CZECH LOMBARD SPEECH DATABASE

CLSD '05

Author(s):	Hynek Bořil, Petr Pollák, Czech Technical University in Prague
Institute:	Dept. of Circuit Theory
Address:	CTU in Prague, FEE K13131 Technická 2 16627 Praha 6 – Dejvice Czech Republic
Email:	borilh@gmail.com pollak@feld.cvut.cz
Date:	22nd Feb 2006
Database owner:	CTU in Prague, FEE K13131 Technická 2 16627 Praha 6 – Dejvice Czech Republic Contact persons: Hynek Bořil, Petr Pollák
Version:	1.4 – 24nd Feb 2006 –Hynek Bořil & Petr Pollák Updated version

1. Introduction

The CLSD05 dabase (Czech Lombard Speech Database 2005) focuses on analysis and modeling of Lombard effect. The database consists of neutral speech and speech produced in various types of simulated noisy background. Recently, 26 speakers participated in the recording.

The CLSD05 design and acquisition are carried out by the Speech Processing Lab at Czech Technical University in Prague, contact persons Hynek Bořil (borilh@gmail.com) and Petr Pollák (pollak@fel.cvut.cz). The CLSD05 is fully owned by Czech Technical University in Prague.

2. Distribution media

2.1 Storage media

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

<database>_<oo>

where <database> is "CLSD05" and <oo> is a two digit code. Currently, there is one DVD named

CLSD05_01

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.
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 label files is

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

where "/" is a generic file system separator symbol. Session numbers are generated automatically by the recording application. The following directories are used for Czech database.

<database>	CLSD05, 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

NOTE! Several neutral speech sessions were used from the Czech SPEECON database. These sessions are stored in the directory ADULT1CS with the same structure as in the CSLD05.

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: “LN“ = Lombard speech sessions, “LO“ = Neutral sessions, “SA“ = SPEECON sessions,
<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“: speech signal files for channels 0 - 1

Table 3 – Signal and label files names

NOTE! Just first two channels were copied from SPEECON database. Also several item categories were not recorded within CLSD05 and therefore these items were not included in this collection.

2.2.4 Corpus Codes

According to the corpus description in D213, the following corpus codes have been defined for CLSD05:

Corpus id.	Item id.	Corpus contents
Read speech		
S	01 – 30	30 phonetically rich sentences
W	01 – 05	5 phonetically rich words
Core words (read), general words and phrases, applic. specific words and phrases		
C C	I1 – I4 30 – 69	44 isolated digits
C C	B1 – B2 00 – 29	32 isolated digit sequences (8 dig.)
C C	C1 – C4 C70 – C99	34 connected dig. seq. (5 dig.)
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	P1	1 proper name
C	O1 – O2	2 city or street names
C	Q1 – Q2	2 questions
C	K1 – K2	2 special keyboard characters
Y	01 – 95	core word synonyms
1	01 – 85	Basic IVR commands
2	01 – 40	Directory navigation
3	01 – 22	Editing
4	01 – 57	Output control
5	01 – 70	Messaging & Internet browsing
6	01 – 33	Organizer functions
7	01 – 39	Routing
8	01 – 12	Automotive
9	01 – 95	Audio & Video

Table 4 – Corpus codes

2.3 Documentation directories and files

2.3.1 Documentation directories

The documentation will be held in a file system in one documentation directory.

/CLSD05/DOC	Documentation

Table 5 – Documentation directory

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	

TXT		ISO 8859-1 DOS-formatted text file
	FILESLO.TXT	
	FILESLN.TXT	
	PAIRLIST.TXT	
	SESLO.TXT	
	SESLN.TXT	
TBL		ISO 8859-2 DOS-formatted text file
	CLSD05LX.TBL	
PDF		Adobe Portable Document Format
	DESIGN.PDF	
	ISO88592.PDF	
	SAMPALEX.PDF	
PS		Adobe PostScript format
	ISO88592.PS	
	SAMPALEX.PS	

Table 6 – Content of the DOC directory

DESIGN.DOC | DESIGN.PDF
This file contains information about design, collection, annotations, and completion of Czech Lombard Speech database. (**This document**)

ISO88592.PS | ISO88592.PDF
Table of Czech characters used in the database.

SAMPALEX.PS | SAMPALEX.PDF
Table of Czech phonemes in the SAMPA format.

FILESLO.TXT | FILESLN.TXT
Contain lists of all included files with Neutral speech (FILESLO.TXT) and with Lombard speech (FILESLN.TXT).

SESLO.TXT | SESLN.TXT
Files containing lists of all included Neutral sessions (SESLO.TXT) and Lombard sessions (SESLN.TXT).

PAIRLIST.TXT
Files containing lists of all Lombard and Neutral speech item pairs. Some Lombard speech items do not have a neutral pair (and vice versa) in cases where SPEECON sessions were used to cover the neutral speech (as CLSD05 session prompts do not contain all SPEECON item types on one side and are extended for digits on the other side).

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 has a separate entry. As the lexicon is derived from the database it uses the same alphabetic encoding for special and accented characters as used in the transcriptions. The CLSD05 lexicon file consists of three mandatory fields:

- Orthography (related to conventions used for annotations, as described in the Section 8),

- Frequency of the occurrence count of a given orthographic form,
- SAMPA pronunciation (Czech SAMPA is described in the following part of this document and at the web page `http://noel.feld.cvut.cz/sampa` or at offical SAMPA web `http://www.phon.ucl.ac.uk/home/sampa/home.htm`),
- Additional optional tab-delimited pronunciation variants fields.

The first line contains 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. 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.

The representation of the SAMPA pronunciations is same as it was used in lexicons for Czech SPEECON and SpeechDat databases.

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. Both speech channels are recorded at 16 kHz/16 b 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 also an additional EPI field which contains phonetic transcription of utterance in SAMPA, i.e. transcription containing pronunciation variant used in the utterance. In the case of mispronunciation, "*" is used in the same way as in the LBO field.

There is one SAM label file assigned to each utterance, i.e. one for both 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	CLSD05	
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>\BLOCK<NN>\SES<NN><M>\	%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	Integer: [16]	%d

	per sample		
QNT	Quantization	Fixed vocabulary item, e.g.: PCM	%s
NCH	Num.of channels	Integer: 2	%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 attribue-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 unknwon, the values may be left empty. However, the correct number of commas must remain.	%s,%s,%s,%s, %s,%s
MIP	Microphone positions	An attribue-value pair list, CHN0= %s, CHN1= %s, it is described in follwoing part of this doument where the recording conditions are desribed.	%s,%s
MIT	Micophone types	An attribue-value pair list, CHN0= %s, CHN1= %s, it is described in follwoing part of this doument where the recording conditions are desribed.	%s,%s
DBA	Noise level	The noise level, as measured by the noise-level meter during recording of the silence word – EMPTY for CLSD05	%f
SNQ	Signal/Noise Quality	Attribute value pair list, CHN0 = %f, CHN1 = %f The SNR values estimated by the recording platform	%f
LBO	Orthographic transcription	<transcription text>	%d, %d, %d, %s
EXP	Labelling expert	Name Surname, Organization	%s
SYS	Labelling system	Software description	%s
DAT	Date of completion of labelling	DD/Mon/YYYY	%s

Table 7 – SAM format – labels used in CLSD05

EPI	Phonetic transcription	Phonetic transcription of the utterance pronunciation (in SAMPA)	%s
NTY	Noise type	%s	Filenames – including noise description code
MNL	Noise level	%f	The noise level – set by measured level from soundcard output
MSL	N/A	N/A	Reserved for future use
DES	Speaker-OperatorDistance	%f	Distance (m) – info for appropriate speech signal attenuation in operator's recording monitor

Table 8 – Additional SAM format labels in Czech Lombard speech database

An example label file is given below.

```
LHD: SAM 6.1
DBN: CLSD05
SES: SES025
CMT: *** Speech Label Information
SRC: LN025302.CS0,LN025302.CS1
DIR: CLSD05\BLOCK02\SES025
CCD: 302
BEG: 0
END: 31232
REP: office_460
RED: 06/Apr/2005
RET: 13:47:20
CMT: *** Speech Data Coding ***
SAM: 16000
SNB: 2 signed
SBF: lohi
SSB: 16
QNT: PCM
NCH: 2
CMT: *** Speaker Information ***
SCD: 025
SEX: M
AGE: 25
ACC: CWNB
CMT: *** Recording Conditions ***
SNQ: CHN0=, CHN1=,
MIP: CHN0=CLOSE_HEADSET,CHN1=CLOSE_LAVALIER
MIT: CHN0=SENNHEISER_ME104,CHN1=NOKIA
SCC: ENV=OFFICE,PLC=office_460,POS=CLOSE_WALL_01,SIZ=SQM_20_30,AUD=,DRV=
DBA:
CMT: *** Labelling information ***
SYS: FTP Transcriber 3.07.1
EXP: Petr Jonas
DAT: 06/Nov/2005, 20:34:33
ORT: poloľka
EPI: poloSka
CMT: *** Label File Body ***
LBD:
LBR: 0,31232,,,, položka
```

```
LBO: 0,15616,31232,položka
CMT: *** Lombard Speech - Noise Information ***
NTY: an_840_2500Hz.wav
MNL: 90
MSL:
DES: 1
ELF:
```

Table 9 – Sample of a label-file

4. Prompting and recording procedure

4.1 Prompts

All recording sessions were prompted the same way using the recording studio developed for CLSD05 collection.

4.2 Recording procedure

First, the speaker is interviewed to ensure the correct speaker profile (suitable age and gender). Then the speaker is instructed about the recordings: to read the prompts when the red light of the recording application appears and to repeat the word/sentence if he or she has mispronounced it. Details of the recording setup are desribed in section 6.

5. Database design

The CLSD05 database consists of 26 adult speakers. The corpus and item identifiers are specified in the section 2.2.4. Items recorded in the database are described in detail in the following sections.

5.1 Read speech

Total number of items in the prompting material in CLSD05 is 205 per session (with an exception of pilot sessions 000 – 002). An extensive set of digits was included into each CLSD05 session to allow statistically significant analyses for relatively small amount of speakers present in the database.

5.1.1 Phonetically rich sentences (S01–30)

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. This number of sentences results from the first pre-filtering removing very long sentences, orthographically or grammatically strange sentences, etc. Further selection was made by CorpusCrt, which can be 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.

5.1.2 Phonetically rich words (W01–05)

The phonetically rich words were added to the corpus to introduce phonemes that are not well covered 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.

These words were chosen from a original corpus of 2798072 words from lexicon used in Linux *ispell* for Czech. The selection was made again by *CorpusCrt*. All words are 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.

5.1.3 Isolated digits (CI1–4, C30–69)

The frequency of occurrence of all tokens is (approximately) uniform. Language specific peculiarities are specified in following Table 10. The digits were presented to the speakers as words.

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 10 – List of basic digit variants

5.1.4 Isolated digit sequences (CB1–2, C00–29)

The isolated digit sequences provide examples of 8 digits in various orders. The sequence of the digits is random. The digits are presented to speakers as words.

5.1.5 Digit strings (CC1–4, C70–99)

Strings of 5 digits to be read continuously are represented as words. The digit odrers are generated randomly.

5.1.6 Natural numbers (CN1-3)

The prompts present numbers X ≤ 10,000,000.

5.1.7 Money amounts (CM1)

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

The items contain small (i.e. including decimal currency units) and larger money amounts (not including decimal currency units). ‘Euro’ and ‘Cent’ are used as a foreign currency.

5.1.8 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, see the list bellow (English equivalents) and Table 11.

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 11 – List of most common Czech analogue time expressions

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

The times were prompted as words.

Example: "`pět hodin dvacet sedum minut`".

5.1.9 Dates (CD1-3)

D1: analogue form.

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

```
čtvrtek, dvacátého devátého února, dva tisíce patnáct
```

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

Table 12 – 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 13 – List of Czech month names

Note: Month names are used in declined form (2nd case of singular). These forms are summarized in the third clumn of Table 13 (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 1414.

Dnes	today (+0d from now)
Zítra	tomorrow (+1d from now)
Pozítří	the day after tomorrow (+2d from now)
Včera	yesterday (-1d from now)
Předevčírem	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 (24th of December)
První_svátek_vánoční | Boží_hod_vánoční	(First) Christmas_Day | (25th 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_dětí	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 14 – List of general and relative date expressions

D3: digital form.

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

An example for the prompt:

```
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 15 – List of names for Czech month names in digital form

5.1.10 Proper Name (CP1)

The proper names were chosen from a set of **150** first and last names or a combination of them.

5.1.11 Cities (CO1)/street names (CO2)

These items contain utterances from the list of 275 most frequent cities and 275 most frequent street names.

5.1.12 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 16 – Yes/no expressions

5.1.13 Special keyboard characters (CK1-2)

The names of the special keyboard characters were prompted to speakers as orthographic words. Bold text is used for mandatory keyboard characters.

zavináč	**'at' sign ('@')**
křížek	**hash ('#')**
zpětné_lomítko	**backslash ('\')**
mezera	**Blank | space (' ')**
dvojtečka	**Colon (':')**
tečka	**dot | period ('.')**
pomlčka | mínus	**hyphen ('-')**
lomítko	**slash ('/')**
hvězdička	**star | asterisk ('*')**
plus	plus | plus sign ('+')
podtržítko	**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 17 – List of keyboards characters

5.1.14 Application specific commands

Recorded application words originate from the SPEECON list. It was not possible to record all application commands within several recording sessions, so we do not present here the complete list. Complete list is available within Czech SPEECON documentation [1].

6. Recording environment, scenarios and setup

6.1 Environment

Room size

- Approx. 7 x 7 x 3.5 meters

Noise level

- LAeq = 30 – 60dBA

Recording conditions

- Operator and speaker were sitting in front of the monitors placed next to the abacus
- Most of the computers were running (fan noise), some of the windows were open
- Room is situated in the 4th floor, windows lead into a relatively calm exterior
- Door is closed
- Except the recorded person, nobody was allowed to talk during the recording

Figure 1 – The room from the recording place perspective

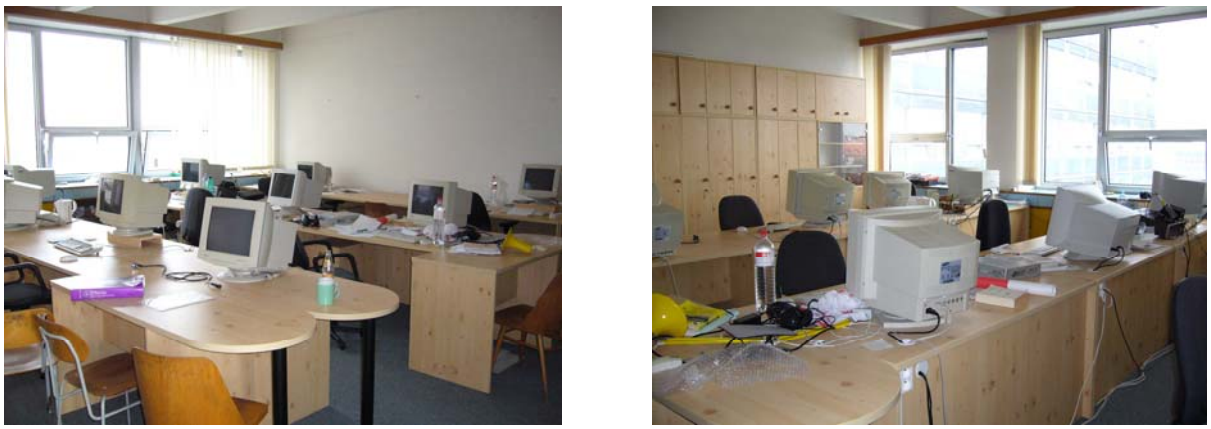

Figure 2 – The room from the door and diagonal to the recording place perspective

6.2 Scenarios

No discussions or meetings were allowed in the office during the recordings.

6.2.1 Neutral speech

Neutral speech recording conditions were equivalent to the SPEECON Office scenario recordings. Speakers read prompts from the display, speech was recorded by two microphones placed in the different distances from the mouth, see details in section 6.3. In this case speakers did not wear headphones.

6.2.2 Lombard speech

In the Lombard speech scenario noisy background was reproduced to the speaker through headphones, hence high SNR of the recorded speech was preserved. An operator qualified intelligibility of the utterances while listening to noise of the same level mixed with the utterance of intensity lowered in proportion to selected virtual speaker-listener distance. This setup motivated speakers to react more to the noise background. See details in section 6.3.

6.3 Recording platform

The database was recorded digitally into hard disc. In case of the noisy conditions scenario, speaker heard his own voice mixed with noise in closed headphones. The level of the speech feedback was adjusted individually to make speaker feel comfortable.

6.3.1 Hardware setup

Recording set, see Figures 3, 4, 5, consists of 2 closed headphones AKG K44 and 2 SPEECON microphones – close talk Sennheiser ME-104 and hands-free Nokia NB2, placed in different distances from the speaker's mouth.

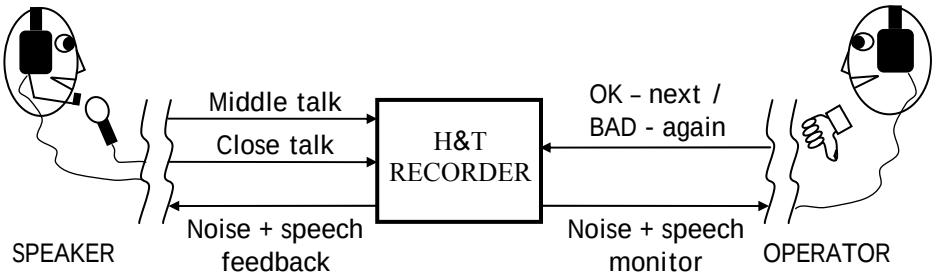

Figure 3 – Recording setup

Figure 4 – Sennheiser ME-104 and Nokia NB2 microphones, detail

Figure 5 – Microphone placement - side and front view

6.3.2 Noise level adjustment

To enable noise level adjustment without need of measuring SPL in the beginning of each recording session, transfer function between sound card open circuit effective voltage V_{RMS_OL} and SPL in headphones was determined. The measurement was performed on a dummy head, see Figure 6, 7. For the chosen noise level, corresponding V_{RMS_OL} was set up at the beginning of each session recording. The transfer function is shown in Figure 8.

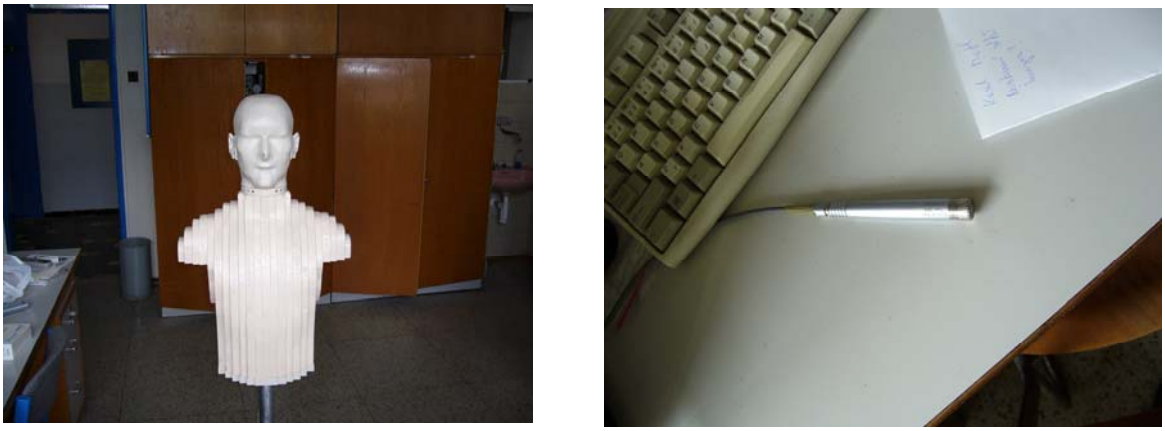

Figure 6 – Dummy head and similar microphone to the dummy head ones

Figure 7 – Measuring amplifier BK 2525

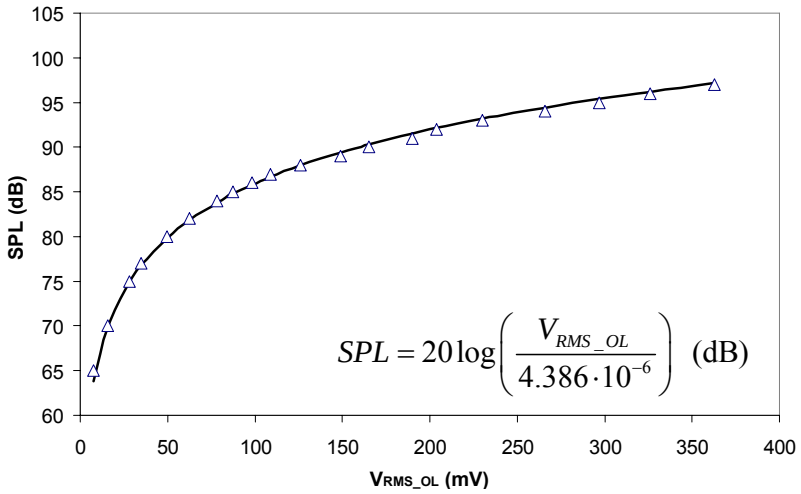

Figure 8 – V_{RMS_OL} – noise SPL dependency

An average of 90 dB SPL and 1–3 meters of virtual distance were chosen as default for Lombard speech recording scenarios. In some cases the settings had to be modified according to particular speaker's capabilities.

6.3.3 Noise backgrounds

For Lombard speech recording, background noises were selected for observations of speech production changes both for natural noisy environment and for artificial band-noises interfering with typical locations of f_0 and first formants occurrence. 25 noises recorded in car environment from CAR2E database and 4 band-pass noises (62-125, 75-300, 220-1120, 840-2500 (Hz)) were chosen. Each car noise sample was about 14 sec long, stationary band-noises were 5 sec long. The noise sample was looped in case the utterance was to exceed the sample length. All noises were RMS normalized to provide corresponding sound pressure level (SPL) during the reproduction.

6.3.4 Recording studio

H&T recorder used for the CLSD05 collection was implemented as a .NET application.

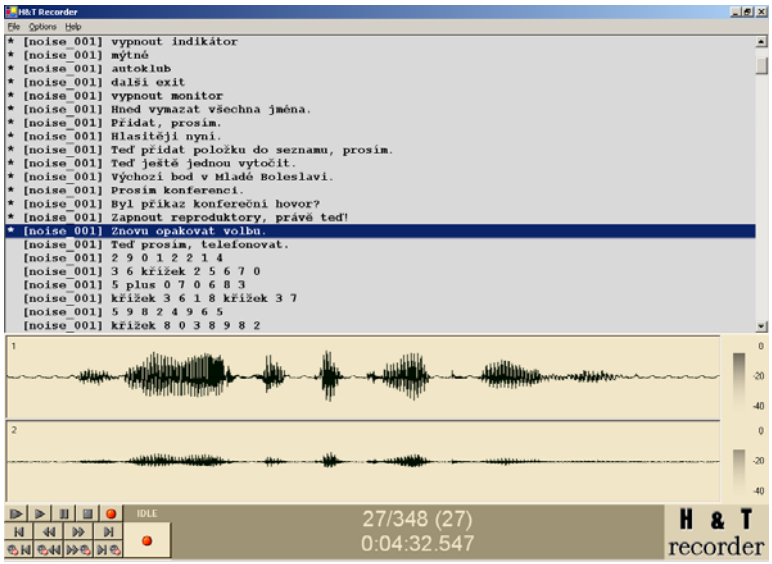

Figure 9 – H&T recorder window

The recorder supports two-channel recording and separate noise/speech monitoring for speaker and operator respecting virtual distance. To each utterance an item from the noise list is assigned during the recording.

7. Speaker distribution

The information about speaker gender, age, and dialect was saved but concenring the speaker coverage just gender was take into account for the proportional coverage. 14 male and 12 female speakers participated in the recording of the database.

The dialect regions were defined for previous database collections (Czech SpeechDat and SPEECON databases). Used regions are defined in the following table.

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

Table 18 – Dialect regions in Czech Republic

8. Annotations

8.1 General description of annotations

CLSD05 was annotated orthographically, the real pronunciation was checked for all utterances. The exceptions from the expected pronunciation were marked and finally used for the generation of the phonetic lexicon. The pronunciation of each utterance is kept in a label-file as an EPI field.

FTP-Transcriber written by Petr Schwarz from VUT Brno was used for the annotation of the database.

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, see details in section 2.3.2.

CLSD05 transcriptions are not case sensitive (except for words like “dlouhý” (long), “přehlasované” (umlaut), etc.). Punctuation marks are removed from the transcriptions with an exception of word level punctuations like “bude-li”, “autorsko-spisovatelský”, etc.

The symbols specified in Table 19 are used to denote word truncations, mispronunciations, non-understandable speech, and non-speech acoustic events. As the recording was strongly controlled by the operator, the frequency of these non-speech marks is relatively small.

word truncations	`~word` or `word~`		at signal begin or end only
mispronunciations	`*word`		
non-understandable speech	`**`		separated by the blank from the rest of text
non-speech acoustic	`[fil]`	filled pause	at the correct location between

events			words
	[spk]	speaker noise	at the correct location between words
	[int]	intermittent noise	at the correct location between words
	[sta]	stationary noise	at the beginning of the noise (placed between words)

Table 19 – Annotation of non-speech marks

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

- [spk] – The most frequently used mark, typically for the events like lip smack, cough, grunt, throat clear, tongue click, loud breath, laugh, loud sigh.
- [fil] – Used for the "filled pause" between words. These events may be well modeled by the speech filled pause model.
- [int] – Used for noises of generally intermittent nature. Typically, it appears once (door slam), or periodically (telephone ringing), etc.
- [sta] – Used for of noises having less or more stable amplitude spectrum during the time. It is not used when such kind of noise is considered to be typical for the given environment and thus appears during whole session. It is used if the noise is present just in several utterances.

9. Lexicon information

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

During the annotation in the FTP-Transcriber, the "regular" Czech phonetic transcription is on-line generated for each orthographic annotation. If the phonetic transcription does not correspond to the real pronunciation, it is marked with a special syntax, from which above mentioned LBO and EPI fields and lexicon are generated consequently.

9.1 Czech SAMPA table

The pronunciation in CSO-files and in the lexicon file is in Czech SAMPA, see Table 20.

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: o:	rád móda	ra:t mo:da	glad fashion RARE
u:	půl, únor	pu:l, u:nor	half, February
Diphthongs			
o_u	mouka	mo_uka	flour
a_u	auto	a_uto	car
e_u	euforie	e_uforie	euphoria RARE
Plosives			
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
Affricates			
t_s	cíl	t_si:l	aim
d_z	leckdy	led_zgdi	at times RARE
t_S	čas	t_Sas	time
d_Z	džbán	d_Zba:n	jug RARE
Fricatives			
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
Liquids			
r	ret	ret	lip
l	led	let	ice
Nasals			
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 20 – Used phonemes from Czech SAMPA

In some particular cases also the following phoneme is used in the pronunciation transcription.

- "@" (schwa) – used in the second unofficial (phonetic) version of spelling. In other utterance types it usually does not appear.

Schwa			
@	DTW	d@ t@ v@	2-nd spelling of DTW

Table 21 – Additional phonemes used in lexicon

9.2 Phonetic annotations – a source for the lexicon generation

All phonetic transcriptions are done on the word level without context. In this sense the words are included in the pronunciation lexicon. However, the change between unvoiced and voiced equivalents of 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.

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.

10. References

[1] P. Pollák and J. Černocký. Czech SPEECON Adult Database – Documentation on the DVDs. IST-1999-10003 Deliverable, 2004.