II. world war)   8th of May
Cyrila_a_Metoděje	Day of St Cyril and St Metodej   5th of July
Mistra_Jana_Husa	Day of Johannes Hus   6th of July (1415)
Svatého_Václava	Day of St Venceslav (Czech duke)   28th of September
Vznik_republiky	Czechoslovak Republic establishment   28th of October (1918)
Sametová_revoluce	Velvet revolution   17th of November (1989)

Table 21 – List of general and relative date expressions

### D3: digital form.

The digital dates cover all month names (uniformly distributed) without weekday specification because the weekday cannot be specified digitally in Czech.

An example for a prompt is the following:

dvacátého devátého třetí, dva tisíce patnáct

January	první
February	druhý
March	třetí
April	čtvrtý
May	pátý
June	šestý
July	sedmý
August	osmý
September	devátý
October	desátý
November	jedenáctý
December	dvanáctý

Table 22 – List of names for Czech month names in digital form

### 5.3.11 Spelled words (CL1-3)

Two of the three spelled words are natural (e.g. proper names, city names). The third word is an artificial sequence which is used to enhance a more uniform distribution of letters.

The handling of Czech spelling has the following specialities.

1. “ch” is in Czech one letter and it has its own spelling variant.
2. Czech pronunciation is quite regular so people are not used to spell words. Consequently, though there are the official spelling rules for each letter, i.e. “á, bé, cé, čé, dé, dě, é, ef, ..., zet, žet”, it is very often that people spell more simply without vowel before or after consonant, i.e. “a, b, c, č, d, d', e, f, ..., z, ž”. Consonants in the second variants are usually followed by brief schwa. This is so called spelling using phonetic realization of the letter, as it is described in annotation rules for SpeeCon .
3. It is difficult to resolve short and long variant of vowels just by spelling, since both short and long pronunciation may appear in the spelling of both variants of vowels, i.e. the problem may be the recognition of „A“ and „/Á“.
4. Letter with accents are sometimes spelled with other additional words as “dlouhé (long), s čárkou (with accent), krátké (short), přehlasované (with umlaut), s kroužkem (with circle) ...”. Although the occurrence of each letter is uniform in prompts, statistics done on transcription will be different because long vowels are usually spelled as “long A”. In fact, this increases occurrence of short vowel spellings and decreases occurrence of pure long vowel spellings. Similar case (however more rare) may appear also for accented consonants.
5. German letters “ä, ö, ü” may appear in names of German origin (Müller, etc.).

The spelling alphabet with annotation and general pronunciation is described in the section 9 in the Table 43 and Table 44. The following letters should be spelled:

A | Á | B | C | Č | D | Ď | E | É | Ě | F | G | H | CH | I | Í  
| J | K | L | M | N | Ň | O | Ó | P | Q | R | Ř | S | Š | T |  
Ě | U | Ú | Ů | U | W | X | Y | Z | Ž | + ( Ä | Ö | Ü )

German letters (Ä, Ö, Ü) do not appear in artificial sequences of letters because these letters do not exist in purely Czech words. They usually appear in names of German origin, then they may appear also in spelling items.

### 5.3.12 Proper Name (CP1)

The proper name was drawn from a set of **150** first or last names or a combination of them. Complete list of used proper names is in the following Table 23. It was generated from 75 most frequent male names and surnames (one half). The second half was generated from 75 most frequent female names and surnames. Note. Surname for males and females is different in Czech language.

Alena Hájková	Aleš Kratochvíl	Anna Čechová	Antonín Pokorný
---------------	-----------------	--------------	-----------------

Bedřich Kopecký	Stanislava Sedláčková	Iva Matoušková	Miroslav Zeman
Blanka Šťastná	Stanislav Kovář	Ivana Novotná	Oldřich Blažek
Bohumila Vávrová	Svatopluk Procházka	Ivona Blažková	Olga Říhová
Bohuslav Černý	Věra Vaňková	Jana Svobodová	Ondřej Krejčí
Božena Bláhová	Vladimír Musil	Jan Staněk	Otto Dušek
Břetislav Bárta	Vlasta Kučerová	Jaroslav Soukup	Petra Kovářová
Daniel Šťastný	Vlastimil Dvořák	Jindra Müllerová	Petr Urban
Eva Mašková	Vratislav Doležal	Jindřich Novák	Radek Vlček
Hana Fišerová	Zbyněk Fiala	Jindřiška Staňková	Radka Strnadová
Ivan Kučera	Adolf Havlíček	Jiří Pospíšil	Radomír Jelínek
Iveta Němcová	Alexandr Růžička	Jitka Procházková	Robert Sýkora
Ivo Kašpar	Alois Svoboda	Josefa Konečná	Romana Šimková
Jarmila Malá	Alžběta Doležalová	Josef Král	Roman Horák
Jaromír Mach	Anežka Tichá	Kamil Novotný	Rostislav Říha
Jaroslava Soukupová	Antonie Horáková	Kateřina Králová	Rudolf Beran
Jiřina Marková	Arnošt Kříž	Květa Moravcová	Růžena Růžičková
Julie Čermáková	Blažena Musilová	Lada Vlčková	Soňa Dušková
Karel Hruška	Bohumil Mašek	Ladislava Poláková	Šárka Krejčová
Květoslava Fialová	Bohumír Ševčík	Leopold Tůma	Štefan Hrubý
Ladislav Bureš	Čestmír Tichý	Libor Müller	Štěpánka Holubová
Lenka Dvořáková	Dagmar Kolářová	Lubomír Štěpánek	Štěpán Konečný
Libuše Urbanová	Dalibor Vávra	Luděk Polák	Tomáš Bartoš
Luboš Němec	Dana Kadlecová	Ludmila Sedláková	Václav Bednář
Marek Holub	Daniela Bartošová	Ludvík Vacek	Viktor Jaroš
Markéta Veselá	Danuše Burešová	Marcela Pokorná	Vilém Sedláček
Michal Kolář	Drahomíra Vacková	Marie Tůmová	Vítězslav Valenta
Milada Pešková	Dušan Kadlec	Marta Jandová	Vít Marek
Monika Černá	Eduard Čech	Martina Zemanová	Vladimíra Jelínková
Naděžda Machová	Eliška Pospíšilová	Martin Hájek	Vladislav Čermák
Otakar Moravec	Emilie Ševčíková	Michaela Hrušková	Vojtěch Veselý
Pavel Šimek	Emil Ježek	Milan Dostál	Zděna Křížová
Pavla Štěpánková	Evžen Navrátil	Milena Kopecká	Zdeněk Beneš
Pavčina Hrubá	František Vaněk	Miloslava Navrátilová	Zdenka Benešová
Přemysl Malý	Františka Beranová	Miloslav Matoušek	Zdeňka Nováková
Radim Sedlák	Gabriela Sýkorová	Miloš Bláha	Zuzana Němečková
Renata Kratochvílová	Helena Valentová	Miluše Havlíčková	
Richard Mareš	Irena Dostálová	Miroslava Marešová	

Table 23 – List of used proper names

### 5.3.13 Cities (CO1)/street names (CO2)

1 item drawn from a set of 275 most frequent cities, see Table 24.

1 item drawn from a set of 275 most frequent street names, see Table 25.

Praha	Pardubice	Zlín	Karlovy Vary
Brno	Hradec Králové	Olomouc	Karviná
Ostrava	České Budějovice	Liberec	Kladno
Plzeň	Ústí nad Labem	Havířov	Jihlava

Teplice	Šumperk	Třeboň	Týn nad Vltavou
Most	Otrokovice	Kaplice	Frenštát pod Radhoštěm
Frýdek-Místek	Sokolov	Slaný	Stříbro
Jablonec nad Nisou	Domažlice	Velké Meziříčí	Soběslav
Třebíč	Valašské Meziříčí	Semily	Zábřeh
Znojmo	Svitavy	Dobříš	Slavkov u Brna
Prostějov	Veselí nad Moravou	Vysoké Mýto	Rýmařov
Přerov	Kadaň	Tišnov	Jeseník
Opava	Říčany	Neratovice	Adamov
Mladá Boleslav	Beroun	Polička	Čelákovice
Tábor	Holešov	Bílina	Duchcov
Česká Lípa	Ústí nad Orlicí	Rosice	Šternberk
Příbram	Poděbrady	Hořovice	Bystřice nad Pernštejnem
Chomutov	Uherský Brod	Tachov	Mnichovo Hradiště
Trutnov	Mariánské Lázně	Mikulov	Kuřim
Děčín	Louny	Hustopeče	Světlá nad Sázavou
Kolín	Ostrov	Litomyšl	Nejdek
Žďár nad Sázavou	Český Těšín	Roudnice nad Labem	Luhačovice
Uherské Hradiště	Vrchlabí	Dobruška	Litovel
Hodonín	Rokycany	Nové Město nad Metují	Rumburk
Kroměříž	Jirkov	Broumov	Napajedla
Strakonice	Bohumín	Jilemnice	Dačice
Chrudim	Prachatice	Dobětice	Staré Město
Orlová	Turnov	Moravské Budějovice	Frýdlant nad Ostravicí
Náchod	Žatec	Vimperk	Votice
Havlíčkův Brod	Český Krumlov	Sezimovo Ústí	Český Brod
Písek	Bruntál	Holice	Varnsdorf
Blansko	Kyjov	Sušice	Pacov
Třinec	Rakovník	Moravská Třebová	Doksy
Kutná Hora	Rožnov pod Radhoštěm	Blatná	Kojetín
Nový Jičín	Čáslav	Nový Bydžov	Ledeč nad Sázavou
Vyškov	Jaroměř	Valašské Klobouky	Studénka
Klatovy	Jaroměř	Valašské Klobouky	Úvaly
Jindřichův Hradec	Vlašim	Veselí nad Lužnicí	Kunovice
Cheb	Kopřivnice	Kláštorec nad Ohří	Frýdlant v Čechách
Břeclav	Hlinsko	Bystřice pod Hostýnem	Dubí
Vsetín	Rychnov nad Kněžnou	Sedlčany	Hulín
Hranice	Kněžnou	Nová Paka	Meziboří
Pelhřimov	Brandýs nad Labem	Hlučín	Bílovec
Litvínov	Krnov	Chodov	Šlapanice
Dvůr Králové nad Labem	Boskovice	Aš	Mimoň
Mělník	Krupka	Lososice	Kdyně
Kralupy nad Vltavou	Česká Třebová	Lovosice	Telč
Benešov	Přelouč	Žamberk	Chropyně
Litoměřice	Chotěboř	Ivančice	Vítkov
Jičín	Chotěboř	Ivančice	Benátky nad Jizerou
Nymburk	Hořice	Nové Město na Moravě	
	Humpolec		

Moravský Krumlov	Horšovský Týn	Lázně Bohdaneč	Volary
Příbor	Valtice	Hrádek nad Nisou	Stará Boleslav
Blovice	Hluk	Jáchymov	Střelice
Kamenice nad Lipou	Lomnice nad	Chrastava	Újezd
Česká Skalice	Popelkou	Vrbno pod Pradědem	Lipník nad Bečvou
Starý Plzenec	Letovice	Uhlířské Janovice	Křenovice
Bojkovice	Plasy	Trhové Sviny	Hřebeč
Hronov	Lysá nad Labem	Skuteč	Vamberk
Vodňany	Prostřední Suchá	Týniště nad Orlicí	Bělá pod Bezdězem
Velešín	Toužim	Vratimov	Kraslice
Březová	Zruč nad Sázavou	Habartov	Plešivec
Červený Kostelec	Štětí	Židlochovice	Lanžhot
Bechyně	Uničov	Pohořelice	Hostinné
Rychvald u Karviné	Odolena Voda	Velké Pavlovice	Bzenec
Dobřany	Chlumeč	Jemnice	Heřmanův Městec
Třešť	Nová Ves	Planá	Úpice
Kostelec nad Orlicí	Františkovy Lázně	Nové Strašecí	Hrádek
Choceň	Třemošná	Vizovice	
Nová Role	Dobřichovice	Řevnice	

Table 24 – List of 275 cities

Komenského	Kosmonautů	Haškova	Výškovická
Nádražní	Nová	Studentská	Oblá
Palackého	Plzeňská	Lutyně	Vítězná
Sokolovská	Luční	Severní	Jabloňová
Pražská	Sadová	Tylova	Opavská
Družstevní	Revoluční	Moravská	Švermova
Husova	Polní	Brožíkova	Jižní
Jiráskova	Gagarinova	Seifertova	Polská
Tyršova	Jungmannova	Na vyhlídce	Dobrovského
Havlíčkova	U stadionu	Vinohradská	K. Čapka
Lidická	Mírová	Poruba	Purkyňova
Okružní	Dělnická	Hornická	Riegrova
Mánesova	Štefánikova	Máchova	E. Krásnohorské
Zahradní	Hlavní	Budovatelů	Sportovní
Nerudova	Slezská	Moskevská	Vančurova
Masarykova	Brněnská	Zborovská	Mlýnská
Bezručova	Koněvova	Dolní	Žitná
Čs. armády	1. máje	kpt. Jaroše	Hlavní tř.
Sídliště	Březinova	Bělehradská	Krátká
Dukelská	Čechova	Na výsluní	Čajkovského
Smetanova	Sokolská	Bulharská	Čtyři Dvory
Dlouhá	Novodvorská	Evropská	SNP
17. listopadu	Lípová	Jablonecká	Růžová
Školní	Ruská	Erbenova	M. Majerové
Dvořákova	Lesní	Francouzská	Fučíkova
Vrchlického	Kollárova	Fibichova	Čapkova
Žižkova	Větrná	Táborská	Písečná
B. Němcové	Wolkerova	Zelená	Česká

Budovatelská	Fügnerova	sídl. Vajgar	Trnovanská
Slovenská	Biskupcova	Zámecká	Havanská
Korunní	Dukelských hrdinů	Spojovací	A. Staška
Stavbařů	Americká	Alešova	V Olšínách
Španielova	Vysočanská	Vondroušova	Bořivojova
Brandlova	Jizerská	Pionýrů	bří. Čapků
Alšova	Lužická	Západní	Pekařská
Tolstého	Na Petřínách	Javorová	Jílovská
Janského	Vojanova	Štěpnická	28. pluku
Průběžná	Trávníky	Bzenecká	Palachova
Tovární	Lábkova	B. Martinů	Březová
Hálkova	Jeseniova	Budějovická	Partyzánská
Rooseveltova	Černokostelecká	Resslova	P. Jilemnického
M. Horákové	Kubelíkova	Ostravská	Brechtova
Merhautova	Počernická	Krymská	Závodní
Vršovická	tř. E. Beneše	Štichova	nám. Míru
Pod lipami	Poděbradova	Nezvalova	Mozartova
Bělohorská	Bellušova	Jílová	Rumunská
J. Palacha	Raisova	Pujmanové	Přecechtělova
Jugoslávská	Černého	Svojsíkova	Jateční
Hradecká	Gorkého	Bieblova	Tachovská
Křížikova	Sládkova	Chomutovská	Jesenická
Chelčického	Volgogradská	Nevanova	Hodonínská
Mládežnická	Čelakovského	Lamačova	Slavíkova
Jasmínová	Poznaňská	Palackého tř.	Žerotínova
Obránců míru	Podlesí	Školská	Záhřebská
Přátelství	Kostelní	Rokycanova	nám. Republiky
Jaselská	Šafaříkova	Kyjevská	Poděbradská
Heyrovského	Pod strání	Hybešova	Veletřní
Šumavská	Opletalova	Rezlerova	Škroupova
Osvobození	Ciolkovského	Zelenohorská	Skupova
Ukrajinská	Sv. Čecha	S. K. Neumanna	Levského
Janáčkova	Úvoz	Makovského	Míru
Kmochova	Příčná	Roháčova	Dvorská
5. května	Jičínská	Majakovského	Vídeňská
T. G. Masaryka	Baarova	Chodská	Sladkovského
J. Zajíce	Kamenická	Kladenská	Žlutická
Údolní	kpt. Nálepky	Blatenská	
Slunečná	Lucemburská	Arbesova	
Na kopci	Frýdlantská	Šeříková	
Horní	Holečkova	Horská	
Masarykovo nám	Třebízského	Slavičkova	

Table 25 – List of 275 street names

### 5.3.14 Yes/No (CQ1-2)

Two items (one Yes, one No) were recorded for each speaker. The following yes/no expressions were used:

Ano	yes
Ne	no

Table 26 – Yes/no expressions

### 5.3.15 Email & Web addresses (CW1-2)

**CW1:** WWW address from a list of 150 URL-addresses, see Table 27

**CW2:** Email drawn from a set of 550 e-mail addresses, see Table 28.

<a href="http://olc3.feld.cvut.cz/ak/jednotky.phtml">http://olc3.feld.cvut.cz/ak/jednotky.phtml</a>	<a href="http://www.pedf.cuni.cz">http://www.pedf.cuni.cz</a>
<a href="http://home.nextra.sk">http://home.nextra.sk</a>	<a href="http://www.msmt.cz">www.msmt.cz</a>
<a href="http://www.pocitace.cx">http://www.pocitace.cx</a>	<a href="http://www.pbko.sk">http://www.pbko.sk</a>
<a href="http://www.hackertrickkiste.de">www.hackertrickkiste.de</a>	<a href="http://www.amd.com">http://www.amd.com</a>
<a href="http://www.littleigloo.org">www.littleigloo.org</a>	<a href="http://ilectric.com">http://ilectric.com</a>
<a href="http://slashdot.org">slashdot.org</a>	<a href="http://www.libranet.com">www.libranet.com</a>
<a href="http://bigbook.com">http://bigbook.com</a>	<a href="http://www.mucl.cz">http://www.mucl.cz</a>
<a href="http://www.palmsoft.cz">www.palmsoft.cz</a>	<a href="http://www.mpsv.cz/scripts/default.asp">www.mpsv.cz/scripts/default.asp</a>
<a href="http://www.marketing1to1.com">http://www.marketing1to1.com</a>	<a href="http://www.bbc.co.uk">http://www.bbc.co.uk</a>
<a href="http://aldebaran.feld.cvut.cz">http://aldebaran.feld.cvut.cz</a>	<a href="http://www.earthlink.com">www.earthlink.com</a>
<a href="http://www.scoreworks.com">www.scoreworks.com</a>	<a href="http://amber.feld.cvut.cz/bio/program460.htm">amber.feld.cvut.cz/bio/program460.htm</a>
<a href="http://linuxtoday.com">http://linuxtoday.com</a>	<a href="http://www.volweb.cz/gohradni/spojeni.htm">www.volweb.cz/gohradni/spojeni.htm</a>
<a href="http://www.reuters.com">http://www.reuters.com</a>	<a href="http://www.moreorless.au.com/heroes/slovo.htm">http://www.moreorless.au.com/heroes/slovo.htm</a>
<a href="http://www.lutlinux.com">http://www.lutlinux.com</a>	<a href="http://images.google.com">http://images.google.com</a>
<a href="http://www.altap.cz/salam_cz/newver.html">www.altap.cz/salam_cz/newver.html</a>	<a href="http://www.overture.com">http://www.overture.com</a>
<a href="http://www.volny.cz/vorisekd/archiv.htm">www.volny.cz/vorisekd/archiv.htm</a>	<a href="http://www.digisound.cz">www.digisound.cz</a>
<a href="http://dewil.ic.cz/clanky/gsm-karty.htm">dewil.ic.cz/clanky/gsm-karty.htm</a>	<a href="http://www.chatavatra.cz">http://www.chatavatra.cz</a>
<a href="http://www.soyo.com">www.soyo.com</a>	<a href="http://www.slackware.com">www.slackware.com</a>
<a href="http://www.dinonet.net">www.dinonet.net</a>	<a href="http://www.interbed.com">www.interbed.com</a>
<a href="http://www.exploratorium.edu">http://www.exploratorium.edu</a>	<a href="http://www.oncolink.com">http://www.oncolink.com</a>
<a href="http://www.ups.com">www.ups.com</a>	<a href="http://www.mmo.cz/far/odd_skol.html">www.mmo.cz/far/odd_skol.html</a>
<a href="http://www.slovo.nn.ru">www.slovo.nn.ru</a>	<a href="http://www.letnany.cz/slovo/default.asp">www.letnany.cz/slovo/default.asp</a>
<a href="http://www.info211.sk">www.info211.sk</a>	<a href="http://www.rsa.com">http://www.rsa.com</a>
<a href="http://meteor.fzu.cz">meteor.fzu.cz</a>	<a href="http://www.sfsite.com">www.sfsite.com</a>
<a href="http://www.kvarnerbanka.hr">www.kvarnerbanka.hr</a>	<a href="http://www.fotostar.cz">www.fotostar.cz</a>
<a href="http://www.tiscali.cz">www.tiscali.cz</a>	<a href="http://www.webzdarma.cz">http://www.webzdarma.cz</a>
<a href="http://slovo.and.ru">slovo.and.ru</a>	<a href="http://www.eff.org">http://www.eff.org</a>
<a href="http://google.icq.com">google.icq.com</a>	<a href="http://www.bayerdiag.com">http://www.bayerdiag.com</a>
<a href="http://slovo.freecd.sk">slovo.freecd.sk</a>	<a href="http://www.slovo.cz">http://www.slovo.cz</a>
<a href="http://lemouillour.eric.free.fr">lemouillour.eric.free.fr</a>	<a href="http://www.rpmfind.net">www.rpmfind.net</a>
<a href="http://http://bazar.technet.cz">http://bazar.technet.cz</a>	<a href="http://www.visa.com">http://www.visa.com</a>
<a href="http://www.zona.cz">www.zona.cz</a>	<a href="http://www.otp.sk">www.otp.sk</a>
<a href="http://www.perina.net">www.perina.net</a>	<a href="http://katalog.centrum.sk">http://katalog.centrum.sk</a>
<a href="http://www.vareni.cz">www.vareni.cz</a>	<a href="http://www.astro.cz/apod/ap010112.html">www.astro.cz/apod/ap010112.html</a>
<a href="http://www.newsday.com">www.newsday.com</a>	<a href="http://www.linux.org">www.linux.org</a>
<a href="http://www.mff.cuni.cz">http://www.mff.cuni.cz</a>	<a href="http://www.cnet.com">www.cnet.com</a>
<a href="http://www.money.com">www.money.com</a>	<a href="http://www.abisource.com">http://www.abisource.com</a>



<a href="http://www.gbkr.si">http://www.gbkr.si</a>	<a href="http://www.kcell.kz/site/telefon_eng.html">www.kcell.kz/site/telefon_eng.html</a>
<a href="http://www.skb.si">www.skb.si</a>	<a href="http://www.tel.hr">www.tel.hr</a>
<a href="http://www.mojenoviny.cz/zahranici/neris20021202F00555.html">http://www.mojenoviny.cz/zahranici/neris20021202F00555.html</a>	<a href="http://www.volny.cz/vorisekd/odkazy.htm">http://www.volny.cz/vorisekd/odkazy.htm</a>
<a href="http://www.investorsolutions.com">http://www.investorsolutions.com</a>	<a href="http://www.scour.com">www.scour.com</a>
<a href="http://www.accuweather.com">www.accuweather.com</a>	<a href="http://www.finweb.com">www.finweb.com</a>
<a href="http://www.adobe.com">www.adobe.com</a>	<a href="http://www.freevibe.com/mj/lowdown.shtml">http://www.freevibe.com/mj/lowdown.shtml</a>
<a href="http://www.mudrc.com/mobily/bazar.php">http://www.mudrc.com/mobily/bazar.php</a>	<a href="http://www.opera.com/download/modules.html">www.opera.com/download/modules.html</a>
<a href="http://www.davidicke.com">www.davidicke.com</a>	<a href="http://www.visto.com">http://www.visto.com</a>
<a href="http://www.wired.com">http://www.wired.com</a>	<a href="http://www.destinacie.sk/casopisy/casopisy.html">http://www.destinacie.sk/casopisy/casopisy.html</a>
<a href="http://www.epa.gov">www.epa.gov</a>	<a href="http://www.semily.cz/radnice/mu_odbory_osksv.htm">http://www.semily.cz/radnice/mu_odbory_osksv.htm</a>
<a href="http://www.bankofalbania.org">www.bankofalbania.org</a>	<a href="http://www.vtourist.com">http://www.vtourist.com</a>
<a href="http://www.novinky.cz">www.novinky.cz</a>	<a href="http://www.pbz.hr">www.pbz.hr</a>
<a href="http://www.mssch.cz/rozcestnik/skolstvi.html">http://www.mssch.cz/rozcestnik/skolstvi.html</a>	<a href="http://www.gotoworld.com">http://www.gotoworld.com</a>
<a href="http://www.kiplinger.com">http://www.kiplinger.com</a>	<a href="http://www.oalib.cz">http://www.oalib.cz</a>
<a href="http://www.sciencemag.org">http://www.sciencemag.org</a>	<a href="http://portal.opera.com">portal.opera.com</a>
<a href="http://www.inc.com">www.inc.com</a>	<a href="http://www.bsi.si">www.bsi.si</a>
<a href="http://www.investoralert.com">http://www.investoralert.com</a>	<a href="http://www.enlightenment.org">www.enlightenment.org</a>
<a href="http://www.fme.vutbr.cz">http://www.fme.vutbr.cz</a>	<a href="http://www.bastyr.edu">http://www.bastyr.edu</a>
<a href="http://slovo.newton.cz">slovo.newton.cz</a>	<a href="http://www.soyo.com.tw">www.soyo.com.tw</a>
<a href="http://www.webdevelopersjournal.com">http://www.webdevelopersjournal.com</a>	<a href="http://www.gama.fsv.cvut.cz">http://gama.fsv.cvut.cz</a>
<a href="http://www.sagar.cz/cesty/zast.phtml">www.sagar.cz/cesty/zast.phtml</a>	<a href="http://www.armed.net">www.armed.net</a>
<a href="http://www.who.int">www.who.int</a>	<a href="http://amber.feld.cvut.cz/bio/diskuse.asp">amber.feld.cvut.cz/bio/diskuse.asp</a>
<a href="http://pdftohtml.sourceforge.net">pdftohtml.sourceforge.net</a>	<a href="http://www.virginmobile.com">http://www.virginmobile.com</a>
<a href="http://www.genhomepage.com">http://www.genhomepage.com</a>	<a href="http://www.comcentral.com">www.comcentral.com</a>
<a href="http://www.mmo.cz">www.mmo.cz</a>	<a href="http://www.inzercemorava.cz">http://www.inzercemorava.cz</a>
<a href="http://vesmir.cts.cuni.cz/97-slepice/slepice2.htm">vesmir.cts.cuni.cz/97-slepice/slepice2.htm</a>	<a href="http://www.eastview.com">www.eastview.com</a>
<a href="http://login.client.tmo.cz">login.client.tmo.cz</a>	<a href="http://www.pf.jcu.cz">http://www.pf.jcu.cz</a>
<a href="http://www.ms.mff.cuni.cz">www.ms.mff.cuni.cz</a>	<a href="http://www.mobily.net">http://www.mobily.net</a>
<a href="http://rhn.redhat.com/errata/RHSA-2000-028.html">rhn.redhat.com/errata/RHSA-2000-028.html</a>	<a href="http://groups.google.com">http://groups.google.com</a>
<a href="http://www.lesstif.org">http://www.lesstif.org</a>	<a href="http://www.linux.com">www.linux.com</a>
<a href="http://www.startech.cz">http://www.startech.cz</a>	<a href="http://www.lateko.lv">www.lateko.lv</a>
<a href="http://http://db.zde.cz">http://db.zde.cz</a>	<a href="http://boxoff.com">boxoff.com</a>
<a href="http://shop.vscom.cz">shop.vscom.cz</a>	<a href="http://www.totem.cz">www.totem.cz</a>
<a href="http://www.bateria.cz">http://www.bateria.cz</a>	<a href="http://www.nerve.com">www.nerve.com</a>
<a href="http://www.everton.com">http://www.everton.com</a>	<a href="http://www.News365.com">http://www.News365.com</a>
<a href="http://www.baznet.freemove.co.uk">http://www.baznet.freemove.co.uk</a>	<a href="http://www.bezvaportal.cz">www.bezvaportal.cz</a>
<a href="http://www.chicagotribune.com">http://www.chicagotribune.com</a>	
<a href="http://www.autobytel.com">www.autobytel.com</a>	
<a href="http://ahasweb.webpark.cz">http://ahasweb.webpark.cz</a>	

Table 27 – List of WEB addresses

<a href="mailto:lukas.szemla@anfdata.cz">lukas.szemla@anfdata.cz</a>	<a href="mailto:p.j.hix@larc.nasa.gov">p.j.hix@larc.nasa.gov</a>
<a href="mailto:jirka@itc.cz">jirka@itc.cz</a>	<a href="mailto:chsthaaa@yahoo.co.kr">chsthaaa@yahoo.co.kr</a>
<a href="mailto:m.schejbal@sh.cvut.cz">m.schejbal@sh.cvut.cz</a>	<a href="mailto:hubicka@tabor.ta.jcu.cz">hubicka@tabor.ta.jcu.cz</a>
<a href="mailto:ghasrsaz@yahoo.com">ghasrsaz@yahoo.com</a>	<a href="mailto:lorenzp@fel.cvut.cz">lorenzp@fel.cvut.cz</a>

mikolandovi@raz-dva.cz	kas@fi.muni.cz
pavel@janik.cz	xserak@fi.muni.cz
xmartan@fi.muni.cz	felzmann@osu.cz
r.varkocek@sh.cvut.cz	platon@pf.jcu.cz
xneveril@fi.muni.cz	jaredh@innovsys.com
sojka@daeron.ics.muni.cz	milan@tmapy.anet.cz
hans@matfyz.cz	rftyj@yahoo.com
sima@aipsafe.cz	leda@leda.cz
jirka.honcu@seznam.cz	cstex-mgr@cs.felk.cvut.cz
tykim@ispl.korea.ac.kr	amdal@tele.ntnu.no
tuerk@phonetik.uni-muenchen.de	toman@artax.karlin.mff.cuni.cz
annick.leroy@irisa.fr	w@linux.fjfi.cvut.cz
podlesak@its.cz	tanel.alumae@aqris.com
dundr@raz-dva.cz	andilek@sh.cvut.cz
aldowsary@hotmail.com	amubadal@lucent.com
risa@gsl.kralupy.cz	free.survey@relay.felk.cvut.cz
sarsonj@upol.cz	lukas_voborsky@kb.cz
ltuuri@cc.helsinki.fi	nguyentindung@hotmail.com
zlatuska@anxur.fi.muni.cz	dusek@port.troja.mff.cuni.cz
xserak@fi.munii.cz	krysar@fykos.mff.cuni.cz
afrox@centrum.cz	vaclav.dvorsky@cssz.cz
mensa@cs.felk.cvut.cz	marina@nero.uni-bonn.de
digitalkoksal@hotmail.com	ee_awh@stu.ust.hk
nii.kommey@berlin.de	kuceram@risc.upol.cz
rdedic@mail.kolej.mff.cuni.cz	uptex-l@risc.upol.cz
tcajtham@students.zcu.cz	eespina@eee.upd.edu.ph
chaoxs@sat.mdc.tsinghua.edu.cn	rouha@iname.com
peter.podlucky@siemens.com	adamcik.inecom@brn.pvt.net.cz
paluch@terezka.ufa.cas.cz	mihokova@fzu.cz
praise@laud.felk.cvut.cz	pavel.pokorny@vscht.cz
csean@cs.jhu.edu	sladky@jablotron.cz
lipann@singnet.com.sg	nikola@fykos.mff.cuni.cz
gemper@fzu.cz	webacc@math.muni.cz
michalg@cmail.cz	info@zastudena.cz
arungopal@telstra.com	peter.podlucky@siemens.at
rob@cc.umist.ac.uk	service@allytrade.cz
rusiriwan@hotmail.com	mikulas.gangur@fek.zcu.cz
paul.warren@vuw.ac.nz	michal.hekrdla@bewotec.cz
michal.poupa@aliatel.cz	tasos.kounoudes@ddl.co.uk
vilag@look.ca	michal@mobil.cz
xkumpost@fi.muni.cz	xrehulka@fi.muni.cz
farkas@fzu.cz	obch.pratele@all.cz
ivonez@icpf.cas.cz	deleni@eagle.vscht.cz
somi@imc.cas.cz	sunil@asp.ogi.edu
pestonji@vsnl.com	lycka@isibrno.cz
rusek@gybon.cz	kas@foresta.cz
mnavratil@fr.vsp.cz	pjanik@inet.cz
rybicka@dahlia.vszbr.cz	petr.cach@atlas.cz

richard.korinek@seznam.cz	listproc@utia.cas.cz
www-admin@estug.cz	m.hnatek@sh.cvut.cz
tomas.mraz@pvt.cz	mestan@zce.cz
jantoniohm@hotmail.com	ubik@fsid.cvut.cz
vpasx@uom.gr	hnojil@geodesie.fsv.cvut.cz
lazarekm@seznam.cz	sweetkiss@wanadoo.es
nawaz_sadaf@hotmail.com	vladimir.naprstek@scplyn.cz
pepa.svob@worldonline.cz	walder@itee.uq.edu.au
pirdos@post.cz	ohk@kolin.cz
vranik@firma.seznam.cz	pon_priya@hotmail.com
filipdesign@typo.cz	jan.vlcek@sasprg.cz
calabek@matnw.upol.cz	i.franke@lingcom.de
mandak@fsik.cvut.cz	press@linux.cz
sanglifeng@yahoo.com	tomas@honzak.cz
bicak@feld.cvut.cz	seide@mac.com
lam.yoke.chiew@unico.com.my	xchoura@hpk.felk.cvut.cz
oto_b@jezek.buchtovi.cz	koukal@uesp.fee.vutbr.cz
qmailr@tygr.kolej.mff.cuni.cz	kkh@icu.ac.kr
dtoman@pef.skoda.cz	rosta@acarus.entu.cas.cz
refikdogan@hotmail.com	clmiao@sycamore.ia.ac.cn
skim@egarden.cz	tomik@dracula.sh.cvut.cz
prazak@dcse.fee.vutbr.cz	l.tenbosch@let.kun.nl
igor.kopriva@seznam.cz	j.madle@sh.cvut.cz
dond@volny.cz	sola.aina@filmakademie.de
lnobel@ibasis.net	bazant@ktf.cuni.cz
xplechac@fi.muni.cz	news@news.dcs.muni.cz
sanger@math.muni.cz	glambr@in.gr
liboska@tereza.uochb.cas.cz	pb@ckk.cz
wtgu@gavo.t.u-tokyo.ac.jp	xvrozina@cslab.felk.cvut.cz
jakl@fyzika.fme.vutbr.cz	xpolak@informatics.muni.cz
miroslav.prkna@wwl.cz	rezacd@fit.vutbr.cz
usenet@news.vutbr.cz	pdm@artemis.dcs.muni.cz
redl@vol.cz	bajer@optnw.upol.cz
chenyuanpeng@sina.com	magor@email.cz
psvs_sai@yahoo.com	matus@cmtfnw.upol.cz
hulan@cs.bu.edu	kas@erebor.ics.muni.cz
paluch@bimbo.fjfi.cvut.cz	a.krsek@sh.cvut.cz
balda@cdm.it.cas.cz	news@news.gts.cz
pmb@pha.pvtnet.cz	xmatl@fi.muni.cz
mutex@vm.ics.muni.cz	chmelik@fyzika.fme.vutbr.cz
fiala@kolej.mff.cuni.cz	iam_bobson@hotmail.com
redhat-cz@linux.cz	luzhache@usa.net
dutka@lazne.cz	brejcha@udv.zcu.cz
mike_bugger@yahoo.com	kween@wo.cz
my_pal@hanmail.net	tkadlec@fel.cvut.cz
rehak@math.muni.cz	silvia.allegro@phonak.ch
kubecek@crs-net.cz	arnaud@prevel.com
miloslav.kosek@vslib.cz	meo_me@hotmail.com

usenet@ns.felk.cvut.cz	j@kubs.cz
adelton@fi.muni.cz	list@km.cz
alokt@iitk.ac.in	adelton@aisa.fi.muni.cz
mathias@voxi.se	kilinc@tis.havelsan.com.tr
usenet@news.cuni.cz	dvorakadam@seznam.cz
docbook-info@linux.cz	jakl@pecka.isibrno.cz
a.foerster@actebis-sro.cz	rudolf.ulc@oku-st.cz
keyvan@stanford.edu	tuba@yunus.mam.gov.tr
flidr@kky.zcu.cz	eeqhhe@scut.edu.cn
xthovt@seznam.cz	tvrdik@sun.felk.cvut.cz
root@ns.felk.cvut.cz	petr.sojka@cstug.cz
p.kral@sh.cvut.cz	adelton@informatics.muni.cz
rada@menza.mff.cuni.cz	ea@cs.felk.cvut.cz
georg.sollers@atlas.cz	dantweedie@yahoo.com
muller@matnw.upol.cz	kiru@u.washington.edu
miroslav.benes@vsb.cz	mojza@seznam.cz
talich@cesnet.cz	gavin.smith@brunel.ac.uk
xpoupa@feld.cvut.cz	pavel@ics.muni.cz
rezek@asu.cas.cz	marecek@elnet.cz
news@ns.felk.cvut.cz	rajput_manoj@yahoo.co.in
thaaa@newton.hanyang.ac.kr	pavel.rychetsky@mpsv.cz
i@email.cz	vasek@matfyz.cz
salvet@muni.cz	pkucera@bart.math.muni.cz
arthur@ust.hk	s.rataj@atlas.cz
jkvdpoel@yahoo.com.br	rany@eng.tau.ac.il
milan.hokr@vslib.cz	egon@www.capitol.cz
marcos.abalo@unifr.ch	jchoate@newyork.com
randa_el_wakeel@hotmail.com	stepan@students.zcu.cz
mugur@uekae.uekae.tubitak.gov.tr	kocar@fzu.cz
chainatsung@yahoo.com	jakl@isibrno.cz
ceplm@seznam.cz	piska@fzu.cz
undisclosed_recipients@anxur.fi.muni.cz	murray.holt@ddl.co.uk
brendan.lynskey@pace.co.uk	dont@math.feld.cvut.cz
saintien@newton.hanyang.ac.kr	jonathanv@post.cz
krela@seznam.cz	viktor.pavlicka@osu.cz
vlk@ltm.pvt.cz	jirri@seznam.cz
colombo@mail.kolej.mff.cuni.cz	rk_tex@centrum.cz
paul.shiller@eng.ped.gmeds.com	vsitar@caramail.com
cernym@vszbr.cz	fikais@mx.bajt.cz
htk-users-request@eng.cam.ac.uk	chiaping@u.washington.edu
usenet@news.eunet.cz	sejda@kn.vutbr.cz
kilian@dbme.fee.vutbr.cz	vladboro@mff.cuni.cz
dobias@labe.felk.cvut.cz	frantisek.nemec@ruk.cuni.cz
olsak@math.feld.cvut.cz	osovsky@law.muni.cz
finews@daeron.fi.muni.cz	mailserv@cs.felk.cvut.cz
ramil.sagum@up.edu.ph	krivka@apollo.karlov.mff.cuni.cz
coufal@uivt.cas.cz	gopi@ele.uri.edu
pgs@dcs.muni.cz	sedlacek@ipp.cas.cz

nancy.chan@student.cityu.edu.hk	weed@sci.muni.cz
lars@einsteinlab.de	jgrim@physics.muni.cz
pavel.pilny@seznam.cz	ktoman@email.cz
pavel.herzog@vslib.cz	wuhai@compapp.dcu.ie
moreau@ftmail.ee.tu-berlin.de	xholecek@informatics.muni.cz
olsakova@karlin.mff.cuni.cz	horsky@semi.mff.cuni.cz
vipdfjp@ms.anet.cz	zdendik@students.zcu.cz
michaelk@volny.cz	honza@ov.sasprg.cz
kocicka@cpress.cz	ceplovam@seznam.cz
holly.kelleher@motorola.com	tom.kucera@sh.cvut.cz
k.kralert@sh.cvut.cz	craddison@yahoo.com
sercetary@custug.cz	xpaier@feld.cvut.cz
pavel.janik@math.muni.cz	osovsk@math.muni.cz
pfranc@retia.cz	drijo@comversens.com
pavel@inet.cz	kpostran@students.zcu.cz
zlamal@ufi.fme.vutbr.cz	lrnliu@sina.com
kapsa@apollo.karlov.mff.cuni.cz	fraga@inatel.br
robot@mrp.cz	livingston@pha.pvtnet.cz
ask@mpa-garching.mpg.de	ericyang@dsp.ee.cgu.edu.tw
marek.olsavsky@sostgvb.cz	syjeong@neuron.kaist.ac.kr
simonik@fme.vutbr.cz	demel@fsv.cvut.cz
stary@klok.cvut.cz	schang@ihanyang.ac.kr
albouy@cict.fr	xznamen@aisa.fi.muni.cz
sabbagh_maha@hotmail.com	kloc@drs.ipm.cz
liselene@lps.usp.br	csinfo-l@earn.cvut.cz
belgiorno@cirte.it	maiser@mat-infnw.upol.cz
t_holenda@advis.cz	plesingr@fm.vse.cz
jazyk@artax.karlin.mff.cuni.cz	jan.buckow@voicecom.de
infojob@centrum.cz	wagner@mbox.cesnet.cz
petr_novak@chmi.cz	milos.balcarik@berit.cz
htk-users@eng.cam.ac.uk	m.bruchanov@sh.cvut.cz
editors@linux.cz	lstefek@logis.cz
sorm@pef.mendelu.cz	murat@research.att.com
xkumpera@feld.cvut.cz	wagner@icpf.cas.cz
listserv@cs.felk.cvut.cz	assaf@svu.co.il
berka@sgi.felk.cvut.cz	jan.hon@hon.cz
MichalM@opencae.cz	copebr@navair.navy.mil
linux@mat-infnw.upol.cz	seminar@utia.cas.cz
pavel.kudrna@mff.cuni.cz	rene.levinsky@cerge.cuni.cz
news@morticia.kit.vslib.cz	jira@mail.kolej.mff.cuni.cz
gregus@ant.fee.vutbr.cz	sgi@cs.felk.cvut.cz
member_service@aol.com	houstek@karlin.mff.cuni.cz
peter.vanroose@esat.kuleuven.ac.be	jmanek@mbox.vol.cz
kamin@atlas.cz	yhifny@hotmail.com
hiddenspeech@yahoo.com	pkucera@hunter.math.muni.cz
chanmo@madang.ajou.ac.kr	paraon@orgio.net
jsed@email.com	zzyka@lab.felk.cvut.cz
kucera@fzu.cz	jokr@phil.muni.cz

oskar@jaist.ac.jp	katz@ipe.et.uni-magdeburg.de
smajlik@cs.felk.cvut.cz	ttkristj@math.uwaterloo.ca
jean-julien.aucouturier@kcl.ac.uk	jjezek@hermes.zcu.cz
setantomang@yahoo.com	manisha_ha@hotmail.com
hadacz@math.muni.cz	petr@ferschmann.cz
kas@informatics.muni.cz	teemu.hirsimaki@hut.fi
petr@iconics.cz	n_psaroudakis@hotmail.com
luzmalvy@telcel.net.ve	jan.tichy@webbie.cz
xdobiasl@seznam.cz	tchen@msrchina.research.microsoft.com
jana.chlebikova@cstug.cz	hlavac@vision.felk.cvut.cz
peter.duke@team.telstra.com	matlas@centrum.cz
dcadenas@polar.es	xbacovsk@informatics.muni.cz
javi@velayos.com	kaja@fykos.mff.cuni.cz
olivier.thyes@swisscom.com	hnojil@tetraodon.fsv.cvut.cz
vs@visla.utia.cas.cz	calabek@aix.upol.cz
xmerta@fi.muni.cz	vrehak@seznam.cz
davidt@ucw.cz	folio@omicron.felk.cvut.cz
xmudra@sun.felk.cvut.cz	zhengke@mail.hz.zj.cn
fasanti@usc.es	jih@itc.cz
lisicky@datis.cdmail.cz	libor@informatics.muni.cz
pfranc@pdns.retia.cz	g.jeddak@volny.cz
zbynek.hubinka@vslib.cz	martinv@mtranslations.cz
phix@alumni.utexas.net	maria.aguilar@mch.siemens.de
skoda@@.felk.cvut.cz	xcaha@fi.muni.cz
stale@volny.cz	robik@veverka.sh.cvut.cz
kozusky@ics.muni.cz	stribrsj@vscht.cz
kenneth@iis.sinica.edu.tw	netol@volny.cz
dalibor.lukas@email.cz	michalr@baal.econ.muni.cz
tm@salome.datron.cz	jinsp@pusan.ac.kr
salvet@ics.muni.cz	bohous@lapeco.cz
cech@student.upce.cz	root@gate.etspraha.cz
japp@golem.gate.fee.vutbr.cz	xmareks@cslab.felk.cvut.cz
windonabc@sina.com	postry@gandalf.fsih.cvut.cz
ggravier@irisa.fr	libor@fi.muni.cz
lukovic@minv.sk	bulant@math.muni.cz
weaam@hotmail.com	bookorders@cstug.cz
ffridrich@telecom.cz	lenka@fykos.mff.cuni.cz
fiala@mail.kolej.mff.cuni.cz	norbert.pomp.fs@vsb.cz
pam@fi.muni.cz	t.cermak@aquasoft.cz
mxserver@cs.felk.cvut.cz	pavel@muni.cz
xobitko@fel.cvut.cz	rybicka@mendelu.cz
johnn@cstr.ed.ac.uk	eeevansn@swansea.ac.uk
xbezpale@sun.felk.cvut.cz	korcak@vol.cz
jan.kos@vslib.cz	hroch@monoceros.physics.muni.cz
gabzili_hanen@yahoo.com	yaser_aseem@yahoo.com
ivo.jancik@winet.cz	zlamal@fyzika.fme.vutbr.cz
bruchm@racom.cz	indad@fce.vutbr.cz
zbynh@seznam.cz	kolman@spc.civ.cvut.cz

sqrt@centrum.cz	lcermak@students.zcu.cz
sluka@ant.fee.vutbr.cz	prema@email.cz
xdufka@fi.muni.cz	tunnelprod@hotmail.com
balda@ufy.zcu.cz	jared_hanson@hotmail.com
omdes@glue.umd.edu	sangbae@icu.ac.kr
b.zwiers@kpn.com	macinnis@fas.harvard.edu
zlatuska@anxur.dcs.muni.cz	lbj@gavo.t.u-tokyo.ac.jp
wolf@umbr.cas.cz	qzhou@research.bell-labs.com
dubouc@ifrance.com	tomas.honzak@st.ms.mff.cuni.cz
beranj@ed.dp-praha.cz	vrana@comp.cz
kropacek@hrk.pvt.cz	amnyousaf@hotmail.com
dan@cas.cz	hgarudad@qualcomm.com
uchac@x-base.cz	phix@odu.edu
nospam-vmichl@cadstudio.cz	mike@dale.uek.cas.cz
beranek@ikaros.zcu.cz	petr.vojkovsky@post.cz
bulik@math.muni.cz	ivop@regionet.cz
weel@www.com	xkvasniv@cslab.felk.cvut.cz
pfy@atlas.cz	xbacovsk@fi.muni.cz
martin.slachta@vsb.cz	havlyk@atlas.cz
uri@paradigmgeo.com	cgattein@cc.gatech.edu
mrz@kpssoft.cz	krajicek@ics.muni.cz
zdenek.wagner@cstug.cz	kloc@ipm.cz
hufhendr@atlas.cz	ivo.panacek@regionet.cz
rdelgado@polar.es	vodicka@cslab.felk.cvut.cz
porizka@econ.muni.cz	david@kuzela.cz
imladris@fi.muni.cz	dast@regionet.cz
hluze@email.cz	rkrejci@jumbo.ped.muni.cz
pepa@vabo.cz	cslatex@seznam.cz
assaf@eng.tau.ac.il	dujeu@tsi.enst.fr
hafez@su.sabanciuniv.edu	bernt.andrassy@mchp.siemens.de
vojta@tunw.upol.cz	
roman.kaspar@seznam.cz	

Table 28 – List of 550 e-mails

### 5.3.16 Special keyboard characters (CK1-2)

Names of **two** special keyboard characters from were recorded for each speaker. The names of the special keyboard characters were prompted to the speakers as orthographic words. Bold text is used for mandatory keyboard characters.

<b>zavináč</b>	<b>‘at’ sign (‘@’)</b>
<b>křížek</b>	<b>hash (‘#’)</b>
<b>zpětné_lomítko</b>	<b>backslash (‘\’)</b>
<b>mezera</b>	<b>blank   space (‘ ’)</b>
<b>dvojtečka</b>	<b>colon (‘:’)</b>
<b>tečka</b>	<b>dot   period (‘.’)</b>
<b>pomlčka   mínus</b>	<b>hyphen (‘-’)</b>
<b>lomítko</b>	<b>slash (‘/’)</b>

hvězdička	star   asterisk ('*')
plus	plus   plus sign ('+')
podtržítka	underscore ('_')
trubka   svislá čára	pipe (' ')
vlnka   tilda	tilde ('~')
vykřičník	exclamation mark   exclamation point ('!')
otazník	question mark ('?')
uvozovky	double quote
apostrof	quote   single quote (' ' '')
procento   procenta	percent   percent sign ('%')
středník	semicolon (';')
čárka	comma (',')

Table 29 – List of keyboards characters

### 5.3.17 Application specific commands

ID	Semantic description	English word	Czech word
<b>Activating and deactivating the system (C1)</b>			
C1.1	activate a device	<i>on</i>	zapnout
C1.2	deactivate a device	<i>off</i>	vypnout
C1.3	put device into standby mode	<i>standby</i>	deaktivovat
C1.4	awaken device from standby mode	<i>wake_up</i>	aktivovat
C1.5	cancel current operation	<i>cancel</i>	zrušit
C1.6	indicate that you want to change password	<i>password</i>	heslo
C1.7	gain access to a protected device/mode	<i>register   login</i>	přihlásit_se   zalogovat_se
C1.8	activate keylock	<i>keylock</i>	zamknout_klávesnici
C1.9	leave a protected mode	<i>logout</i>	odhlásit_se   odlogovat_se
C1.10	activate speech recognition mode	<i>voice_control</i>	hlasový_vstup
C1.11	lock current settings   device etc.	<i>lock</i>	zamknout
C1.12	unlock current settings   device   etc.	<i>unlock</i>	odemknout
C1.13	finalize an input (as <CR>)	<i>enter</i>	potvrdit
C1.14	quit a running application	<i>exit   quit</i>	ukončit
C1.15	to adjust settings	<i>set</i>	nastavit
<b>Devices</b>			



C1.16	device name: radio	<i>radio</i>	rádio
C1.17	device name: camcorder	<i>camcorder</i>	kamera   videokamera
C1.18	device name: camera	<i>camera</i>	fotoaparát
C1.19	device name: cassette	<i>cassette</i>	kazeta
C1.20	device name: CD player	<i>CD</i>	CD
C1.21	device name: CD writer	<i>CD-R</i>	CD_vypalovačka
C1.22	device name: DAT player/recorder	<i>DAT</i>	DAT
C1.23	device name: DVD player/writer	<i>DVD</i>	DVD
C1.24	device name: HiFi system	<i>HiFi</i>	HiFi
C1.25	device name: MiniDisk player/recorder	<i>MD</i>	minidisk
C1.26	device name: MP3 player	<i>MP3</i>	MP3_přehrávač
C1.27	device name: microphone	<i>microphone</i>	mikrofon
C1.28	device name: television	<i>TV</i>	televize
C1.29	device name: mobile telephone   cellular phone	<i>mobile   cellular</i>	mobil
C1.30	device name: PDA	<i>PDA</i>	PDA
C1.31	device name: set-top box	<i>set-top_box</i>	set-top_box   STB
C1.32	device name: video	<i>video</i>	video
C1.33	device name: computer	<i>computer</i>	počítač
C1.34	device name: phone	<i>phone</i>	telefon
<b>Operating devices</b>			
C1.35	start service   device   function	<i>start   activate</i>	aktivovat   spustit
C1.36	start again	<i>restart</i>	restartovat
C1.37	end service   function	<i>abort   end   stop</i>	konec   ukončit   zrušit
C1.38	temporarily interrupt the current function	<i>pause</i>	pauza
C1.39	resume the current function	<i>continue</i>	pokračovat
C1.40	open an object (file/item/window)	<i>open</i>	otevřít
C1.41	close an object (file/item /window)	<i>close</i>	zavřít
C1.42	confirm an operation	<i>confirm</i>	potvrdit
C1.43	activate help mode	<i>help</i>	nápověda
C1.44	ask for information	<i>information</i>	informace
C1.45	activate a SW assistant or contact a human assistant	<i>assistant</i>	pomocník
C1.46	verify an operation	<i>verify</i>	ověřit
C1.47	check if any event has happened   e.g. check mail- box	<i>check</i>	zkontrolovat
C1.48	switch on or off or between   e.g. functions/menu items   etc.	<i>switch # switch_on # switch_off</i>	přepnout # zapnout # vypnout

<b>Voice control settings</b>			
C1.49	greeting the system (e.g. saying <i>Hello</i> enables the system to define English as dialogue language)	<i>hello</i>	dobrý_den
C1.50	list command & control vocabulary	<i>vocabulary</i>	slovník
C1.51	start a demo	<i>demo</i>	demo
C1.52	name of the current language (e.g. for German recordings select word <i>Deutsch</i> )	< <i>language</i> >	čeština   česky
C1.53	tell the recognizer to learn a new word	<i>learn</i>   <i>train</i>	trénovat   učit_se
C1.54	enter spelling mode	<i>spell</i>	hláskovat
C1.55	enter digit mode	<i>digit</i>	číslo
C1.56	to record a command or message	<i>record</i>	nahrát
<b>Timer control</b>			
C1.57	address timer function; activate/program timer function	<i>timer</i>	časovač
C1.58	address clock function; activate clock	<i>clock</i>	hodiny
C1.59	address time function; announce/display/set time	<i>time</i>	čas
C1.60	address date function; announce/display/set date	<i>date</i>	datum
C1.61	address alarm function; activate alarm function	<i>alarm</i>	alarm   budík
C1.62	activate device or function daily at a preset time	<i>daily</i>	denně   každý_den
C1.63	activate device or function weekly at a preset time	<i>weekly</i>	týdně   každý_týden
C1.64	activate device or function monthly at a preset time	<i>monthly</i>	měsíčně   každý_měsíc
C1.65	go to an earlier time	<i>earlier</i>	dříve   předchozí
C1.66	go to a later time	<i>later</i>	později   následující
<b>Programming</b>			
C1.67	program advanced features or options   or enter sub-menu to do this	<i>program</i>	programovat
C1.68	select automatic mode	<i>automatic</i>	automaticky

C1.69	select manual mode	<i>manual</i>	ručně
C1.70	reset to the system to default	<i>reset</i>	reset   výchozí_nastavení
C1.71	save current item   configuration   etc.	<i>save   store</i>	uložit
C1.72	initiate recording/activation of a macro	<i>macro</i>	makro
C1.73	enter settings menu; select settings functions	<i>settings</i>	nastavení
C1.74	noun   activate device/system control	<i>control</i>	ovládání
<b>System status</b>			
C1.75	show the status of the system	<i>status</i>	stav
C1.76	show battery status	<i>battery</i>	baterie
C1.77	display signal strength	<i>signal_strength</i>	signál
C1.78	activate checking procedure	<i>self-test</i>	kontrola_systému
<b>Connectivity</b>			
C1.79	connect to network	<i>network</i>	síť
C1.80	infrared interface	<i>infrared</i>	infraport
C1.81	select bluetooth operation	<i>bluetooth</i>	bluetooth
C1.82	address (internet) host	<i>host</i>	hostitel   počítač
C1.83	go to network function	<i>network</i>	síť
C1.84	synchronize data with PC entries	<i>synchronize</i>	synchronizovat
C1.85	try again to activate/redo the last (unsuccessful) action (command   query   ...) again	<i>redo   retry</i>	opakovat
<b>Directory navigations (C2)</b>			
<b>Calling up menus</b>			
C2.1	go to a menu	<i>menu</i>	menu   nabídka
C2.2	enter directory (e.g. of files) or list directory entries	<i>directory</i>	adresář
C2.3	show options from a menu   etc.	<i>options</i>	volby
C2.4	go to a sub-menu   i.e. give more details	<i>detail</i>	podrobnosti
C2.5	to show an inventory of items	<i>list</i>	seznam
<b>Menu browsing</b>			
C2.6	go to the beginning of a	<i>home</i>	hlavní_menu

	dialogue or to top-level menu		hlavní_nabídka
C2.7	noun or adverb   go to the beginning of list	<i>start</i>	začátek
C2.8	noun or adverb   go to the beginning of list	<i>top</i>	nahoře
C2.9	noun or adverb   go to the beginning of list	<i>front</i>	dopředu
C2.10	noun or adverb   go to the beginning of list	<i>first</i>	první
C2.11	go to end of list	<i>bottom</i>	dole
C2.12	go to end of list	<i>end</i>	na_konec
C2.13	go to end of list	<i>last</i>	poslední
C2.14	go up one item	<i>up</i>	nahoru
C2.15	go up one item	<i>plus</i>	
C2.16	go down one item	<i>down</i>	dolů
C2.17	go down one item	<i>minus</i>	
C2.18	go left from present position	<i>left</i>	doleva   nalevo
C2.19	go right from present position	<i>right</i>	doprava   napravo
C2.20	go forward one item in the list	<i>next</i>	další
C2.21	go forward one item in the list	<i>forward</i>	vpřed
C2.22	go back one item in the list	<i>previous</i>	předchozí
C2.23	go back one item in the list	<i>back</i>	zpět
C2.24	color: blue	<i>blue</i>	modrý   modrá
C2.25	color: red	<i>red</i>	červený   červená
C2.26	color: yellow	<i>yellow</i>	žlutý   žlutá
C2.27	color: green	<i>green</i>	zelený   zelená
C2.28	color: black	<i>black</i>	černý   černá
C2.29	color: white	<i>white</i>	bílý   bílá
C2.30	search for an item	<i>search</i>   <i>lookup</i>   <i>find</i>	hledat   najít
C2.31	ask the system to suggest something (e.g. hotel or a route)	<i>suggest</i>	navrhnout
C2.32	browse through a list	<i>browse</i>	procházet
C2.33	conduct a fuzzy search based on written or spoken input (e.g. Krzysztof ⇔ sounds like Kschischtof   reads like Christof)	<i>like</i>	něco_jako
C2.34	go to item   identified by name   number etc.	<i>go_to</i>	jít_na   skočit_na
C2.35	go to row	<i>row</i>	řádek
C2.36	go to column	<i>column</i>	sloupec
C2.37	go to page (+ number)	<i>page</i>	stránka   stránku

### Selecting items

C2.38	select menu/item	<i>select   choose</i>	vybrat   označit
C2.39	select all items	<i>all</i>	vybrat_vše   vše
C2.40	skip a marked item	<i>skip   ignore</i>	přeskočit   ignorovat
<b>Editing (C3)</b>			
C3.1	create a new item	<i>new</i>	nový   nová
C3.2	object of operation: entry; e.g. entry in a list or directory	<i>entry</i>	položka
C3.3	input an item	<i>enter</i>	zadat
C3.4	dictate text	<i>dictate</i>	diktovat
C3.5	change an entry	<i>modify   change</i>	změnit
C3.6	correct an entry	<i>correct</i>	opravit
C3.7	copy (to clipboard)	<i>copy</i>	kopírovat
C3.8	cut from chosen source to chosen destination	<i>cut</i>	vyjmout
C3.9	paste an entry	<i>paste</i>	vložit
C3.10	insert an entry or item	<i>insert</i>	vložit
C3.11	add an entry or an item	<i>add</i>	přidat
C3.12	create an entry or an item	<i>create</i>	vytvořit
C3.13	delete or remove an entry/item	<i>delete   remove   erase</i>	vymazat   odstranit   smazat
C3.14	highlight points of interest	<i>mark</i>	označit
C3.15	sort elements by name   date   ...	<i>sort</i>	seřadit
C3.16	undo last operation	<i>undo</i>	zpět
C3.17	repeat last function	<i>repeat</i>	opakovat
C3.18	replace an item	<i>replace</i>	nahradit
C3.19	accept current status/settings or confirm correctness	<i>okay   OK</i>	OK
C3.20	edit a text   business card   entry   etc.	<i>edit</i>	upravit   editovat
C3.21	edit/delete/go to number xxx	<i>number</i>	číslo
C3.22	politely ask for something	<i>please</i>	prosím
<b>Output control (C4)</b>			
<b>General</b>			
C4.1	personal picture and/or sound	<i>profile</i>	profil   nastavení
C4.2	set settings to normal/neutral	<i>neutral   normal   default</i>	standardní   implicitní   normální
C4.3	reach the top value of a certain range	<i>maximum</i>	maximum
C4.4	reach the smallest value of a certain range	<i>minimum</i>	minimum

C4.5	set level (+ number   range etc.)	<i>level</i>	úroveň   stupeň
C4.6	set level high	<i>high</i>	horní   vysoká
C4.7	set level low	<i>low</i>	spodní   nízká
C4.8	set level medium	<i>medium</i>	střední
C4.9	increase or get more	<i>more</i>	přidat
C4.10	decrease or get less	<i>less</i>	ubrat
C4.11	select channel/program (+ number or name)	<i>channel   program</i>	kanál   program
<b>VISUAL OUTPUT</b>			
<b>Display format</b>			
C4.12	display results   file etc.	<i>show   display</i>	zobrazit
C4.13	enlarge window to full-scale	<i>maximize</i>	maximalizovat
C4.14	hide window on a display	<i>hide   minimize</i>	minimalizovat   skryt
C4.15	clear display	<i>clear</i>	vymazat
C4.16	zoom in	<i>zoom_in</i>	zvětšit
C4.17	zoom out	<i>zoom_out</i>	zmenšit
C4.18	determine format or scale of displayed object	<i>format   picture_size</i>	formát   rozměry
C4.19	noun   e.g. scale of the map	<i>scale</i>	stupnice   měřítko
C4.20	e.g. of an image in pixels	<i>resolution</i>	rozišení
C4.21	refresh current display	<i>update   refresh</i>	obnovit
C4.22	hold picture	<i>freeze</i>	zmrazit   zastavit
<b>Brightness</b>			
C4.23	initiate control of 'light' function	<i>light</i>	světlo   osvětlení
C4.24	enter brightness menu	<i>brightness</i>	jas
C4.25	set picture more bright	<i>brighter</i>	zesvětlit
C4.26	set picture less bright	<i>darker</i>	ztmavit
<b>Color balance</b>			
C4.27	enter and manipulate color balance menu	<i>color</i>	barva
<b>Contrast</b>			
C4.28	enter contrast menu	<i>contrast</i>	kontrast
C4.29	set contrast to be sharper	<i>sharper</i>	zvýšit kontrast
C4.30	set contrast to be softer	<i>softer</i>	snížit kontrast
<b>AUDIO OUTPUT</b>			

<b>Volume</b>			
C4.31	enter volume menu	<i>volume</i>	hlasitost
C4.32	increase volume	<i>louder</i>	zesílit
C4.33	decrease volume	<i>quieter</i>	ztlumit   zeslabit
C4.34	turn sound off	<i>mute   silent</i>	vypnout_zvuk
C4.35	turn sound on again	<i>unmute</i>	zapnout_zvuk
C4.36	let volume fade	<i>fade</i>	utlumit
<b>Sounds and signal control</b>			
C4.37	choose sounds	<i>sounds</i>	zvuky
C4.38	select melody/ring-tone function	<i>melody   ring_tone</i>	melodie   vyzvánění
C4.39	vibrator signal	<i>vibrating_alert</i>	vibrace
<b>Tone control</b>			
C4.40	enter tone menu	<i>tone</i>	barva_zvuku
C4.41	set bass level (+ number   range   etc.)	<i>bass</i>	hloubky   basy
C4.42	set treble level (+ number   range   etc.)	<i>treble</i>	výšky
C4.43	bass boost	<i>bass-boost</i>	zvýraznit_hloubky   zvýraznit_basy
C4.44	access equalizer	<i>equalizer</i>	ekvalizér
<b>Output mode</b>			
C4.45	output to loudspeakers (+on/off)	<i>loudspeaker</i>	reproduktory
C4.46	output to headphones (+on/off)	<i>headphones</i>	sluchátka
C4.47	data source	<i>line-in</i>	linkový_vstup
C4.48	data target	<i>line-out</i>	linkový_výstup
C4.49	switch to hands-free (muting head-set)	<i>hands-free</i>	hands-free
<b>Play</b>			
C4.50	playback	<i>play   playback</i>	přehrát
C4.51	play at slower speed	<i>slow</i>	pomaleji
C4.52	play at faster speed	<i>fast</i>	rychleji
C4.53	play in shuffle mode	<i>shuffle   random_play</i>	náhodný_výběr
C4.54	change side of tape	<i>reverse</i>	druhá_strana
C4.55	eject medium	<i>eject</i>	vyjmout   vysunout
<b>Print</b>			

C4.56	create hardcopy of ticket   map   time table   etc...	<i>print</i>	tisk
<b>Read</b>			
C4.57	Read (aloud) a file/item	<i>read</i>	číst
<b>Browsing (C5)</b>			
<b>General message functions</b>			
C5.1	enter message mode	<i>message</i>	zpráva
C5.2	make a connection	<i>connect</i>	připojit_se
C5.3	answer the call/message	<i>answer   reply</i>	odpovědět   odpověď
C5.4	answer to everyone in a list	<i>answer_all   reply_all</i>	odpovědět_všem
C5.5	forward message or call (+ name   number etc.)	<i>forward   transfer   divert</i>	přeposlat   forwardovat
C5.6	send message	<i>send</i>	poslat   odeslat
C5.7	receive a message or broadcast or phone call	<i>receive</i>	přijmout
C5.8	declare message as urgent	<i>urgent</i>	urgentní
C5.9	call someone back	<i>callback</i>	zavolat_zpět
<b>Voice Memo</b>			
C5.10	enter voice memo program	<i>memo   voice_memo</i>	hlasové_připomenutí   záznamník
C5.11	invoke vacation mode; set vacation notification	<i>vacation</i>	dovolená
<b>Email</b>			
C5.12	send/receive email	<i>email</i>	e-mail
C5.13	enter mailbox	<i>mailbox</i>	schránka   mailbox
C5.14	get an item   e.g. get an e-mail from a server	<i>get   receive</i>	přijmout   stáhnout
C5.15	go to Inbox	<i>inbox</i>	inbox   příchozí_pošta
C5.16	ask for information on mail header	<i>header</i>	hlavička
<b>Fax</b>			
C5.17	enter fax menu program	<i>fax</i>	fax
<b>SMS</b>			
C5.18	enter SMS menu program	<i>SMS</i>	SMS



<b>DTMF</b>			
C5.19	activate transfer by DTMF	<i>DTMF</i>	tónová_volba
<b>Infrared/Bluetooth messaging</b>			
C5.20	transfer e.g. the electronic business card via infrared interface	<i>beam   transfer</i>	poslat   přenést
C5.21	set / edit the electronic business card	<i>business_card</i>	vizitka
<b>Internet browsing</b>			
C5.22	enter/leave internet program	<i>internet</i>	internet
C5.23	activate/deactivate internet browser	<i>browser</i>	prohlížeč
C5.24	set bookmark	<i>bookmark</i>	záložky
C5.25	follow hyperlink	<i>follow</i>	následovat
C5.26	address hyperlink	<i>hyperlink   link</i>	odkaz   link
<b>TELEPHONE</b>			
<b>Dialing options</b>			
C5.27	a telephone call (noun)	<i>call</i>	hovor   telefon
C5.28	place a call (verb)	<i>dial</i>	volit   vytočit
C5.29	place a call (verb)	<i>call</i>	zavolat
C5.30	call last number again	<i>redial</i>	opakovat_volbu
C5.31	place a call by saying a name	<i>name_dialing</i>	volit_jméno
C5.32	place a call digit by digit	<i>digit_dialing</i>	volit_číslo
C5.33	initiate call by its short number	<i>speed_dialing</i>	rychlé_vytáčení
C5.34	choose an item from a finite list	<i>choose</i>	vybrat
C5.35	set a call-back reminder	<i>reminder</i>	připomenout
C5.36	make a conference phone call	<i>conference</i>	konference
C5.37	use prepaid operation	<i>prepaid</i>	předplacené_služby
C5.38	list recent calls   incoming and outgoing	<i>call-list</i>	seznam_hovorů
<b>Shortcuts</b>			
C5.39	emergency call(s)	<i>emergency</i>	tísňové_volání
C5.40	call customer care	<i>customer_care  </i>	zákaznické_centrum

		<i>support</i>	
C5.41	transfer call to human operator	<i>operator</i>	operátor
<b>Extensions &amp; User profiles</b>			
C5.42	activate extension mode; e.g. “train new extension: ‘boat’” which can then be used for setting up a call by saying “call Anton boat”	<i>extension</i>	rozšíření
C5.43	connect to extension; i.e. make a call to the ‘car phone’ of the selected name	<i>car</i>	auto
C5.44	connect to extension; i.e. call the person “at home”	<i>home   private</i>	domů
C5.45	connect to extension; i.e. call a person’s mobile number	<i>mobile</i>	mobil
C5.46	connect to extension; i.e. call a person’s office number	<i>office</i>	kancelář
C5.47	make a call to the ‘pager’ for the selected name	<i>pager</i>	pager
C5.48	connect to extension; call the person’s assistant	<i>assistant</i>	asistent   asistentka
C5.49	make a call to an unspecified other number of the selected name	<i>other</i>	další
C5.50	connect to extension; call the person’s secretary	<i>secretary</i>	sekretářka
C5.51	connect to extension; call a person “at work”	<i>work</i>	práce   zaměstnání
C5.52	use an international phone number extension	<i>international</i>	mezinárodní
C5.53	activate/change user profile setting (e.g. from normal to meeting)	<i>profile</i>	profil
C5.54	activate user profile	<i>home</i>	domácí
C5.55	activate user profile	<i>outdoors</i>	venkovní
C5.56	activate user profile	<i>city</i>	městský
C5.57	activate user profile	<i>hands-free</i>	hands-free
C5.58	activate user profile	<i>meeting</i>	schůzka
<b>Call handling</b>			
C5.59	recording a greeting for answering to the incoming call	<i>greeting</i>	pozdrav
C5.60	hang-up the phone	<i>hang-up</i>	zavěsit
C5.61	put the call on hold	<i>hold</i>	podržet

C5.62	accept call	<i>accept_call</i>	přijmout
C5.63	refuse call	<i>reject_call</i>   <i>refuse_call</i>   <i>busy</i>	odmítnout
C5.64	list missed calls	<i>missed_calls</i>	nepřijatá_volání   zmeškaná_volání
C5.65	list received calls	<i>received_calls</i>	přijatá_volání
C5.66	getting the time length of the call	<i>duration</i>	délka
C5.67	go to voice mail	<i>voice_mail</i>   <i>answering_machin</i> <i>e</i>	hlasová_schránka   záznamník
<b>Making reservations</b>			
C5.68	ask for arrival time	<i>arrival</i>	příjezd   přilet
C5.69	ask for departure time	<i>departure</i>	odjezd   odlet
C5.70	make a reservation	<i>book</i>	rezervovat
<b>Organiser functions (C6)</b>			
<b>Organiser</b>			
C6.1	invoke calendar application	<i>calendar</i>	kalendář
C6.2	display today's tasks   appointments etc.	<i>today</i>	dnes
C6.3	display tomorrow's tasks   appointments etc.	<i>tomorrow</i>	zítra
C6.4	display calendar of week + number	<i>week</i>	týden
C6.5	display calendar of month + month or number	<i>month</i>	měsíc
C6.6	display calendar of week + year	<i>year</i>	rok
C6.7	call agenda function	<i>agenda</i>	agenda   program
C6.8	set appointment date	<i>appointment</i>   <i>meeting</i>	schůzka   jednání
C6.9	enter tasks menu	<i>tasks</i>	úkoly
C6.10	display all birthdays	<i>birthday</i>	narozeniny
C6.11	display date(s) of delivery	<i>schedule</i>	harmonogram   program
C6.12	go to contacts	<i>contacts</i>	kontakty
C6.13	remind of task   appointment etc.	<i>reminder</i>	připomenout
<b>Accessories</b>			
C6.14	activate accessories menu	<i>accessories</i>   <i>utilities</i>	příslušenství   nástroje
C6.15	activate calculator	<i>calculator</i>	kalkulačka

C6.16	activate word processor	<i>text-editor</i>	textový_editor
C6.17	activate notes editor	<i>notes</i>	poznámky
C6.18	mode name: games	<i>games</i>	hry
<b>Address book</b>			
C6.19	go to phonebook operations	<i>phone-book</i>	telefonní_seznam
C6.20	go to address book operations	<i>address-book</i>	adresář
C6.21	activate professional address or number list	<i>business</i>	pracovní
C6.22	activate private address or number list	<i>private</i>	osobní   soukromý
C6.23	add name to address book	<i>name</i>	jméno
C6.24	add address (company   street   P.O. box   city   zip code   state   country) to address book	<i>address</i>	adresa
C6.25	add address categories	<i>company</i>	firma
C6.26	address category	<i>street</i>	ulice
C6.27	address category	<i>city</i>	město
C6.28	address category	<i>zip-code</i>	PSC   směrovací_číslo
C6.29	address category	<i>province/state</i>	kraj
C6.30	address category	<i>country</i>	stát
C6.31	add phone number(s) to address book	<i>phone_number</i>	telefon
C6.32	add fax number(s) to address book	<i>fax_number</i>	fax
C6.33	add email address(es) to address book	<i>email_address</i>	e-mail
<b>Routing (C7)</b>			
<b>Positioning</b>			
C7.1	enter/state the actual position	<i>locate   position</i>	pozice   lokalizace
C7.2	show actual position's map	<i>map</i>	mapa
C7.3	enter GPS (global positioning system) menu	<i>GPS</i>	GPS
<b>Journey planning</b>			
C7.4	acoustic guidance on/off	<i>acoustic_guidance</i>	akustická_navigace
C7.5	activate navigation program	<i>guide   navigate</i>	navigace
C7.6	directly guide to pre-stored destination	<i>guide_to</i>	navigace_do
C7.7	suggest route	<i>route</i>	cesta

C7.8	suggest fastest route	<i>fastest_route</i>	nejrychlejší_cesta
C7.9	suggest shortest route	<i>shortest_route</i>	nejkratší_cesta
C7.10	suggest alternative route	<i>alternative_route</i>	jiná_cesta
C7.11	calculate a new route	<i>new_route</i>	nová_cesta
C7.12	show distance to destination	<i>distance</i>	vzdálenost
C7.13	suggest/book/guide to a hotel/accommodation	<i>hotel</i>	hotel
C7.14	suggest/guide to a car lot	<i>parking</i>	parkoviště
C7.15	suggest and/or provide info about rest areas	<i>rest_area</i>	odpočívadlo
C7.16	propose a gas/service station	<i>service_station</i>	servis
C7.17	suggest route with points of interest	<i>sightseeing</i>	prohlídka
C7.18	calculate a route avoiding traffic jams	<i>traffic_jam</i>	zácpa
C7.19	suggest points of interest	<i>points_of_interest</i>	zajímavosti
<b>Enter landmarks</b>			
C7.20	enter start point	<i>start_point</i>	začátek_cesty
C7.21	enter destination	<i>destination</i>	konec_cesty   cíl
C7.22	enter airport	<i>airport</i>	letiště
C7.23	enter border-point	<i>border</i>	hranice
C7.24	enter car rental station	<i>car_rental</i>	autopůjčovna
C7.25	enter city centre	<i>centre</i>	centrum
C7.26	enter name of a crossing street	<i>crossing   junction_point</i>	křižovatka   exit
C7.27	enter place of exhibition	<i>fair   trade_show   exhibition</i>	výstava   veletrh   výstaviště
C7.28	enter ferry crossing point	<i>ferry</i>	trajekt
C7.29	enter car service (garage)	<i>garage</i>	servis
C7.30	enter gas / Petrol station	<i>gas_station</i>	čerpací_stanice   pumpa
C7.31	enter hospital	<i>hospital</i>	nemocnice
C7.32	enter railway station	<i>railway_station</i>	nádraží
C7.33	enter restaurant	<i>restaurant</i>	restaurace
C7.34	Get general traffic information	<i>traffic_information</i>	dopravní_informace
<b>Maps</b>			
C7.35	toggle to pictogram mode	<i>pictogram</i>	piktogram
C7.36	move east from current position	<i>east</i>	východ
C7.37	move west from current position	<i>west</i>	západ
C7.38	move north from current position	<i>north</i>	sever
C7.39	move south from current	<i>south</i>	jih

	position		
<b>Automotive (C8)</b>			
<b>General</b>			
C8.1	recall driver settings	<i>recall_settings</i>	obnovit_nastavení
<b>Cabin Control</b>			
C8.2	enter or leave air-conditioning menu	<i>air-conditioning</i>	klimatizace
C8.3	activate or deactivate heater	<i>heater</i>	topení
C8.4	air re-circulation	<i>re-circulation</i>	cirkulace
C8.5	choose air flow/fan/ventilation/blower	<i>ventilation</i>	větrání
C8.6	enter seat menu	<i>seat</i>	sedadlo
C8.7	set cabin temperature	<i>temperature</i>	teplota
C8.8	open or close window	<i>window</i>	okno
<b>Vehicle control</b>			
C8.9	enter auto PC menu	<i>PC</i>	počítač   PC
C8.10	activate or deactivate defroster	<i>defroster</i>	rozmrazování
C8.11	give status of fuel	<i>fuel   gas</i>	benzín   nafta
C8.12	give status of maintenance	<i>maintenance</i>	údržba
<b>Audio &amp; Video (C9)</b>			
<b>Sound control</b>			
C9.1	select surround level	<i>surround</i>	surround   prostorový_zvuk
C9.2	set balance level	<i>balance</i>	vyvážení
<b>Videotext</b>			
C9.3	activate and deactivate videotext	<i>videotext</i>	videotext
C9.4	switch subtitles on and off	<i>subtitles</i>	titulky
<b>Programming channels</b>			
C9.5	select next memory position	<i>slot   position</i>	předvolba

<b>Connectivity</b>			
C9.6	input from terrestrial antenna	<i>antenna</i>	anténa
C9.7	input from satellite dish	<i>satellite</i>	satelit
C9.8	input from cable network	<i>cable</i>	kabel
C9.9	select auxiliary input number	<i>AUX</i>	AUX
<b>Electronic Program Guide + Radio Data system</b>			
C9.10	enter EPG menu	<i>EPG</i>	EPG
C9.11	enter RDS menu	<i>RDS</i>	RDS
C9.12	show EPG/RDS information on artist	<i>artist</i>	umělec
C9.13	show EPG information on program title	<i>title</i>	název
C9.14	show EPG information on director of the program	<i>director</i>	režisér
C9.15	show EPG information on actor name	<i>actor</i>	herec
<b>Record</b>			
C9.16	store with normal capacity	<i>normal</i>	běžný_záznam
C9.17	store with extended capacity	<i>longplay</i>	dlouhý_záznam
C9.18	media for information storage	<i>mini-disk</i>	minidisk
C9.19	media for information storage	<i>diskette   floppy_disc</i>	disketa
C9.20	media for information storage	<i>memory</i>	paměť
C9.21	media for information storage	<i>disc</i>	disk
<b>Winding and search commands</b>			
C9.22	rewind medium in the device	<i>rewind</i>	přetočit
C9.23	wind medium in the device	<i>fast_forward</i>	rychle_vpřed
C9.24	object: index mark	<i>index</i>	index   značka
C9.25	object: counter	<i>counter</i>	počítadlo
C9.26	activate changer	<i>changer</i>	měníč
<b>Noise reduction options</b>			
C9.27	Select noise reduction system	<i>noise_reduction</i>	potlačení_šumu
<b>Portable devices options</b>			

C9.28	set portable mode	<i>carrying_mode</i>	přenosný_mód
C9.29	set battery charge on	<i>battery_charge</i>	nabíjení_baterie
<b>Indications</b>			
C9.30	show remaining capacity of named medium (tape   MD   CD   etc.)	<i>capacity</i>	kapacita
<b>Video device / camera</b>			
C9.31	record still picture instead of movie	<i>still_picture</i>	zastavit_obraz
C9.32	audio-video track	<i>AV</i>	audio-video
C9.33	insert sound only	<i>sound</i>	zvuk
C9.34	insert video only	<i>video</i>	obraz
<b>Radio #</b>			
C9.35	select DAB (digital audio broadcasting)	<i>D.A.B.</i>	digitální_rádio   DAB
C9.36	ask for stored traffic information messages (TiM)	<i>traffic_messages</i>	dopravní_zprávy   zelená_vlna
<b>DAB categories</b>			
C9.37	select news channel	<i>news</i>	zprávy
C9.38	select actual matters	<i>current_affairs</i>	aktuality
C9.39	select sport channel	<i>sport</i>	sport
C9.40	select education channel	<i>education</i>	vzdělávání
C9.41	select drama program	<i>drama</i>	drama
C9.42	select culture channel	<i>culture</i>	kultura
C9.43	select science channel	<i>science</i>	věda
C9.44	select pop music	<i>pop</i>	pop
C9.45	select rock music	<i>rock</i>	rock
C9.46	select easy music	<i>easy_listening</i>	oddech
C9.47	select classical music	<i>classical</i>	klasika
C9.48	select jazz music	<i>jazz</i>	jazz
C9.49	select country music	<i>country</i>	country
C9.50	select national music	<i>national</i>	domácí_scéna
C9.51	select oldies	<i>oldies</i>	oldies
C9.52	select folk music	<i>folk</i>	folk
C9.53	choose religion program	<i>religion</i>	náboženství
C9.54	select travel channel	<i>travel</i>	cestování
C9.55	select documentary channel	<i>documentary</i>	dokumenty
<b>DVB categories</b>			
C9.56	select movie channel	<i>movie</i>	filmy



C9.57	select detective movie	<i>detective</i>	detektivka
C9.58	select thriller movie	<i>thriller</i>	thriller
C9.59	select adventure program	<i>adventure</i>	dobrodružný_film
C9.60	select western movie	<i>western</i>	western
C9.61	select war movie	<i>war</i>	válečný_film
C9.62	select science fiction channel	<i>science_fiction</i>	sci-fi
C9.63	select fantasy movie	<i>fantasy</i>	fantastický_film
C9.64	select horror film	<i>horror</i>	horor
C9.65	select comedy channel	<i>comedy</i>	komedie
C9.66	select soap opera	<i>soap</i>	telenovela   seriál
C9.67	select melodrama movie	<i>melodrama</i>	melodram
C9.68	select romance movie	<i>romance</i>	romantický_film
C9.69	select serious movie	<i>serious</i>	vážný_film
C9.70	select historical program	<i>historical</i>	historický_film
C9.71	select discussion program	<i>discussion</i>	diskusní_pořad
C9.72	select game show	<i>game_show   quiz   contest</i>	hra   soutěž
C9.73	select variety show	<i>variety_show</i>	varieté   estráda
C9.74	select talk show	<i>talk_show</i>	talk-show
C9.75	select football match	<i>football   soccer</i>	fotbal
C9.76	select tennis match	<i>tennis</i>	tenis
C9.77	select basketball match	<i>basketball</i>	basketbal
C9.78	select athletics show	<i>athletics</i>	atletika
C9.79	select motor sport	<i>motor_sport</i>	motoristické_sporty
C9.80	select water sport	<i>water_sport</i>	vodní_sporty
C9.81	select winter sport	<i>winter_sports</i>	zimní_sporty
C9.82	select martial sport	<i>martial_sports</i>	bojové_sporty
C9.83	select cartoons channel	<i>cartoons</i>	kreslené_filmy
C9.84	select music	<i>music</i>	hudba
C9.85	select culture program	<i>arts   culture</i>	umění   kultura
C9.86	select fashion show	<i>fashion</i>	móda
C9.87	select economy program	<i>economics   financial_services</i>	ekonomika
C9.88	select program on nature	<i>nature   animals   environment</i>	přírodopisný_pořad
C9.89	select fitness show	<i>medicine   health   fitness</i>	zdraví
C9.90	select hobby channel	<i>hobby</i>	hobby
C9.91	select cooking program	<i>cooking</i>	vaření
C9.92	select advertisement show	<i>advertisement   shopping</i>	reklama
C9.93	select gardening show	<i>gardening</i>	zahrádkářství
C9.94	select original language of the movie	<i>original_language</i>	původní_znění
C9.95	select live event	<i>live</i>	živé_vysílání

For the corpus selection, the corpus identifier of synonyms of application specific commands is replaced by a special corpus identifier Y to ensure an unequivocal discrimination of the

recordings of the synonyms. The replaced item identifiers of synonyms are listed in the following table:

<b>Adult database</b>					
<b>ID</b>	<b>Primary word</b>	<b>Y-ID</b>	<b>Synonym 1</b>	<b>Y-ID</b>	<b>Synonym 2</b>
C1.7	přihlásit_se	Y01	zalogovat_se		
C1.9	odhlásit_se	Y02	odlogovat_se		
C1.17	kamera	Y03	videokamera		
C1.31	set-top_box	Y04	STB		
C1.35	aktivovat	Y05	spustit		
C1.37	konec	Y06	ukončit	Y07	zrušit
C1.48	přepnout	Y08	zapnout	Y09	vypnout
C1.52	čeština	Y10	česky		
C1.53	trénovat	Y11	učit_se		
C1.61	alarm	Y12	budík		
C1.62	denně	Y13	každý_den		
C1.63	týdně	Y14	každý_týden		
C1.64	měsíčně	Y15	každý_měsíc		
C1.65	dříve	Y16	předchozí		
C1.66	později	Y17	následující		
C1.70	reset	Y18	výchozí_nastavení		
C1.82	hostitel	Y19	počítač		
C2.1	menu	Y20	nabídka		
C2.6	hlavní_menu	Y21	hlavní_nabídka		
C2.18	doleva	Y22	nalevo		
C2.19	doprava	Y23	napravo		
C2.24	modrý	Y24	modrá		
C2.25	červený	Y25	červená		
C2.26	žlutý	Y26	žlutá		
C2.27	zelený	Y27	zelená		
C2.28	černý	Y28	černá		
C2.29	bílý	Y29	bílá		
C2.30	hledat	Y30	najít		
C2.34	jít_na	Y31	skočit_na		
C2.37	stránka	Y32	stránku		
C2.38	vybrat	Y33	označit		
C2.39	vybrat_vše	Y34	vše		
C2.40	přeskočit	Y35	ignorovat		
C3.1	nový	Y36	nová		
C3.13	vymazat	Y37	odstranit	Y38	smazat
C3.20	upravit	Y39	editovat		
C4.1	profil	Y40	nastavení		
C4.2	standardní	Y41	implicitní	Y42	normální
C4.5	úroveň	Y43	stupeň		
C4.6	horní	Y44	vysoká		
C4.7	spodní	Y45	nízká		
C4.11	kanál	Y46	program		

C4.14	minimalizovat	Y47	skrýt		
C4.18	formát	Y48	rozměry		
C4.19	stupnice	Y49	měřítka		
C4.22	zmrazit	Y50	zastavit		
C4.23	světlo	Y51	osvětlení		
C4.33	ztlumit	Y52	zeslabit		
C4.38	melodie	Y53	vyzvánění		
C4.41	hloubky	Y54	basy		
C4.43	zvýraznit hloubky	Y55	zvýraznit_basy		
C4.55	vyjmout	Y56	vysunout		
C5.3	odpovědět	Y57	odpověď		
C5.5	přeposlat	Y58	forwardovat		
C5.6	poslat	Y59	odeslat		
C5.10	hlasové připomenutí	Y60	záznamník		
C5.13	schránka	Y61	mailbox		
C5.14	přijmout	Y62	stáhnout		
C5.15	inbox	Y63	příchozí pošta		
C5.20	poslat	Y64	přenést		
C5.26	odkaz	Y65	link		
C5.27	hovor	Y66	telefon		
C5.28	volit	Y67	vytočit		
C5.48	asistent	Y68	asistentka		
C5.51	práce	Y69	zaměstnání		
C5.64	nepřijatá volání	Y70	zmeškaná volání		
C5.67	hlasová schránka	Y71	záznamník		
C5.68	příjezd	Y72	přilet		
C5.69	odjezd	Y73	odlet		
C6.7	agenda	Y74	program		
C6.8	schůzka	Y75	jednání		
C6.11	harmonogram	Y76	program		
C6.14	příslušenství	Y77	nástroje		
C6.22	osobní	Y78	soukromý		
C6.28	PSC	Y79	směrovací číslo		
C7.1	pozice	Y80	lokalizace		
C7.21	konec cesty	Y81	cíl		
C7.26	křižovatka	Y82	exit		
C7.27	výstava	Y83	veletrh	Y84	výstaviště
C7.30	čerpací stanice	Y85	pumpa		
C8.9	počítač	Y86	PC		
C8.11	benzín	Y87	nafta		
C9.1	surround	Y88	prostorový zvuk		
C9.24	index	Y89	značka		
C9.35	digitální rádio	Y90	DAB		
C9.36	dopravní zprávy	Y91	zelená vlna		
C9.66	telenovela	Y92	seriál		
C9.72	hra	Y93	soutěž		
C9.73	varieté	Y94	estráda		
C9.85	umění	Y95	kultura		

For the corpus selection, multiple occurrences of application specific commands or their synonyms are removed to avoid multiple recordings of the same word. The removed duplicates are listed in the following table.

<b>Word/Phrase</b>	<b>Recorded as</b>	<b>Also found as...</b>	
<b>Adult database</b>			
<b>General Commands</b>			
adresář	<b>C2.2</b>	C6.20	
aktivovat	<b>C1.4</b>	C1.35	
číslo	<b>C3.21</b>	C1.55	
další	<b>C5.49</b>	C2.20	
e-mail	<b>C5.12</b>	C6.33	
fax	<b>C5.17</b>	C6.32	
hands-free	<b>C4.49</b>	C5.57	
kultura	<b>C9.42</b>	Y95	
minidisk	<b>C9.18</b>	C1.25	
mobil	<b>C5.45</b>	C1.29	
nastavení	<b>Y40</b>	C1.73	
opakovat	<b>C1.85</b>	C3.17	
označit	<b>C3.14</b>	Y33	
počítač	<b>Y19</b>	C8.9	C1.33
poslat	<b>C5.6</b>	C5.20	
potvrdit	<b>C1.13</b>	C1.42	
profil	<b>C4.1</b>	C5.53	
program	<b>Y46</b>	Y74	Y76
předchozí	<b>Y16</b>	C2.22	
přidat	<b>C3.11</b>	C4.9	
přijmout	<b>C5.7</b>	C5.14	C5.62
připomenout	<b>C5.35</b>	C6.13	
servis	<b>C7.16</b>	C7.29	
schůzka	<b>C6.8</b>	C5.58	
síť	<b>C1.83</b>	C1.79	
telefon	<b>Y66</b>	C6.31	C1.34
ukončit	<b>Y06</b>	C1.14	
vložit	<b>C3.10</b>	C3.9	
vybrat	<b>C2.38</b>	C5.34	
vyjmout	<b>C3.8</b>	C4.55	
vymazat	<b>C3.13</b>	C4.15	
vypnout	<b>C1.2</b>	Y09	
zapnout	<b>C1.1</b>	Y08	
záznamník	<b>Y60</b>	Y71	
zpět	<b>C2.23</b>	C3.16	
zrušit	<b>C1.5</b>	Y07	

## 6. Recording platforms

Basically the recording platform consists of a suitcase which contains all necessary audio equipment (microphones, preamplifiers, loudspeaker, CD player) as well as cables and adapters. Due to the different recording environments (public places, cars,...) the platform is mobile and contains a battery to make the platform independent of the main power supply. Four channels can be recorded simultaneously and the following table gives an overview of the used microphones and their mounting positions used within the different recording scenarios.

SCENARIO	CLOSE DISTANCE		MEDIUM DISTANCE		FAR DISTANCE
office, entertainment	Sennheiser ME 104	Nokia Lavalier HDC-6D	Sennheiser ME 64	-	Mikrofonbau Haun MBNM-550 E-L
public places	Sennheiser ME 104	Nokia Lavalier HDC-6D	Sennheiser ME 64	Mikrofonbau Haun MBNM-550 E-L	-
car	Sennheiser ME 104	Nokia Lavalier HDC-6D	AKG Q400 Mk3 T	Peiker ME15/V520-1	-

Table 30 – Microphone types in different environments

For the recordings an *Acer TravelMate, Type 273LC* (Mobile Intel Pentium 4, 1.7GHz, 512MB DDR SDRAM, 30GB Ultra ATA/100 HDD, DVD/CD-RW combo-drive) was used in Prague, *Acer TravelMate, Type 533LC* (Mobile Intel Pentium 4, 2 GHz, 256 MB DDR SDRAM, 30GB Ultra ATA/100 HDD, DVD/CD-RW combo-drive) was used in Brno. To lengthen battery life of notebook computers, special power devices were mounted both in Prague and in Brno, containing 3, resp. 2 traction batteries (12V, 50 Ah), and voltage controllers. Two VXpocket V2 interface cards and as recording software an adapted version of the ‘Mobile Recording Studio’ from Sony were used. The software provides full control, prompting and monitoring facilities during the recording and was placed at SpeeCon partners’ disposal by Sony. The software allows a monitoring and prompting on different screens. A very detailed description of the recording platform as well as the definitions of the recording scenarios and environments can be found in D2.1.2-v3.3 which is public and available on the SpeeCon Website.

**High-pass filter was not used in all recorded sessions in all environments .**

## 7. Recording environments

NOTE: All numbers of recorded speakers in particular categories and places are relevant to the main 550 speakers. Additional speakers (sessions 550-589) are not counted here. These sessions are described separately with recording conditions at the end of this document.

**Office:**

An office, i.e. a room where people are working at desks, usually or possibly with a computer. No discussions or meetings should be held in the office during the recordings.

$$L_{Aeq} = 30 - 60\text{dBA}$$

**Entertainment:**

Living room i.e. a room with some furniture, places where people may sit down. A table, a TV or some audio equipment may be present. Instead of a living room also a hotel room may be possible.

$$L_{Aeq} = 30 - 65\text{dBA}$$

**Public Place :**

A very large room (hall) or open-air.

A hall should have at least 3 walls and a ceiling; more or less busy people, but not too quiet.

An open area has no walls and no closed ceiling. Of course, it can be marked off by the walls of the surrounding building. In such a case, at most 2 walls may be closer than 10 meters. This allows recordings at the corner of two buildings. In all cases trees, small shops, an open cafe area, traffic as well as a pedestrian way are possible.

$$L_{Aeq} = 45 - 90\text{dBA}$$

**Car :**

Vehicle for 4 or 5 passengers,  $L_{Aeq} = 28 - 80\text{dBA}$

## 7.1 Office environment

The *Office* environment was divided into several sub-environments. The relevant parameters for the subdivision are the office room itself (Place) and the distance of the speaker to the wall. The speaker was considered close to the nearest wall if the distance was 1m or less.

General overview of speakers in OFFICE environment is specified in Table 31, more detailed specification for each of 15 recorded offices is in Table 32.

	<i>Min</i>	<i>Max</i>	Recorded
<i>CLOSE WALL</i>	80	120	104
<i>FAR WALL</i>	80	120	100
<i>MALES</i>	60	140	99
<i>FEMALES</i>	60	140	105
<b>TOTAL</b>	<b>200</b>		<b>204</b>

**Table 31 – General speaker coverage for OFFICE environment**

<b>OFFICE_01</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	6	5	2	0
	<i>Females</i>	2	4	1	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	9	4	3	10
	<i>Females</i>	1	6	1	0
<b>OFFICE_02</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	3	0	0	0
	<i>Females</i>	0	1	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	1	1	0	0
	<i>Females</i>	0	2	0	0
<b>OFFICE_03</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	1	3	0	0
	<i>Females</i>	9	6	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	1	0	0	0
	<i>Females</i>	4	0	0	0
<b>OFFICE_04</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	5	0	0	0
	<i>Females</i>	3	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	2	0	0	0
	<i>Females</i>	3	0	0	0
<b>OFFICE_05</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	3	0	0	0
	<i>Females</i>	3	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	1	0	0	0
	<i>Females</i>	5	0	0	0
<b>OFFICE_06</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	2	0	0	0
	<i>Females</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	1	0	0	0
	<i>Females</i>	1	0	0	0
<b>OFFICE_07</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	4	2	0	0
	<i>Females</i>	6	2	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	5	0	0	0
	<i>Females</i>	5	0	0	0
<b>OFFICE_08</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	2	0	0	0
	<i>Females</i>	1	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	1	0	0	0
	<i>Females</i>	2	0	0	0

OFFICE_09		CLOSE_1	CLOSE_2	CLOSE_3	CLOSE_4
	Males	3	0	0	0
	Females	6	0	0	0
		FAR_1	FAR_2	FAR_3	FAR_4
	Females	10	0	0	0
OFFICE_10		CLOSE_1	CLOSE_2	CLOSE_3	CLOSE_4
	Males	1	0	0	0
	Females	2	0	0	0
		FAR_1	FAR_2	FAR_3	FAR_4
	Females	1	0	0	0
OFFICE_11		CLOSE_1	CLOSE_2	CLOSE_3	CLOSE_4
	Males	3	0	0	0
	Females	0	0	0	0
		FAR_1	FAR_2	FAR_3	FAR_4
	Females	1	0	0	0
OFFICE_12		CLOSE_1	CLOSE_2	CLOSE_3	CLOSE_4
	Males	0	0	0	0
	Females	2	0	0	0
		FAR_1	FAR_2	FAR_3	FAR_4
	Females	0	0	0	0
OFFICE_13		CLOSE_1	CLOSE_2	CLOSE_3	CLOSE_4
	Males	3	0	0	0
	Females	2	0	0	0
		FAR_1	FAR_2	FAR_3	FAR_4
	Females	5	0	0	0
OFFICE_14		CLOSE_1	CLOSE_2	CLOSE_3	CLOSE_4
	Males	1	0	0	0
	Females	3	0	0	0
		FAR_1	FAR_2	FAR_3	FAR_4
	Females	1	0	0	0
OFFICE_15		CLOSE_1	CLOSE_2	CLOSE_3	CLOSE_4
	Males	0	0	0	0
	Females	2	0	0	0
		FAR_1	FAR_2	FAR_3	FAR_4
	Females	2	0	0	0

Table 32 – Detail description of recordings per recorded places in OFFICE environment

## 7.2 Entertainment environment

The *Entertainment* environment was divided into several sub-environments. The relevant parameters for the subdivision are the use of audio equipment, and the distance of the speaker to the wall. The speaker was considered close to the nearest wall if the distance was less 1m.

General overview of speakers in ENTERTAINMENT environment in specified in Table 33, more detailed specification for each from 16 recorded places is in Table 33.



	<i>Min</i>	<i>Max</i>	Recorded
CLOSE_WALL	25	50	40
FAR_WALL	25	50	34
MALES	22	50	32
FEMALES	22	50	42
AUDIO_ON	25	50	41
AUDIO_OFF	25	50	33
<b>TOTAL</b>	<b>72</b>		<b>74</b>

Table 33 – General speaker coverage for ENTERTAINMENT environment

<b>ENTERTAIN_01</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	1	0	0	0
	<i>Females</i>	1	0	0	0
	<i>Audio on</i>	2	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	1	0	0	0
	<i>Females</i>	0	0	0	0
<b>ENTERTAIN_02</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	1	0	0	0
	<i>Females</i>	0	0	0	0
	<i>Audio on</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
<b>ENTERTAIN_03</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
	<i>Audio on</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	3	0	0	0
	<i>Females</i>	5	0	0	0
<b>ENTERTAIN_04</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
	<i>Audio on</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	1	0	0	0
	<i>Females</i>	3	0	0	0
<b>ENTERTAIN_05</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	6	0	0	0
	<i>Females</i>	0	0	0	0
	<i>Audio on</i>	3	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
	<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>	
<i>Males</i>	0	0	0	0	
<i>Females</i>	0	0	0	0	
<i>Audio on</i>	0	0	0	0	

<b>ENTERTAIN_06</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	Males	0	0	0	0
	Females	1	0	0	0
	Audio on	0	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
Audio on	0	0	0	0	
<b>ENTERTAIN_07</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	Males	0	0	0	0
	Females	4	0	0	0
	Audio on	4	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
Audio on	0	0	0	0	
<b>ENTERTAIN_08</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
	Audio on	0	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	Males	2	0	0	0
	Females	1	0	0	0
Audio on	3	0	0	0	
<b>ENTERTAIN_09</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	Males	3	0	0	0
	Females	3	0	0	0
	Audio on	6	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
Audio on	0	0	0	0	
<b>ENTERTAIN_10</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	Males	1	0	0	0
	Females	2	0	0	0
	Audio on	0	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
Audio on	0	0	0	0	
<b>ENTERTAIN_11</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
	Audio on	0	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	Males	5	0	0	0
	Females	4	0	0	0
Audio on	0	0	0	0	

<b>ENTERTAIN_12</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	1	0	0	0
	<i>Females</i>	5	0	0	0
	<i>Audio on</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
	<i>Audio on</i>	0	0	0	0
<b>ENTERTAIN_13</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
	<i>Audio on</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	2	0	0	0
	<i>Females</i>	6	0	0	0
	<i>Audio on</i>	4	0	0	0
<b>ENTERTAIN_14</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	2	0	0	0
	<i>Females</i>	3	0	0	0
	<i>Audio on</i>	1	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
	<i>Audio on</i>	0	0	0	0
<b>ENTERTAIN_15</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
	<i>Audio on</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	1	0	0	0
	<i>Audio on</i>	1	0	0	0
<b>ENTERTAIN_16</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	3	0	0	0
	<i>Females</i>	3	0	0	0
	<i>Audio on</i>	6	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
	<i>Audio on</i>	0	0	0	0

Table 34 – Detail description of recordings per recorded places in ENTERTAINMENT environment

### 7.3 Public environment

The *Public* environment was divided into several sub-environments. The relevant parameters for the subdivision are the room itself (Place), the place category (Hall or Open), and the distance of the speaker to the wall. The speaker was considered close to the nearest wall if the distance was 2 m or less; the speaker was considered as speaking without wall reverberations (No Wall) if the distance was larger than 10 m.

General overview of speakers in PUBLIC environment is specified in Table 35, more detailed specification for each from 10 recorded halls and 12 recorded open places is in Table 36.

	<i>Min</i>	<i>Max</i>	<i>Recorded</i>
MALES	60	140	95
FEMALES	60	140	104
CLOSE_HALL	40	70	49
FAR_HALL	40	70	45
CLOSE_OPEN	20	45	35
FAR_OPEN	20	45	44
NO_OPEN	20	45	26
TOTAL_HALL	90	110	94
TOTAL_OPEN	90	110	105
<b>TOTAL</b>	<b>200</b>		<b>199</b>

Table 35 – General speaker coverage for PUBLIC environment

<b>PUBHALL_01</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
<b>PUBHALL_02</b>	<i>Males</i>	1	0	0	0
	<i>Females</i>	0	0	0	0
		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	3	0	0	0
<b>PUBHALL_03</b>	<i>Females</i>	5	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
<b>PUBHALL_04</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	9	2	0	0
	<i>Females</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
<b>PUBHALL_05</b>	<i>Males</i>	6	0	0	0
	<i>Females</i>	3	0	0	0
		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	3	0	0	0
<b>PUBHALL_06</b>	<i>Females</i>	2	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
<b>PUBHALL_07</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
<b>PUBHALL_08</b>	<i>Males</i>	0	4	0	0
	<i>Females</i>	10	6	5	0
		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	0	0	0	0
<b>PUBHALL_09</b>	<i>Females</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	2	3	0	0
	<i>Females</i>	4	1	0	0

<b>PUBHALL_07</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	Males	5	0	0	0
	Females	1	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
<b>PUBHALL_08</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	Males	5	0	0	0
	Females	3	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
<b>PUBHALL_09</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	Males	2	0	0	0
	Females	4	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
<b>PUBHALL_10</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	Males	2	0	0	0
	Females	3	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	Males	0	0	0	0
	Females	0	0	0	0

<b>PUBOPEN_01</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	Males	3	0	0	0
	Females	1	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
		<b>NO_1</b>	<b>NO_2</b>	<b>NO_3</b>	<b>NO_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
<b>PUBOPEN_02</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	Males	1	0	0	0
	Females	3	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
		<b>NO_1</b>	<b>NO_2</b>	<b>NO_3</b>	<b>NO_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
<b>PUBOPEN_03</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	Males	0	0	0	0
	Females	0	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	Males	2	0	0	0
	Females	8	0	0	0
		<b>NO_1</b>	<b>NO_2</b>	<b>NO_3</b>	<b>NO_4</b>
	Males	3	0	0	0
	Females	6	0	0	0

<b>PUBOPEN_04</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	<i>Males</i>	1	0	0	0
	<i>Females</i>	2	2	0	0
		<b>NO_1</b>	<b>NO_2</b>	<b>NO_3</b>	<b>NO_4</b>
	<i>Males</i>	2	0	0	0
<i>Females</i>	4	0	0	0	
<b>PUBOPEN_05</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	<i>Males</i>	2	2	0	0
	<i>Females</i>	2	5	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
		<b>NO_1</b>	<b>NO_2</b>	<b>NO_3</b>	<b>NO_4</b>
	<i>Males</i>	0	0	0	0
<i>Females</i>	0	0	0	0	
<b>PUBOPEN_06</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	<i>Males</i>	3	0	0	0
	<i>Females</i>	3	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	<i>Males</i>	1	0	0	0
	<i>Females</i>	4	0	0	0
		<b>NO_1</b>	<b>NO_2</b>	<b>NO_3</b>	<b>NO_4</b>
	<i>Males</i>	0	0	0	0
<i>Females</i>	0	0	0	0	
<b>PUBOPEN_07</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	<i>Males</i>	4	0	0	0
	<i>Females</i>	3	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	<i>Males</i>	5	0	0	0
	<i>Females</i>	1	0	0	0
		<b>NO_1</b>	<b>NO_2</b>	<b>NO_3</b>	<b>NO_4</b>
	<i>Males</i>	0	0	0	0
<i>Females</i>	0	0	0	0	
<b>PUBOPEN_08</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
		<b>NO_1</b>	<b>NO_2</b>	<b>NO_3</b>	<b>NO_4</b>
	<i>Males</i>	4	0	0	0
<i>Females</i>	2	0	0	0	
<b>PUBOPEN_09</b>		<b>CLOSE_1</b>	<b>CLOSE_2</b>	<b>CLOSE_3</b>	<b>CLOSE_4</b>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
		<b>FAR_1</b>	<b>FAR_2</b>	<b>FAR_3</b>	<b>FAR_4</b>
	<i>Males</i>	1	0	0	0
	<i>Females</i>	0	0	0	0
		<b>NO_1</b>	<b>NO_2</b>	<b>NO_3</b>	<b>NO_4</b>
	<i>Males</i>	0	0	0	0
<i>Females</i>	0	0	0	0	

<b>PUBOPEN_10</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	4	0	0	0
	<i>Females</i>	2	0	0	0
		<i>NO_1</i>	<i>NO_2</i>	<i>NO_3</i>	<i>NO_4</i>
	<i>Males</i>	0	0	0	0
<b>PUBOPEN_11</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	0	0	0	0
	<i>Females</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	1	0	0	0
	<i>Females</i>	4	0	0	0
		<i>NO_1</i>	<i>NO_2</i>	<i>NO_3</i>	<i>NO_4</i>
	<i>Males</i>	0	0	0	0
<b>PUBOPEN_12</b>		<i>CLOSE_1</i>	<i>CLOSE_2</i>	<i>CLOSE_3</i>	<i>CLOSE_4</i>
	<i>Males</i>	3	0	0	0
	<i>Females</i>	0	0	0	0
		<i>FAR_1</i>	<i>FAR_2</i>	<i>FAR_3</i>	<i>FAR_4</i>
	<i>Males</i>	4	0	0	0
	<i>Females</i>	2	0	0	0
		<i>NO_1</i>	<i>NO_2</i>	<i>NO_3</i>	<i>NO_4</i>
	<i>Males</i>	2	0	0	0
<i>Females</i>	3	0	0	0	

**Table 36 – Detail description of recordings per recorded places in PUBLIC environment**

## 7.4 Car environment

General overview of speakers in CAR environment is specified in Table 37, more detailed specification for each from 7 recorded MIDDLE class cars and 3 recorded UPPER class cars is in Table 38.

All categories in both car classes were recorded. For each category is minimal representation above 5. In three cases (CARUPPER & COUNTRY, CARMIDDLE & ENGINE\_OFF, CARMIDDLE & CITY) we over-recorded these category when 10 speakers were recorded (it was allowed to record 9 speakers). It happens within re-recording of bad sessions and we hope that this inconsistency is not too critical, moreover, when the coverage of other conditions is not extremely un-balanced.

	<i>Min</i>	<i>Max</i>	<i>Recorded</i>
<b>CATEGORY CAR MIDDLE</b>			
<i>ENGINE_OFF</i>	3	9	10
<i>ENGINE_ON</i>	3	9	6
<i>CITY_30_70</i>	3	9	10
<i>COUNTRY_60_100</i>	3	9	8
<i>HIGHWAY_90_130</i>	3	9	5
<b>MIDDLE_TOTAL</b>	<b>30</b>	<b>45</b>	<b>39</b>
<b>CATEGORY CAR UPPER</b>			
<i>ENGINE_OFF</i>	3	9	5
<i>ENGINE_ON</i>	3	9	5
<i>CITY_30_70</i>	3	9	7
<i>COUNTRY_60_100</i>	3	9	10
<i>HIGHWAY_90_130</i>	3	9	7
<b>UPPER_TOTAL</b>	<b>30</b>	<b>45</b>	<b>34</b>
<b>CAR TOTAL BALANCE</b>			
<i>MALES</i>	22	50	49
<i>FEMALES</i>	22	50	24
<b>CAR TOTAL</b>	<b>72</b>		<b>73</b>

Table 37 – General speaker coverage for CAR environment

		<i>OFF</i>	<i>ON</i>	<i>CITY</i>	<i>COUNTRY</i>	<i>HIGHWAY</i>
<b>CAR_MIDDLE_01</b>	<i>Males</i>	2	1	1	0	1
	<i>Females</i>	2	3	2	1	0
<b>CAR_MIDDLE_02</b>	<i>Males</i>	0	0	1	2	0
	<i>Females</i>	1	0	1	0	0
<b>CAR_MIDDLE_03</b>	<i>Males</i>	0	0	0	1	0
	<i>Females</i>	0	0	0	0	0
<b>CAR_MIDDLE_04</b>	<i>Males</i>	0	2	4	1	1
	<i>Females</i>	0	0	1	1	0
<b>CAR_MIDDLE_05</b>	<i>Males</i>	0	0	0	1	0
	<i>Females</i>	0	0	0	0	0
<b>CAR_MIDDLE_06</b>	<i>Males</i>	5	0	0	0	0
	<i>Females</i>	0	0	0	0	0
<b>CAR_MIDDLE_07</b>	<i>Males</i>	0	0	0	0	2
	<i>Females</i>	0	0	0	1	1
<b>CAR_UPPER_01</b>	<i>Males</i>	0	0	3	4	1
	<i>Females</i>	0	0	3	2	4
<b>CAR_UPPER_02</b>	<i>Males</i>	5	4	0	0	0
	<i>Females</i>	0	0	0	0	0
<b>CAR_UPPER_03</b>	<i>Males</i>	0	0	1	4	2
	<i>Females</i>	0	1	0	0	0

Table 38 – Detail description of recordings per recorded places in CAR environment



## 8. Speaker distribution

### 8.1 Age and Gender distribution

Speaker distribution was controlled from the point of view of age and gender. Three age group for adult speakers were defined in D215, with minimal required coverage in database which is specified in the second column of the following table:

For adult speakers uniform gender coverage is required globally for the entire database. Also for each age category, gender distribution must be with 5% tolerance. Therefore, the percentage of males and females speakers must be in the range 45%-55% globally and in each age category. Good speaker age and gender distributions were finally reached in ADULT1CS database, see the following table:

Age group	Required %	Recorded Spks		In %
		Males	Females	
(VC) - 30		133	130	
		50,57%	49,43%	
	>= 30 %	263		47,82%
31 - 45		91	98	
		48,15%	51,85%	
	>= 30 %	189		34,36%
46 +		51	47	
		52,04%	47,96%	
	>= 10 %	98		17,82%
Total		275	275	
		50,00%	50,00%	
		550		100,00%

Table 39 – Age and Gender distribution for ADULT1CS database

### 8.2 Dialect distribution

Four dialect regions were defined for the collection of SpeeCon in Czech Republic, see Table 40. The dialects vary little in phoneme pronunciations, partially in vocabulary and there are only very few and small differences in the grammar. Speaking any of the above dialects, people will still understand each other; therefore we do not need to translate a dialect to the official Czech for those not speaking it.

The dialect regions were defined with co-operation of phonetic and linguistic specialist prof. Zdena Hladka from Masaryk University in Brno during the collection of Czech SpeechDat database. For the purposes of SpeeCon project, dialect regions East Moravia (EM) and Silesia (S) were joined into one region (EMS). The main reason for this is the fact, that fulfilling the requirements defined in D215 could lead to unreasonable enhancement of these two dialects in comparison to the percentage of the population in these regions. Similarly, West Bohemia (Plzen) and South Bohemia (České Budejovice) dialects were previously joined into one SB region.

CWNB	<b>Central-West-North Bohemia</b> (Prague, Cheb, Liberec, Hradec Kralove)
SB	<b>South Bohemia</b> (Ceske Budejovice, Plzen, Jihlava)
CM	<b>Central Moravia</b> (Brno, Olomouc)
EMS	<b>East Moravia and Silesia</b> (Uherske Hradiste, Ostrava, Opava)

**Table 40 – Dialect regions in Czech Republic**

Minimal number of speakers in each dialect is given again by D215, and for 4 dialect regions and 550 adult speakers it is 97. For each dialect, both genders must be uniformly covered with 20% tolerance, i.e. percentage of males and females must be always in the range 30%-70% in each dialect group. Finally reached speaker distribution from the dialect point of view is shown in the following table:

<i>Dialect</i>	<i>Required</i>	<i>Recorded spks</i>	
		<i>Males</i>	<i>Females</i>
CWNB		76	86
		46,91%	53,09%
	97	162	
SB		62	45
		57,94%	42,06%
	97	107	
CM		82	68
		54,67%	45,33%
	97	150	
EMC		55	76
		41,98%	58,02%
	97	131	
<b>Total</b>		<b>275</b>	<b>275</b>
		<b>50,00%</b>	<b>50,00%</b>
		<b>550</b>	

**Table 41 – Final dialect coverage in ADULTICS database**

For the OFFICE and PUBLIC environments, the minimal coverage of each dialect is required. For 200 speakers per each of those environments and 4 dialect regions, it is minimally 25 speakers per dialect both in OFFICE and PUBLIC environment. Reached speaker coverage in these two environment per dialect is shown in the Table 42 .

<i>Dialect</i>	<i>Required</i>	<i>Recorded Office</i>	<i>Recorded Public</i>
CWNB	25	76	76
SB	25	31	40
CM	25	32	42
EMC	25	65	41

**Table 42 – Dialect representation in OFFICE and PUBLIC environment**

## 9. Annotations

### 9.1 General description of annotations

Czech SpeeCon database was annotated orthographically with check of real pronunciation of all utterances. The exceptions from pronunciation were marked and finally used for the generation of phonetic lexicon. The pronunciation of each utterance is kept in label-file as EPI field.

The transcription software was FTP-Transcriber written by Petr Schwarz from VUT Brno for annotation of SpeechDat database. It was updated for the annotation of SpeeCon. It is running under Windows 95/98/NT. It was created using the programming environment C-Builder from Borland, with specialised components for FTP-connection from NET-Manage. The annotation is possible either via FTP connection directly in recorded data structure or locally with copied block of the data. Usually, the stand-alone mode (without FTP) was preferred by most annotators due to the speed of the transfers, and annotated data were transferred to server in one block.

For quality assurance, all transcriptions were automatically checked for syntax, spelling, etc. These checks are based on comparison with already checked lexicons for the SpeechDat database and with lexicon generated from SpeeCon prompts. Each difference in spelling or pronunciation was always checked by experts. Only successfully checked annotations were accepted.

Finally, several additional annotations were hand checked, especially for usage of marks, and mistakes had to be corrected commonly with similar mistakes in other files. The block of annotation is accepted only when just several minor errors appear. Otherwise, the block is re-checked with other selection to avoid the correction of previously checked annotations only.

The character set used for the transcription is ISO 8859-2 (ISO Latin 2). A sample ISO 8859-2 table in PostScript and PDF format can be found in the DOC directory.

Transcriptions are not case sensitive, except for spelling transcriptions - those are strictly uppercase (except for normal words as “dlouhý” (long), “přehlasované” (umlaut), etc). Punctuation marks are removed from the transcription except of word level punctuation as “bude-li”, “autorsko-spisovatelský”, etc.

Artificial words, word fragments, syllables with no sense, or naturally accented Czech words which are present in WEB and e-mail addresses without accents; all these Internet expressions are annotated with preceding underscore (e.g. \_usenet, \_muni, \_com, etc.). These entries are included in the lexicon with the underscore and they may have often several pronunciation variants. Words with underscore may be present only in items CW1 and CW2. The word may also appear in the lexicon in two forms, with and without underscore. It is typical for foreign words. The rule for usage is the following:

- 1) **underscore is not used** if the word is pronounced with correct regular pronunciation  
(download - d a \_ u n l o \_ u t, einstein - a j n S t a j n)
- 2) **underscore is used** were the word is pronounced with un-natural Czech-like pronunciations:  
(\_download - d o v n l o a t, \_einstein - e i n s t e i n)

Several syllables used in e-mails and WEB addresses are correct Czech words (se, to, ...). In these cases underscore was not used.

There are not alternative spellings for words in database so file SPELLALT.DOC is not included in the documentation directory.

Some English words (also from other languages) used mainly as several application words have sometimes no Czech equivalent and they are used with adopted Czech pronunciation. This pronunciation is marked and it is always expressed only with Czech phonetic set.

## 9.2 Annotations of spontaneous speech

Very special problem had to be solved for the annotation of spontaneous free speech items. Many slang expressions are used within these utterances. It is not possible to fix these slang expressions to written form of the word so the solution using pronunciation variants was not suitable. Finally, it was decided to transcribe these expressions “as it was said” with the closest form respecting as much as possible the rules of written Czech language. Typical examples of more or less strange words may be “seš, dobrej, vejvar, brýden, nashle, todleto, ...” etc.

## 9.3 Annotations of spelled items

The annotation of spelled utterances is summarised in the following table. The Czech spelling is used very irregularly. It is described with more details in Czech part of D216 (LSP). From the point of view of annotations, the following notes are important.

**Note 1.** Only the variants of spelling mentioned in Table 43 and Table 44 are included.

**Note 2.** Additional words connected to spelling as „dlouhé, krátké, s čárkou, s háčkem, ...“ (in English „long, short, with accent, etc...“) may vary. The variants mentioned in previous table are just examples, other words are allowed in annotations, to.

**Note 3.** It is difficult to distinguish short and long variant of vowels just by spelling, since both short and long pronunciation may appear in the spelling of both variants of vowels, i.e. the problem may be the recognition of „A“ and „/Á/“.

**Note 4.** Letter with accents are sometimes spelled with other additional words as “dlouhé, s čárkou, krátké, přehlasované, s kroužkem ...”. Although the occurrence of each letter is uniform in prompts, statistics done on transcription will be different because long vowel are usually spelled as “long A”. In fact it increases occurrence of short vowel spellings and decreases occurrence of pure long vowel spellings. Similar case (however more rare) may appear also for accented consonants.

No.	Letter	Spelling 1		Spelling 2		English spelling	
		Annotation	Pronunciation	Annotation	Pronunciation	Annotation	Pronunciation
1	A	A	a :	/A/	a	EJ	ej

2	Á	dlouhé A	dlo_uh\e: a:	/Á/	a:		
3	B	B	be:	/B/	b@	BÝ	bi:
4	C	C	t_se:	/C/	t_s@	SÝ	si:
5	Č	Č	t_Se:	/Č/	t_S@		
6	D	D	de:	/D/	d@	DÝ	di:
7	Ď	Ď	J\e:	/Ď/	J\@		
8	E	E	e:	/E/	e	EE	i:
9	É	dlouhé E	dlo_uh\e: e:	/É/	e:		
10	Ě	E s háčkem	e: s h\a:t_Skem	/Ě/	ije   je		
11	F	F	ef	/F/	f@		
12	G	G	ge:	/G/	g@	DŽÝ	d_zi:
13	H	H	h\a:	/H/	h\@	EJČ	ejt_s
14	Ch	Ch	xa:	/Ch/	x@		
15	I	I	i:	/I/	i	AJ	aj
16	Í	dlouhé I	dlo_uh\e: i:	/Í/	i:		
17	J	J	je:	/J/	j@	DŽEJ	d_zej
18	K	K	ka:	/K/	k@	KEJ	kej
19	L	L	el	/L/	l@		
20	M	M	em	/M/	m@		
21	N	N	en	/N/	n@		
22	Ň	Ň	eJ	/Ň/	J@		
23	O	O	o:	/O/	o	OU	o_u
24	Ó	dlouhé O	dlo_uh\e: o:	/Ó/	o:		
25	P	P	pe:	/P/	p@	PÝ	pi:
26	Q	Q	kve:	/Q/	kv@	KJÚ	kju:
27	R	R	er	/R/	r@	ÁR	a:r
28	Ř	Ř	eQ\	/Ř/	P\@		
29	S	S	es	/S/	s@		
30	Š	Š	eS	/Š/	S@		
31	T	T	te:	/T/	t@	TÝ	ti:
32	Ě	Ě	ce:	/Ě/	c@		
33	U	U	u:	/U/	u	JÚ	ju:
34	Ú	dlouhé U	dlo_uh\e: u:	/Ú/	u:		
35	Ů	U s kroužkem	u: s kro_uzkem	/Ů/	u:		
36	V	V	ve:	/V/	v@	VÝ	vi:
37	W	dvojitě V   W	dvojite: ve:   dvojve:	/W/	v@	double JÚ	dabl ju:
38	X	X	iks	/X/	ks@	EX	eks
39	Y	Y	ipsilon	/Y/	i	WAJ	uaj

40	Ý	dlouhé Y	dlo_uh\e: ippsilon	/Ý/	i:		
41	Z	Z	zet	/Z/	z@		
42	Ž	Ž	zet	/Ž/	z@		

**Table 43 – Annotation and pronunciation of Czech spelled letters**

Additional German letters					
No.	Letter	Spelling 1		Spelling 2	
		Annotation	Pronunciation	Annotation	Pronunciation
43	Ä	přehlasované A	pP\eh\lasovane: a:	/Ä/	E:
44	Ü	přehlasované U	pP\eh\lasovane: u:	/Ü/	y:
45	Ö	přehlasované O	pP\eh\lasovane: o:	/Ö/	2:

**Table 44 – Annotation and pronunciation of German spelled letters appearing in some surnames**

The following symbols specified in Table 45 are used to denote word truncations, mispronunciations, non-understandable speech, non-speech acoustic events.

word truncations	~word or word~		at signal begin or end only
mispronunciations	*word		
non-understandable speech	**		separated by blank from rest of text
non-speech acoustic events	[fil]	filled pause	at correct location between words
	[spk]	speaker noise	at correct location between words
	[int]	intermittent noise	at correct location between words
	[sta]	stationary noise	at the beginning of the noise (placed between words)

**Table 45 – Annotation of non-speech marks**

Word **truncation mark** – ‘~’ -describes truncation of the word at the signal beginning or end due to bad start or stop of utterance recording. This mark is not used for truncation by speaker. This is always marked as mispronunciation.

Examples of usage truncation mark are bellow.

[spk] čerpací stanice~

~český krumlov

[sta] ~můžem [spk] rád bych ovládal hlasem řízení celého automobilu [spk]

**Mispronunciation mark – ‘\*’** - is used in the cases in two different situations:

- 1) when some exact word (mostly prompted) was slightly mispronounced and when this variant cannot be mentioned as pronunciation variant - \*word, regular pronunciation variant also with \* mark is used in EPI field in this case;
- 2) when the pronounced word is unintelligible or just short fragment of the word is pronounced, mark \*\* replacing this word is used; this appears frequently in spontaneous free speech items;
- 3) extremely strange pronunciations of foreign words (bluetooth, pager, talk-show, etc.) were also marked as mispronunciations.

Examples of usage mispronunciation marks are bellow.

skláním \*rudou tvář před matkou

[spk] domluvil vzal žabku nařezal kytici růží maminka je svázala a \*zabalila

[spk] \*bluetooth

[spk] třeba nějaké \*\* nanebevzetí nebo jakubův žebřík řekl neurčitě

\*\* minimalizovat

[spk] varujeme všechny řidiče mezi ševětínem a českými budějovicemi \*\* mají policajti tři mikrovlňky

**Non-speech acoustic events** were annotated by four different marks: [spk], [fil] – for **speaker events**; [int], [sta] – **environment events**.

[spk] – This is the most frequently used mark, typically for the events like lip smack, cough, grunt, throat clear, tongue click, loud breath, laugh, loud sigh.

[fil] – This mark is used for “filled pause” between words. This kind of the event may be well modeled by speech filled pause model. That is the reason why this kind of the speaker noise is not covered by general [spk] mark. This mark was very frequently used in spontaneous free speech utterances.

[int] – This mark is used for the noise of generally intermittent nature. Typically, it appears once (door slam), or have pauses (telephone ringing), etc.

[sta] – This mark is used for of the noise which has less or more stable amplitude spectrum during the time. This mark is not used when such kind of the noise is typical for given environment and if it appears during whole session. It is used when the noise is present just in several utterances.

Examples of usage these marks are bellow.

[spk] pět hodin dvacet sedm minut [spk]

[spk] V V V tečka M P S V tečka C Z lomno [spk] skript lomno [spk]  
\_default tečka A S P

[spk] dobrý den mám zájem o [fil] váš zájezd do itálie [spk] a chtěla jsem  
[spk] se zeptat [fil] jak je jaké stravování jestli je tam polopenze nebo  
plná penze

[spk] [fil] dobrý den [fil] chci se zeptat jestli na [fil] jméno [spk]  
otáhalová nepřišel doporučený dopis

[spk] dobrý den chtěla jsem se zeptat [fil] jestli [spk] jsou ještě volné  
lístky na koncert [spk] shizuky ishikhavy [spk] v květnu v brně [int]

[sta] [spk] [fil] dobrý den chci se zeptat jestli jste dostali encyklopedii  
zvířat [spk] a kolik [fil] asi tak stojí

[spk] osum [int] osm nula čtyři [int] čtyři sedum dva dvě

[int] tři dva čtyry šest dva

## 10. Lexicon information

The pronunciation lexicon was derived from the annotations which is done commonly with marked exception from regular pronunciation. The pronunciation of typical Czech word is quite regular and it may be generated by several rules for conversion of orthographic transcription into the phonetic one. However, there is important number of words with exceptional pronunciation.

During the annotation with the above described FTP-Transcriber, the “regular” Czech phonetic transcription is on-line generated for each orthographic annotation. If the phonetic transcription is not corresponding with real pronunciation it is marked with special syntax, from which above mentioned LBO and EPI fields are then generated, and consequently also the pronunciation lexicon.

### 10.1 Czech SAMPA table

The pronunciation in CSO-files and in the lexicon is in Czech SAMPA which is summarized in the following Table 46.

Consonants			
Vowels			
Symbol	Word	Transcription	English translation & Remarks
i	myš, liška	miS, liSka	mouse, fox
e	les	les	forest
a	pas	pas	passport



o	rok	rok	year
u	kus	kus	piece
i:	pít, být	pi:t, bi:t	to drink, to beat
e:	lék	le:k	drug
a:	rád	ra:t	glad
o:	móda	mo:da	fashion RARE
u:	půl, únor	pu:l, u:nor	half, February
<b>Diphthongs</b>			
o_u	mouka	mo_uka	flour
a_u	auto	a_uto	car
e_u	euforie	e_uforie	euphoria RARE
<b>Plosives</b>			
p	pes	pes	dog
b	bota	bota	shoe
t	tam	tam	there
d	dům	du:m	house
c	tito	cito	these
J\	děd	J\et	grandfather
k	kolo	kolo	bike
g	kde	gde	where
<b>Affricates</b>			
t_s	cíl	t_si:l	aim
d_z	lecky	led_zgdi	at times RARE
t_S	čas	t_Sas	time
d_Z	džbán	d_Zba:n	jug RARE
<b>Fricatives</b>			
f	forma	forma	form
v	vak	vak	bag
s	sen	sen	dream
z	zub	zup	tooth
P\	řád	P\a:t	order
S	šaty	Sati	clothes
Z	žal	Zal	regret
j	jas	jas	brightness
x	chata	xata	cottage
h\	had	h\at	snake
<b>Liquids</b>			
r	ret	ret	lip
l	led	let	ice
<b>Nasals</b>			
m	mák	ma:k	poppy
n	noc	not_s	night
N	banka	baNka	bank
J	nic	Jit_s	nothing

Additional allophones			
F	tramvaj	traFvaj	tram RARE
Q\	tři	tQ\i	three (UNVOICED variant of P\)

**Table 46 – Used phonemes from Czech SAMPA**

It is useful to use also the following phonemes specified in Table 47 in pronunciation transcription, however these phonemes are not balanced in phonetic rich material. The reasons for it are the following.

- “@” - schwa - It is used just in the second unofficial (phonetic) version of spelling. In other utterances it practically does not appear. On the other hand, the set of spelling items will give sufficient training material for these phones in the most probable context.
- German phones “E: 2: y:” (letters ä, ö, ü) -These letters appear in Czech just in the names of German origin (Müller) etc. In spoken language their pronunciation is mapped to Czech phonemes, i.e. “E: -> e”, “2: -> e”, “y: -> i” or their long variants. The special German-like pronunciations appear therefore only within spelling items where some people would like to express umlaut version of the vowel.

Schwa			
@	DTW	d@ t@ v@	2-nd spelling of DTW
German vowels			
E:	ä	E:	spelling of umlaut vowels
2:	ö	2:	
y:	ü	y:	

**Table 47 – Additional phonemes used in lexicon**

## 10.2 Phonetic annotations – source for lexicon generation

All phonetic transcription are done at word level without the context. In this sense the words are included in the pronunciation lexicon. However, the change between unvoiced and voiced equivalents of the phonemes may appear due to cross-word context. This dependence was not used nor marked because it is out of scope of this project and consequently it may lead to several pronunciation variants for each entry in the lexicon.

As mentioned in the previous section, all items in the lexicon are lowercase except spelled letter which are strictly uppercase.

Multiple transcription was not provided.

No information about stress is supplied. In Czech the stress is regularly at the first syllable of the word or at the preposition before the word. Also any other linguistic tags were not used.

The results of frequency occurrence analysis of phonemes in the phonetically rich sentences, phonetically rich words and in the full database (at transcription level) is provided in the following tables. No statistics were evaluated for diphones and triphones. The analysis were provided on real pronunciations saved in EPI fields.

The following phonemes are considered as rare: **o:**, **e\_u**, **d\_z**, **d\_Z**, **F**. All phonemes are present in the database with the required coverage.

Phoneme	Sentences		Words		Sentences+Words		All DB	
	Occur	In %	Occur	In %	Occur	In %	Occur	In %
@	5	0,001	0	0,000	5	0,001	3765	0,137
<b>a_u</b>	2506	0,355	0	0,000	<b>2506</b>	<b>0,345</b>	5237	0,191
<b>a</b>	50003	7,082	1375	6,673	<b>51378</b>	<b>7,070</b>	197443	7,209
<b>a:</b>	15270	2,163	580	2,815	<b>15850</b>	<b>2,181</b>	69055	2,521
<b>b</b>	11922	1,688	399	1,936	<b>12321</b>	<b>1,696</b>	39375	1,438
<b>t_s</b>	8188	1,160	185	0,898	<b>8373</b>	<b>1,152</b>	44369	1,620
<b>t_S</b>	6147	0,871	123	0,597	<b>6270</b>	<b>0,863</b>	31011	1,132
<b>d</b>	18798	2,662	332	1,611	<b>19130</b>	<b>2,633</b>	99694	3,640
<b>J\</b>	4081	0,578	52	0,252	<b>4133</b>	<b>0,569</b>	10831	0,395
<b>e_u</b>	914	0,129	497	2,412	<b>1411</b>	<b>0,194</b>	1696	0,062
<b>e</b>	69041	9,778	1703	8,265	<b>70744</b>	<b>9,735</b>	264078	9,642
<b>e:</b>	6663	0,944	136	0,660	<b>6799</b>	<b>0,936</b>	30519	1,114
<b>f</b>	6145	0,870	559	2,713	<b>6704</b>	<b>0,923</b>	22931	0,837
<b>g</b>	2968	0,420	38	0,184	<b>3006</b>	<b>0,414</b>	9047	0,330
<b>x</b>	5697	0,807	82	0,398	<b>5779</b>	<b>0,795</b>	17782	0,649
<b>h\</b>	8659	1,226	132	0,641	<b>8791</b>	<b>1,210</b>	32258	1,178
<b>j</b>	21265	3,012	519	2,519	<b>21784</b>	<b>2,998</b>	71530	2,612
<b>k</b>	25214	3,571	487	2,364	<b>25701</b>	<b>3,537</b>	87940	3,211
<b>l</b>	36408	5,156	612	2,970	<b>37020</b>	<b>5,094</b>	106546	3,890
<b>F</b>	926	0,131	539	2,616	<b>1465</b>	<b>0,202</b>	1543	0,056
<b>m</b>	26282	3,722	871	4,227	<b>27153</b>	<b>3,737</b>	95159	3,475
<b>N</b>	2741	0,388	0	0,000	<b>2741</b>	<b>0,377</b>	5472	0,200
<b>n</b>	31092	4,403	1142	5,542	<b>32234</b>	<b>4,436</b>	114773	4,191
<b>J</b>	13820	1,957	554	2,689	<b>14374</b>	<b>1,978</b>	48249	1,762
<b>o_u</b>	4682	0,663	62	0,301	<b>4744</b>	<b>0,653</b>	13778	0,503
<b>o</b>	50007	7,082	2107	10,226	<b>52114</b>	<b>7,171</b>	183317	6,694
<b>o:</b>	2785	0,394	12	0,058	<b>2797</b>	<b>0,385</b>	4860	0,177
<b>p</b>	21657	3,067	462	2,242	<b>22119</b>	<b>3,044</b>	80540	2,941
<b>r</b>	21246	3,009	758	3,679	<b>22004</b>	<b>3,028</b>	98463	3,595
<b>Q\</b>	2749	0,389	0	0,000	<b>2749</b>	<b>0,378</b>	12421	0,454
<b>P\</b>	6773	0,959	115	0,558	<b>6888</b>	<b>0,948</b>	27772	1,014
<b>s</b>	33475	4,741	210	1,019	<b>33685</b>	<b>4,635</b>	143130	5,226
<b>S</b>	8826	1,250	113	0,548	<b>8939</b>	<b>1,230</b>	30905	1,128
<b>t</b>	37073	5,251	809	3,926	<b>37882</b>	<b>5,213</b>	201413	7,354
<b>c</b>	5920	0,838	49	0,238	<b>5969</b>	<b>0,821</b>	19927	0,728
<b>u</b>	19854	2,812	538	2,611	<b>20392</b>	<b>2,806</b>	60866	2,222
<b>u:</b>	3476	0,492	126	0,612	<b>3602</b>	<b>0,496</b>	10922	0,399
<b>v</b>	23403	3,315	735	3,567	<b>24138</b>	<b>3,322</b>	109931	4,014

<b>i</b>	46958	6,651	1242	6,028	<b>48200</b>	<b>6,633</b>	171579	6,265
<b>i:</b>	22522	3,190	616	2,990	<b>23138</b>	<b>3,184</b>	92314	3,371
<b>d_z</b>	923	0,131	1020	4,950	<b>1943</b>	<b>0,267</b>	1950	0,071
<b>z</b>	11719	1,660	163	0,791	<b>11882</b>	<b>1,635</b>	45753	1,671
<b>d_Z</b>	1141	0,162	527	2,558	<b>1668</b>	<b>0,230</b>	2294	0,084
<b>Z</b>	6134	0,869	24	0,116	<b>6158</b>	<b>0,847</b>	16278	0,594

**Table 48 – Frequency occurrence analysis of the phonemes for ADULTICS database**

All phones are covered according to the specification within phonetically rich sentences and words. Schwa (“@”) is not mandatory Czech phone, it appears just within spelled items, in sentences only exceptionally, when some spelling is present.

### 10.3 List of foreign words

This section describes foreign words which have appeared in the database. It should be noted that we assume the word as foreign also in the case when it is declined form of originally foreign word. This is typical for names, e.g. “wood, woodová, woodové“, but generally it may appear in any noun of foreign origin.

We do not assume the word as foreign if it is some strange name of a Czech company.

Following foreign words have appeared in the database.

abakus	airbagy	ambassador	astro
abalo	aires	amfiteatre	atto
abba	alain	amores	audry
abbé	alanis	amundsen	au-pair
abby	albania	amway	austin
acana	albex	andrassy	aux
acapulca	albouy	andre	badman
acer	alcatel	angeles	bambino
adams	alcron	annie	band
adamse	aldo	ansi	bank
adeco	alert	ansil	bartuli
adidas	alf	apes	baseball
adobe	alfa	apollo	bass
advertising	alfrost	apple	bašce
agassi	all	aqua	bayer
agatha	allan	aquabar	beach
agnes	allegro	aquapark	bean
aguilar	allen	aquasoft	beatles
aida	ally	aquatis	beaujolais
aidu	alnos	aral	becker
aidy	alstom	armstrong	beckham
aikido	alvin	arromanches	beethoven
airbag	amadeo	artemis	bekaert
airbagů	amadeus	arthur	bell

bellevue
belmondem
belmondo
belveder
benedict
benedictové
benetton
bern
bernt
best
betty
beverly
big
bill
billa
bille
billy
bismarck
björk
black
blade
bleu
bluetooth
boards
bob
bobby
bobcat
bocelli
bocelliho
bodysun
bonaparte
bond
bono
boo
book
borland
bosch
boss
bowie
brandy
bregoviče
brendan
briana
bridget
bristolu
britney
brosman
brosmana

bruce
buenos
bukowsky
bulgakov
bytes
caen
café
cage
cache
callasovou
cambridge
capital
carla
carlu
carmen
carpaccio
carrefour
carrey
carriera
carruso
cash
casina
castle
cavalera
cave
central
cine
cinzano
circa
citroen
clapton
claptona
clarke
clevel
client
clooney
cobain
coca-cola
coca-coly
cocker
cohen
colin
collection
collinse
collinsem
color
colu
comment

common
commonrail
communication
compact
company
compaq
computer
computerpressu
computers
computeru
comsystem
connery
consulting
contactel
conti
continental
cornflake
cornflakes
cornflakesy
cornflejky
corrs
costner
country
cross
cruise
cugidži
cukidži
culligan
curling
curlingu
cyber
czech
dab
dacia
daewoo
dallas
damme
dap
darell
dave
davem
david
day
daytoně
de
deck
deckendorfu
defacto

default
dell
denzel
depeche
design
detroitu
diamond
diary
dictionary
dion
discman
discmana
discovery
displej
double
doug
douglas
doveru
down
download
dream
dresink
drive-in
druckluft
dublinu
dundee
durrel
dustin
dustinem
dylan
dylana
džabana
džajv
džakarty
džamala
džami
džanašája
džanašija
džavik
džegymu
džejára
džirlo
dživas
dživasu
džubgy
džumové
earthlink
eberharter

ecca
ed
eiffelova
eiffelovu
einstein
electronic
electronics
elo
elra
else
elton
eltona
e-mail
e-mailu
e-mailly
end
englewoodu
enlightenment
enyi
equipment
era
eric
erica
erich
ernest
ery
eukaryotem
eukaryotně
eukaryoton
eulerovu
eunet
eurest
euroland
exit
exitem
exitu
exodus
exuperyho
eyes
falcon
famila
family
farber
fasanti
fauvé
federer
feinmanovy
fender

ferrari
fersschmann
feynmannových
fiat
fiction
fiesta
figo
find
fischer
flag
flash
floorball
fondue
ford
fordu
forest
foresta
forst
forsyth
forsytha
forwardovat
frames
france
franciscu
francise
frank
franke
frankovi
franků
freddie
fredericka
free
freuda
freudovsky
freudu
frida
fridy
fronius
fugees
fuji
funai
funes
funesem
funny
gadžové
gadžovskou
gadžů
gahan

galilei
galileo
gandalf
gandhí
gangur
garcia
garden
garpa
garrigua
gavin
georg
george
georgese
gerald
gere
gibson
gibsonem
ginsberg
globe
godzilla
google
goran
gorana
gordon
góro
gott
gotta
gracia
graffová
grand
grandu
grant
green
gregory
grenoblu
gretzky
grosseto
groups
growing
guano
gump
gun
guns
haas
hackman
haikiki
hakkinen
hallmark

hama
handžárům
hannawald
hanson
happy
hard
hardsoft
harleyi
harrison
harry
harryho
harvard
hase
hat
hawking
head
headway
helsinki
helsinek
helsinki
hemenex
hemingway
henriho
hepburnovou
herbalife
heroes
herzog
hewlett
hifi
hills
hitchcock
hobby
hobbycentrum
hobbypark
hoffman
hoffmanem
hogan
höger
hochštafl
holding
holiday
holzmana
holzner
holznerová
home
homelesákům
hood
hortis
hostel

hot
hotmail
how
huga
hugh
huntera
hypo
charles
charlie
charlieho
chenyuanpeng
chiaping
chicaga
chicago
chicagu
chock
chrise
christie
christopher
chrysler
chuck
churchill
iconics
ideas
idžů
iljič
imaxu
imperial
impo
in
inbox
inclusive
info
infraport
intel
interbed
interconex
intercontinental
interhotelu
interview
invest
iron
isac
ishikhavy
isoleucin
istanbulu
istebné
ivrit

jack
jackson
jacksonville
jag
jake
jam
james
janeira
janeiro
janis
java
jazz
jean
jean-claude
jean-paul
jeep
jeneweinově
jenkins
jensene
jeromem
jerry
jerryho
jerrym
jigoro
jim
jimem
joe
johannes
john
johna
johns
johnson
johns's
joker
jolly
jon
jonathan
jones
joplin
jordan
journal
jovi
joy
juda
judo
jules
julia
julien

kahna
kaiser
kaiserová
kamikadze
karabach
karadžič
karadžiče
karen
kasparov
katya
kaufland
kauflandu
keanu
keazin
ken
kenneth
kevin
kibucu
kick-box
kidmanová
kimi
kimono
kindžál
kindžálem
king
kinga
kiplinger
kirken
kiss
kivatos
kobain
kodak
königu
kraken
kraši
krawits
krišna
kung-pao
lancelot
landia
langhamer
laserovou
last
leara
leasing
lennox
lenny
lennyho

lenz
leonard
leonardo
lesbiens
less
leucinech
leucitech
leucitová
leucitové
leucitový
leukocytu
leukocyty
leukózový
levkada
lewise
lidl
like
lincoln
link
linkin
linking
linus
linux
literal
livigera
livingston
login
loife
long
louis
love
low
lowdown
luis
lumbarda
lynch
macdonaldové
mackennovo
maclarenu
macquirem
madeiru
madonna
madrid
madridu
maestra
maestro
maharádža
mahárádža

maharádži
mahárádži
maharádžo
mahárádžo
maharádžu
mahárádžu
mahárádžů
mahler
mahlerovi
maiden
mail
mailbox
mailboxu
mailová
mailové
mailovou
mailu
mailů
maily
maiser
majakovského
mallorcu
mandžuská
mandžuské
mandžuský
manchester
manikotrada
manowar
manson
mansona
manuela
maradona
marcos
maree
maria
marianne
marilyn
marketing
marley
marques
marseille
mash
masters
matejčuk
materialismus
matisse
matrix
matrixu

matters
mattonku
maurice
mauritus
may
mayer
mayerer
mayů
mcbain
mcbeallová
mcdonald
mcdonalda
mccray
mederu
meetingu
meinl
meintscher
mel
men
mercedes
mercury
mercy
merilyna
merlin
meryl
metallica
metallicu
metallicy
mettenu
mexicana
miami
microcom
microoil
microsoft
michael
michaela
michiganu
mika
mike
miller
minár
minute
moby
mobyho
mode
mojbacher
money
monitor

monitorem
monitoru
monitorů
monitory
monte
montmartru
montoya
montrealu
moore
more
morrisette
morse
motorola
moulin
movie
mudžáhidy
murat
murphyho
musaku
nabucco
nadžáfi
nagana
naganu
nagye
nancy
nandžunský
nasa
natalija
natura
naumov
navair
navy
neil
neocortex
nero
nestlé
neuhausenu
neumann
neumanna
neumannová
neumannovou
neuromancer
něvjem
new
newman
news
newspday
newton

nick
nico
nicol
nicolas
nicolson
nikola
nikósii
niro
nirvana
nirvanou
nirvany
nissan
nissany
niveco
no
nokie
nomurou
noname
normana
norris
notebook
notebooku
nothing
notre-dame
of
oldfield
oldfielda
oldies
oleg
olivier
omicron
ondrej
online
opel
open
opla
or
orlando
orlandu
orwel
orwella
osbourne
oscar
ostende
othello
over
overture
ozzy



packard
paegas
paegasa
page
pager
panasonic
papette
parade
paradise
pearl
peckem
perros
petangue
pete
peter
peugeot
peugeota
pfeffersteak
pharlap
phil
phila
philem
philips
physics
pierce
pierre
piguins
pink
pitt
pizza
pizzeria
pizzeria
pizzerii
pizzu
play
player
playstation
poe
popularis
porsche
potter
pottera
potterovi
power
powers
pratchett
preiss
pretty

prix
psaroudakis
pulp
quanto
quarter
queen
queenovou
queenu
queenyy
quelliova
quelliovy
quick
raikkonen
rail
rainman
rallye
ready
real
receiver
receiverem
recorder
recycling
red
redford
redhat
reeves
reich
relay
reloaded
remarque
renault
republic
research
reuters
rex
ride
rimbaux
robbie
robbieho
roberts
rockwool
rocky
roche
rochesteru
rolling
rooseveltova
rooseveltové
roses

rossi
roswell
rouaulta
rouge
roxette
royal
rulanta
rullack
runner
rush
russolia
ryana
saint
saintien
sampras
san
sandonorico
sanger
sauce
scala
scarie
science
scientific
sci-fi
scooter
score
scorpions
scorpionsû
screamers
scripts
sea
seagal
sean
seattlu
securitas
serene
services
shakespear
shakespeare
shakespeareovské
share
sharingham
shawn
shawshank
sheen
shiller
shizuky
shop

show
shrek
schafer
schalt
schang
schürman
schwarzenegger
schwarzeneggere m
sieben
siemens
simmel
simon
simpsona
simpsoni
simpsonova
simpsonovi
simpsonovy
simpsons
simpsony
singer
single
sinuhet
sitcom
site
smells
smith
smitheho
snipes
snowboarding
so
soft
sollers
solutions
spears
spielberga
spirit
squash
stallone
steak
steelová
steelový
steffi
stephen
stephena
steve
steven
stevena

stevenem
steward
sting
stinga
stones
störig
strasse
streep
streepová
stuttgart
success
sunil
superbikú
support
surround
survey
sven
sydney
tai-chi
talk-show
tampa
tapedeck
tauben
team
teen
telecom
telecomu
teletubbies
terry
terryho
the
theory
thomson

thriller
tiffanyho
tiscali
titanic
today
tolkien
tolkiena
tolkienovu
torpa
torres
tortilli
toshiba
tour
tourist
trade
trading
trainspotting
translations
transpotting
travellers'
travesti
tropez
trutnigholmu
tugendhat
two
umprofor
underground
up
valleau
van
vancouver
vangelis
verne

vertical
victora
viewegh
viewegha
villachu
ville
virgin
volksbank
volkswagen
vonegut
voyager
vurwork
wagner
waldemar
walkman
walkmana
walkmanů
wall
warriors
was
washington
washingtonu
waterloo
wayne
web
wellington
wesley
west
who
wilbury
wilde
willbour
william

williams
williamse
williamsová
williamsovy
willis
willise
willisem
windsurfing
winfred
winky
winston
winstona
winstone
winter
winterová
winterové
wolf
woman
won
woodová
woodové
works
worků
world
yahoo
yacht
york
yorku
zakki
zappa
zeppelin

## 10.4 Lexicon over-completeness

The lexicon is generated for all recorded session, included the additional ones. For main block of regular 550 sessions, we may obtain over-completed lexicon. Also mispronounced words are included in the lexicon, because they were generated on the basis of LBO an EPI field in SAM label files.

## 11. Double checks

The possible mistakes in transcription were detected by semi-automatic tests, i.e. running of scripts based on analysis of spelling and corresponding pronunciation, followed by manual

checking of not-passed words. The error rates during the first overview of annotation package were typically within the range 0.5 – 1%. Found errors were always corrected.

Additional 2 % randomly selected files were double checked for the accuracy of orthographic transcription and for accuracy of usage of non-speech event marks, mispronunciation and truncation marks. Just minor errors of annotation were detected during this step. The correction of these errors (incompatibilities) should lead to more coherent results of annotation among all participating annotators.

## 12. Other information

### 12.1 Impulse response measurement

For each new environment and position the impulse responses were measured, however, due to the recording efficiency, it was sometimes done after the speaker recording. At several places and recording positions we may find two measurement of impulse response. It is due to long pause between the recording at this position the first time (for mini-database) and the second time (for complete DB collection).

The rule to find proper impulse response is very simple. It is that measurement which is present in the nearest session recorded at same place and in same position.

### 12.2 Description of additional sessions 550-589 joined to the final database

40 additional sessions are joined to the database. Descriptions of these sessions, with the reason of these additional recordings are summarized in the following part. All those sessions are fully annotated and underwent the annotation checking process.

Sessions from Table 1 contain generally completely additional recording with new speakers. Sessions 550-559 were recorded as compensation for low DBA in 30 sessions recorded at too quiet public places. The sessions 560-567 were discarded from basic 550 set due other errors (over-recording per place, repeated recording of one prompt sheet). These errors were corrected (sheets were re-recorded) but already performed recordings had good signal quality so we offer them also as another compensation for found problems.

<i>Session No.</i>	<i>Sheet No.</i>	<i>Speaker No.</i>	<i>Note</i>
550	000	800	<b>Compensation of low DBA in PUBLIC environment.</b>
551	001	801	Recorded in PUBHALL environment.
552	002	802	Recorded with new speakers.
553	003	803	
554	004	804	
555	400	700	<b>Compensation of low DBA in PUBLIC environment.</b>
556	401	701	Recorded in PUBOPEN environment.
557	402	702	Recorded with new speakers.
558	403	703	
559	404	704	

560	464	464	<b>Discarded sessions due to over-recording per one place and environment.</b>
561	465	465	
562	466	466	
563	467	467	
564	472	472	Used prompt sheets were re-recorded with new speakers in new environment. Speakers: 705/733, 706/734, 707, 708, 709/717
565	353	354	<b>Discarded sessions due to repeated recording of same prompt sheet.</b>
566	438	439	
567	399	421	Missed prompt sheets (354, 439, 399) were recorded with new speakers (711, 712, 710) in new environment.

**Table 49 – Overview of additional sessions with good signal quality**

The sessions summarized in Table 50 were discarded and re-recorded due to bad signal quality (microphone problems and too weak speech especially in car recordings). Because all these sessions were annotated, we include them as additional.

568	403	403	<b>Discarded sessions due to very bad signal quality.</b> Sheets were re-recorded with new speaker (713, 731) in new environment.
569	467	708	
570	375	375	<b>Discarded sessions due to bad signal quality in CH0 (very low SNQ) – problems with microphone, high noise level, and too quiet speakers.</b> All these sessions are in CAR environment. Except 2 discarded sessions No. 571, 584 (it was re-recorded as 377, 527), sheets were re-recorded with same speakers in CAR environment, but slightly different driving conditions. Signals in CH1, CH2, CH3 should be OK.
571	377	377	
572	378	378	
573	380	380	
574	496	496	
575	498	498	
576	510	510	
577	511	511	
578	512	512	
579	513	513	
580	515	515	
581	517	517	
582	518	518	
583	526	526	
584	527	527	
585	528	528	
586	539	539	
587	464	705	
588	465	706	
589	472	709	

**Table 50 – Overview of additional sessions with some problems in signal quality**

### 12.3 Comments to prompt & lexicon and mini-database pre-validations

The errors specified in VALREP.DOC were eliminated.

As the consequence of the pre-validation remarks, number of digits in CB items was increased and too long utterance were splitted into items CB1 and CB2. Missing description of splitting CB-digit sequence into two items CB1 and CB2 was added in section 5.3.4.

### Explanations to the remarks within prompt lexicon pre-validation:

- španielova S p a J i j e l o v a
  - *As it is personal name it is quite hard to determine what is the right pronunciation, people often tend to "twist" such a name somehow. However the above said pronunciation is possible although highly probably rare.*
  - This is a street name, in final database it appears two times with pronunciations: "S p a J i j e l o v a" and "S p a n j e l o v a".
- surround s u r r a \_ u n t
  - *Not Czech word. Very hard to say how would people who don't know English read it. Most probably (according to e.g. TV news custom) the reading would be s e r a \_ u n t but also s @ r a \_ u n t and even s a r a \_ u n t and s i r a \_ u n t are possible I guess.*
  - This is really not Czech word but it is required application word. No suitable and frequently used Czech word is used for this audio function. Pronunciation may be really different. Most frequently used reasonable pronunciations are in the lexicon.
- lankaš l a N k a S
  - *Haven't ever seen such word. If it is a personal name then it is correctly transcribed otherwise it is nonsense.*
  - It is a proper name.
- třešť t P \ e S c
  - *Strange word. Maybe in some poetry expression or in local dialects but not a common Czech word. Anyway there cannot be P\ after t. It must always be unvoiced Q\: t Q \ e S c.*
  - This is a city in Moravia. Correct pronunciation is really "t Q \ e S c"; P\ was present due to formal mistake in lexicon generation.
- kormutlivý k o r m u t l i v i:
  - *Terribly archaic expression. No one would say that. What was the source for your corpus? ;)*
  - It is maybe really not so frequent expression. It was used in phonetic rich material collected from newspaper text and several classical novels.
- dr d r
  - *Not a Czech word. Probably short for doctor but nobody would really say [d r] in spoken utterance.*
  - It is really short for doctor. Was present just in prompts, nobody pronounced it this way.
- džamala d \_ Z a m a l a
  - *Have never heard. What is it? In Czech it can only be a personal name (of somebody from Africa :) I guess).*
  - Proper foreign name recorded in sentences originated from newspaper texts or novels. Some of these words had to be added to have good coverage of rare phone d \_ Z
- jeromem Z e r o m e m
  - *Also d \_ Z e r o m e m is possible and probably more commonly used.*
  - Finally "d \_ Z e r o m e m" and "j e r o m e m" were used.
- sv s v
  - *Exists only in written language (short for saint). Cannot be pronounced.*
  - One speaker pronounced it this way.
- džmurou d \_ Z m u r o \_ u
  - *Strange word which I don't know.*
  - Surname of Czech sportsman.
- amfiteatre a F f i t e a t r e
  - *Not Czech word: Czech variant is "amfiteátr a F i t e a: t r".*
  - It was prompted as foreign word, one speaker pronounced it really this way.

Other general notes to the lexicon and used vocabulary:

By providing orthographic transcription with pronunciation, we had to solve several problems described in section 9.1. Consequently, some strange words may appear in lexicon, typically foreign words, slang explanations, proper names, street names, strange company names, words from internet addresses, etc. It must be taken into account that the lexicon contains pronunciations of recorded words, it is not the lexicon of written Czech vocabulary with word pronunciations.

## 12.4 Comments to the errors found during final database validation

> *Too little spontaneous speech was found for sessions 242 (93 seconds), 507 (106 s) and 568 (87 s). These can be compensated for by the additional 40 sessions.*

OK. We also hope that this error is minor and well compensated by additional sessions.

> *CBI-2: For the digit form 'osm' only 398 repetitions were found (min is 467). This deficiency is recognized in DESIGN.DOC.*  
> *CC1-4: For the following digit forms the number of repetitions is < min 584: ctyry (569), sedm (543) and osm (488). This deficiency is recognized in DESIGN.DOC.*

OK. We found it already during completion and explanation is given on the page 30 of this document.

> *CE1: 2 numbers are 14-digit long.*

It should be the following numbers, i.e. international codes.

00497131673636  
00493613464280

We collected these numbers from Internet to obtain some realistic numbers. Unfortunately, the check for length was performed when those numbers were collected. The error appeared after replacement of the universal international dial-character "+" by Czech version, i.e. "00".

We hope that when it appears just in two items it is not critical problem.

> *CW2: 548 instead of 550 email addresses were found.*

This is given by discarding of very bad recording during the annotation process. Since each speaker should have uttered unique e-mail address, finally two of them are missing.

> 24 instead of 20 different keyboard characters were found.

This error is given by small misunderstanding. We have really worked with just 20 different keyboard characters but in Czech, 4 of them have two different (but widely used) names. It yielded this final number of 24 different words for keyboard characters.

> In 30% of the files in the CAR environment SNR < 10 dB (max 20%). Extra sessions have been added to compensate for this error.

Yes. This problem should be really compensated by additional recordings. We have found big problem with close-talk microphone in CH0 in several recorded car sessions and these sessions were re-recorded. Bad sessions were kept in the DB (recordings in CH1-3 are OK). SNR for car sessions bellow 550 should be OK (86% above 10 DB according to our statistics).

> 82% of the sessions are in the required noise range for public place (min > 90%). Extra sessions have been added to compensate for this error.

This is a similar problem. On contrary to the previous case, all realized recordings have good signal quality here but approx. 30 sessions were recorded in too quiet environment. After discussion with the members of the consortium, it was compensated by 10 additional recordings. Still, the specifications are not fully met both for the 550 speakers and for the whole database.

> In CARMIDDLE, ENGINE\_OFF 10 sessions were found (max is 9).  
> In CARMIDDLE, CITY\_30\_70 15 sessions were found (max is 9) and 5 come from  
> the additional sessions, but there is still one too many.  
> In CARUPPER, COUNTRY\_60\_100 12 sessions were found (max is 9) and 2 come  
> from the additional sessions, but there is still one too many.

This error appeared during the re-recording of bad car sessions - we missed maximal session numbers in these three environment. We hope that even if we do not slightly comply with the specification, it will be acceptable. Moreover, the driving conditions are not critically unbalanced according to our point of view (min. 5 speakers is at each driving condition and summing CAR\_MIDDLE and CAR\_UPPER in each driving conditions were recorded 15-11-17-18-12 speakers).

> There is a discrepancy between the spelling in DESIGN.DOC and in the database:  
> výstavište vs. výstaviště

It was caused by typographic error in DESIGN.DOC and it was corrected.

> There are some inconsistencies in the spelling at the transcription level:  
> audio video vs. audio-video  
> rádio vs. radio

“audio-video” is the correct spelling however “audio video” gives same pronunciation, and may be principally also used. We tried to unify the way of spelling during the transcription, but in this case it was missed.

“radio” is bad spelling, it is typographic error which appears just in 4 cases, while in 756 cases correct spelling “radio” is used.

*> SNQ is 0 in at least one of the channel in 21 files*

It seems to be the consequence of estimation algorithm failure. The mentioned cases are generally signals with relatively high level of the background noise.

## **12.5 Modification after validation**

Any modification on signal and annotation files has not been done after validation.

Concerning the database contents following, just changes has been done:

- Correction of badly formatted SUMMAR0.TXT file, bad tabulators were replaced by spaces.
- README.TXT will be placed only on the documentation disk and bad reference to CDs was replaced by the reference to DVDs in this file.
- In COPYRIGHT.TXT file were also corrected bad reference to CD-ROM.
- Corrections in this documentation file:
  - correction of mistakes mentions in section 1 of validation report (VALREP.DOC).
  - these new two subsections 12.4 and 12.5.