SPEX / Dept. of Language and Speech
University of Nijmegen
Erasmusplein 1
NL-6525 HT Nijmegen
The Netherlands
E-mail: spex@spex.nl

SUBJECT:	Validation Czech SPEECON Corpus (Adults)
AUTHORS:	Dorota Iskra
VERSION:	1.1
DATE:	17 March 2009

INTRODUCTION

The speech databases made within the SPEECON project were validated by SPEX, Nijmegen, the Netherlands, to assess their compliance with the SPEECON format and content specifications, as documented in Deliverables 2.1, and 4.1 of the project.

The validation results of the Czech SPEECON database (adults) are contained in this document. This database is approved by the SPEECON consortium.

In the validation procedure we systematically check a list of validation criteria for a range of topics. In the following sections we will evaluate these criteria one by one. Validation results that call for attention because of deviations from the SPEECON specifications are marked by

⇒

so that they can be easily found.

CONTENTS

1 DOCUMENTATION

– File DESIGN.DOC is present

OK, DESIGN.DOC describes the first 550 sessions only. All the provided 590 sessions were validated.

– Language of doc file: English

OK

– Contact person: name, address, affiliation

OK

– Collectors and owners of the database

OK

– Number of disks

OK, section 2.4

– Contents of each disk

OK, section 2.4

– Formats of speech files

OK, section 3.1

– The directory structure of the disks
  - Database, block and session orderings
  - Directories DOC, INDEX, TABLE (and optionally HTML, PROMPT, SOURCE)

⇒ README.TXT should only be found on the documentation disk (section 2.2.1).

– File nomenclature
  - Root files
  - Names of speech files and label files

- Files in directories DOC, INDEX, TABLE (and optionally HTML, PROMPT, SOURCE)

OK

– Reference to the validation report made by SPEX (VALREP.DOC)

OK, section 2.3.2

– Contents and format of the label files
  - Clarification of attributes (three letter mnemonics)
  - Example of label file

OK, section 3.2
⇒ The labels MIP and MIT are missing from Table 12.

– Recording platform
  - Hardware set-up
  - Software set-up
  - Microphone types and positions

OK, section 6

– Speaker recruitment

OK, section 4.4

– Prompting
  - Presentation design
  - Prompting example for one recording session

OK, section 4.1 and 4.2

– Description of all the items in the database
  - Specification of the individual items in the database
  - Connection of prompted items to corpus codes in the database (in titles of subsections of individual corpus items)

OK, section 5
⇒ Not clear which corpus id is used for city and which for street names (section 5.3.13).
⇒ The symbols for rare phonemes in section 10.2 do not correspond to the ones used in the database.

– For spontaneous and elicited items the texts prompted to the speakers should be included together with an English translation

  OK

– A list of prompted digits in the language

  OK

– Tables with frequencies of the phones represented in the phonetically rich sentences, and in the phonetically rich words (at transcription level)

  OK, Table 48

– Transcription conventions
  - Procedure used
  - Quality assurance
  - Character set used for annotation (transcription) (ISO-8859 or other if needed)
  - Conventions used for transcription of spoken words
  - Annotation symbols for non-speech acoustic events: Filled Pause, Speaker Noise, Stationary Noise, Intermittent Noise, and additional, optional ones if used.
  - List of symbols used to denote word truncations, mispronunciations and not understandable speech
  - Case sensitivity of transcriptions
  - Use of punctuation

  OK, section 9

– Lexicon information
  - Case sensitivity of transcriptions
  - Procedures to obtain phonemic forms from orthographic input
  - List of SAMPA phone symbols
  - List of PinYins and Hepburn Romaji syllables (if applicable)
  - Statement whether or not the transcription and the lexicon are case sensitive
  - Information captured in the phone transcriptions (assimilation and reduction rules)
  - Statement whether multiple transcriptions are supported
  - Statement whether stress information is supplied
  - Statement whether there are any tags, and if so, the tagging conventions used, e.g., record (noun) vs. record (verb)
  - List of words that are from a foreign language
  - List of rare phonemes

- Analysis of frequency of occurrence of the phonemes represented in the COMBINED phonetically rich sentences and phonetically rich words, and in the full database (at transcription level); optional for statistics of diphones, triphones.
- Any other language-dependent information or conventions

OK, section 10

– Speaker demographics
  - Which regions, how many of each
  - Motivation for selection of regions
  - Which age groups, how many of each
  - Sexes: males, females; how many of each.
  - Number of speakers
  - Number of sessions

OK, section 8

– Recording environments
  - Description of the environments of the sessions:
    Office
    Entertainment
    Public place
    Car (use of high pass filter should be specified)
  - Description of sub-environments per environment
  - Distribution of speakers over environments and sub-environments

OK, section 7. No filter used.

2 DATABASE STRUCTURE, FORMATS AND FILE NAMES

– Directory / subdirectory conventions
  Format of directory tree should be
  \<database>\<block>\<session>
  - Database: defined as <dbName><#><language code>
    where <dbName> is ADULT for the adult speaker and CHILD for the children speaker database, <#> is 1 for SpeeCon, <language code> is the ISO 639 2-letter language code
  - Block: defined as BLOCK<nn> where <nn> is a progressive number from 00 to 99.
    Block numbers are unique over all disks.
    They correspond to the first two digits of <nnm> below.
  - Session: defined as SES<nnm> where <nnm> is the session code also appearing in file name

  OK

– File naming conventions
  - All file names should obey the following pattern: DDNNMCCC.LLF
  - DD: database identification code
    For SPEECON: SA for adult and SC for child speakers
  - NNM: session code 000 to 999
  - CCC: item code;
  - LL: ISO-639 language code (with extensions)
  - F: speech file type
    0,1,2,3 is for signal files of the four channels
    O is for label file

  OK

– NNM in filenames is not in conflict with BLOCK and SES numbers in pathname

  OK

– Contents lowest level subdirectories should be of one recording session only

  OK

– All text files should be in MS-DOS format (<CR><LF> at line ends)

  OK

– A README.TXT file should be in the root of each database describing all (documentation) files. (This file is also allowed as README.HTM).

  ⇒ *In the text CDs are referred to whereas the list contains a DVD distribution.*

– A file containing a shortened version of the volume name (11 chars max.) should be in the root directory. The name of this file is DISK.ID. This file supplies the volume label to UNIX systems that cannot read the physical volume label. Example of contents: ADULT1EN_01.

  OK

– A copyright statement should be present in the file COPYRIGH.TXT (root)

  ⇒ *The copyright statement includes a reference to CD-ROM.*

– Documentation should be in \<database>\DOC
  - DESIGN.DOC
  - PLATFORM.DOC (optional)
  - TRANSCRP.DOC (optional)
  - SPELLALT.DOC (optional)
  - SAMPALEX.PS
  - ISO8859<n>.PS
  - SUMMAR0.TXT
  - SUMMAR{1|2|3}.TXT (optional, only needed if files are missing in other channels than 0)
  - VALREP.DOC

  OK

– Tables should be in \<database>\TABLE
  - SPEAKER.TBL
  - LEXICON.TBL
  - REC_COND.TBL
  - SESSION.TBL

  OK

– Index files should be in \<database>\INDEX
  - CONTENT0.LST
  - CONTENT{1|2|3}.LST (optional, only needed for annotated channels)

  OK

– Prompt sheet files (optional) should be in \<database_name>\PROMPT

Not provided

– Empty (i.e. zero-length) files are not permitted

OK

– All table files, and index files should report the field names as the first row in the files using tabs as in the data records following.

OK

– The contents of the database as given in CONTENT{0|1|2|3}.LST should have the following order of attributes:
- full pathname (DIR:)
- speech file name (SRC:)
- corpus code (CCD:)
- speaker code (SCD:)
- speaker sex (SEX:)
- speaker age (AGE:)
- speaker accent (ACC:)
- scenario code (SCC:)
- orthographic transcription of uttered item (LBO:)

The first line should be a header specifying the information in each record.
This file must be supplied as an ASCII TAB delimited file.
Note: The contents of the CONTENT{0|1|2|3}.LST files are not disk-dependent.

OK

– The contents of the SUMMAR{0|1|2|3}.TXT files should have the following order of attributes:
- the full directory name where speech and label files are to be found (DIR)
- the session number (SES)
- two strings of typically N codes (CCD). Each item present in a string is represented by its code, separated by commas. The first string contains the item list as intended; the second string contains the item list as recorded. If the item is missing, a '---' should appear. The two strings are separated by a space.
- recording date (RED)
- recording time of first item (RET)
- optional comment text
- all these fields are separated by spaces
- also the noise recordings and silent word recordings should be included

Note: The contents of the SUMMAR{0|1|2|3}.TXT files are not disk-dependent.

⇒ *SUMMAR0.TXT should contain spaces instead of tabs between the different fields.*

– All sessions indicated in the documentation SUMMAR{0|1|2|3}.TXT are present

  OK

– Missing items per speaker should correspond to missing files reported in SUMMAR{0|1|2|3}.TXT)

  OK

– The database should be free of viruses.

  OK

3 CORPUS ITEMS: DESIGN AND COMPLETENESS

A. Check on mandatory corpus items

– Free spontaneous speech (F01-F30)
- Prompts are taken from the fixed list of 30
- 10 items per speaker or at least 2 minutes of speech (derived from END label)

OK, prompts from all 30 categories were found. All the prompts match the ones in DESIGN.DOC.
⇒ 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.

– Elicited dates (ED1-ED3)

OK, all the prompts match the ones in DESIGN.DOC.

– Elicited times (ET1-ET2)

OK, all the prompts match the ones in DESIGN.DOC.

– Elicited city names (EC1-EC2)

OK, all the prompts match the ones in DESIGN.DOC.

– Elicited proper name (EP1-EP3)

OK, all the prompts match the ones in DESIGN.DOC.

– Elicited spelling of proper name (EL1)

OK, all the prompts match the ones in DESIGN.DOC.

– Elicited Yes/No (EQ1-EQ2)

OK, all the prompts match the ones in DESIGN.DOC.

– Elicited language (EO1)

OK, all the prompts match the ones in DESIGN.DOC.

– Elicited telephone numbers (EN1-EN3)

OK, all the prompts match the ones in DESIGN.DOC.

– Phonetically rich words (W01-W05)
- Read
- Max. 10 repetitions per word
- Min. 300 different words

OK, 300 different words were found
32 words were found > 10 occurrences, which is probably due to extra speakers.

– Phonetically rich sentences (S01-S30)
- Read
- Max. 5 repetitions per sentence
- Min. of 3300 different sentences
- At least 500 samples per phone at transcription level (except for rare phones), counted with phon. rich words
- Each phone occurs in min. 90% of 550 sessions (except rare phones), counted with phon. rich words

OK, 3300 different sentences were found. All the phonemes except for the rare phoneme /@/ were found in at least 99% of the sessions.
929 sentences were found > 5 occurrences, which is probably due to extra speakers.

– Isolated digits (CI1-CI4)
- Prompted in words
- Min. occurrences per digit at transcription level: 1870/D

OK, all the 16 digit forms were found with 119-159 repetitions each (min is 116).

– Isolated digit string (CB1)
- Prompted in words
- Min. occurrences per digit at transcription level: 467

OK, an extra item CB2 was found.
⇒ For the digit form "osm" only 398 repetitions were found (min is 467). This deficiency is recognised in DESIGN.DOC.

– Connected digit strings (CC1-CC4)
- Prompted in words
- Min. occurrences per digit at transcription level: 9350/D

⇒ *For the following digit forms the number of repetitions is < min 584: čtyry (569), sedm (543) and osm (488). This deficiency is recognised in DESIGN.DOC.*

– Telephone number (CE1)
  - Prompted as 9-13 digits number
  - Including GSM numbers

OK, GSM numbers included.
⇒ *2 numbers are 14-digit long.*

– Natural numbers (CN1-CN3)
  - Prompted in words

OK

– Money amount (CM1)
  - Prompted in words
  - In 50% of prompts local currency should be used (EURO in EC languages)

OK, in 53% of the prompts local currency was found.

– Time phrases (CT1-CT2)
  - Prompted in words
  - One in analog format (CT1)
  - One in digital format (CT2)
  - CT1: Max. of 20 specific time words should be used

OK, 20 time-specific words were found

– Date phrases (CD1-CD3)
  - Prompted in words
  - One in analog format (CD1)
  - One in digital format (CD3)
  - CD2: max. 50 phrases
  - CD1,3: At transcription level:
    Each day name: min. 60 times
    Each month name: min. 35 times

OK, CD2: 39 different phrases were found; CD1: all the day names were found with 79-95 repetitions each, all the month names with 45-56.

– Spelt words (CL1-CL3)
  - Prompted in letter sequences

- One is artificial word
- Min. occurrence per letter at transcription level: 5775/L

OK, all the 42 letters were found with > 137 repetitions each.

– Person name (CP1)
- Set of 150

OK

– City/street names (CO1-CO2)
- Set of 275 cities
- Set of 275 street names

OK

– Yes/No (CQ1, CQ2)
- Min. of 465 yes/ 465 no at transcription level

OK, 588-590 repetitions of yes and no were found.

– E-mail and Web-address (CW1, CW2)
- CW1: set of 150 Web-addresses
- CW2: set of 550 E-mail addresses

⇒ CW2: 548 instead of 550 email addresses were found.

– Keyboard characters (CK1-CK2)
- Max. 20 characters
- Fixed set of 12 characters

⇒ 24 instead of 20 different keyboard characters were found.

– Application words (101-995)
- Set of 450-500 words from 9 categories
- Min. of 190 occurrences per word at transcription level

OK, 506 different words at the prompt level were found with >190 repetitions each. All the words in the specifications match the ones in DESIGN.DOC and these in turn match the ones in the database.

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

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

B. Checks on presence of corpus files (prompt level)

The following completeness checks are performed on obligatory SPEECON items only. Results are presented only for the annotated channel 0. Results of other channels are reported only in case of serious deviations.

1. Files that are not there

The following histograms present an overview per item code merging the distributions of items not present in the database, and empty files:

OK

2. Effectively missing files

For the adults' part, **6** files with empty transcriptions were found (i.e. files with only silence, noise symbols or ** in their transcriptions). When merged with the distributions of missing files given above, the following distribution of effectively missing files is obtained.

536: 1
910: 1
EC2: 2
EO1: 1
EP3: 1

– For the adults' part, SPEECON allows a maximum of 5% of the files for each mandatory corpus item as effectively missing.

OK, max 0.3%

3. Corrupted speech files

For the adults' part, **259** corrupted files were found (utterances which have only truncated or mispronounced words). When merged with the effectively missing files under B.2., the following distribution is obtained (only items with frequency > 2 are included):

103:3
181: 30
413: 3
425: 5
444: 3
449: 8

547: 5
556: 8
901: 8
903: 3
951: 4
958: 10
962: 10
974: 7
CK1: 3
EC2: 4
W01: 10
W02: 13
W03: 9
W04: 6
W05: 7
Y41: 3
Y58: 15
Y61: 3

– For the adults', part SPEECON allows a maximum of 7% of the files for each mandatory corpus item as either effectively missing or corrupted.

OK, max 5.1% per item

C. Automatic checks at transcription level

1. Match between prompt and transcription (only for isolated words; corpus codes: W01-W05, CI1-CI4, CP1, CO1-CO2, CQ1, CQ2, CK1-CK2, 001-995, Y01-Y..)

A mismatch between prompt and transcription is scored if the word in the prompt does not appear in the transcription; if there is no speech at all or only other word(s).

For the adults' part, **274** files with a mismatch between prompt and transcription were found. When merged with the effectively missing or corrupted files under B.3, the following distribution is obtained (only items with frequency > 2 are included):

103 3
109 3
131 14
181 31
223 3
413 3
425 5
444 3
449 8
547 5
556 8

567 5
628 5
901 8
903 3
909 3
910 4
951 4
958 11
962 10
974 7
CO2 14

– For the adults' part, SPEECON allows a mismatch between prompt and transcription text in a maximum of 10% of the files for mandatory isolated word items (effectively missing and/or corrupted items included).

OK, max 5.3% per item

2. Files containing truncation and mispronunciation marks

Truncation and mispronunciation marks (*, **, ~) are counted in the transcriptions of the individual items to obtain an idea of distorted speech data. This will not be used to reject or approve a database, but it will be supplied as supplementary information.

*For the adults' part, **3098** transcriptions containing a truncation or mispronunciation mark were found.*

4 SPEECH DATA FILES

– The speech files should be coded as PCM, 16 bit, 16 kHz, no compression

OK

– At least 90% of the files of all sessions (minus car sessions) has an SNR of 15 dB or more for the *close talk* channel (SNR value as measured by Sony software during the recording; SNQ label) (1)

OK, 97%

– Car recordings: If the 80 Hz high-pass pre-amplifier filter is switched off, then at least 80% of the files recorded in *Car* environment must have an SNR of 10 dB(A) or more for the close talk channel (SNQ label). If the HP filter is switched on, then (1) above is valid.

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

– At least 90% of the sessions recorded in the *Office* environment must have a noise range between 30-60 dB(A) (DBA label)

OK, 99%

– At least 90% of the sessions recorded in the *Entertainment* environment must have a noise range between 30-65 dB(A) (DBA label)

OK, 100%

– At least 90% of the sessions recorded in the *Public Place* environment must have a noise range between 45-90 dB(A) (DBA label)

⇒ *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.*

– At least 90% of the sessions recorded in the *Car* environment must have a noise range between 28-80 dB(A) (DBA label)

OK, 100%

– For every new environment and position a set of room impulse responses are required:

Recording environment	Number of measurements respective to positions	
	***'medium distance'* positions**	***'far distance'* positions**
Office	3	3
Entertainment	3	3
Public places	3	n.a.
Car	n.a.	n.a.

OK

5 ANNOTATION FILES

– Each line must be delimited by <CR><LF>

OK

– Mandatory (SAM) mnemonics for label files:
LHD: SAM 6.1
DBN: SPEECON_<LL>
SES: <session number>
DIR: <with backslashes and no final backslash>
SRC: <filename of speech file>
CCD: <corpus code = item code>
REP: <PLC value of the SCC-label>
RED: <recording date, in format DD/Mmm/YYYY>
RET: <recording time, in format HH:MM:SS>
BEG: <begin sample, 0>
END: <end sample>
SAM: 16000 <sampling freq.>
SNB: 2, signed <number of bytes per sample>
SBF: lohi} <sample byte order, meaningless with single bytes>
SSB: 16 <number of significant bits per sample>
QNT: PCM <quantisation>
NCH: 4 <number of channels>
SCD: <speaker code>
SEX: {M|F|UNKNOWN}
AGE: <in years|unknown>
ACC: <regional accent, place of growing up>
SNQ: CHN0=, CHN1=, … <signal quality, SNR, per channel>
MIP: CHN0=CLOSE_HEADSET, CHN1=CLOSE_LAVALIER, CHN2=MEDIUM, CHN3={MEDIUM|FAR}
MIT: CHN0=SENNHEISER_ME104,CHN1=NOKIA, CHN2={ SENNHEISER_ME64| AKG},CHN3={MBF_HAUN|PEIKER}
SCC: ENV={ OFFICE | ENTERTAINMENT | CAR | PUBLIC_PLACE },PLC=<env>_<nr>, POS={ CLOSE_WALL_nn | FAR_WALL_nn | NO_WALL_nn | CODRIVER }, SIZ={SQM_00_10 | SQM_10_20 | SQM_20-30 | SQM_30+| SQM_100_200 | SQM_200+}, AUD={ON|OFF},DRV={ ENGINE_ON | ENGINE_OFF | CITY_30_70 | COUNTRY_60_100 | HIGHWAY_90_130}
DBA: <dB (A) value>
LBD:
LBR: <start>, <end>, [gain], [minimum value], [maximum value], <orthographic prompt>
LBO: <start sample>, [centre sample], <end sample>, <transliteration>
ELF:

– Optional (SAM) mnemonics (i.e. may be omitted or left empty)
REG: <region of session>
TYP: orthographic
TXF: <name of the prompt sheet text file>
CMT: <comment>
ARC: <region or area code of session>
SHT: <sheet number for prompts>
CMP: <compression, should be empty if used>
EXP: <labelling expert>
SYS: <labelling system>
DAT: <date of completion of labelling>
SPA: <SAMPA version>
EDU: <education level>
SOC: <Socio Economic Status>
HLT: <health>
TRD: <tiredness>
ASS: <assessment code>

– Only legal mnemonics (labels) are used

OK, an extra mnemonic EPI has been used for phonetic transcription. Also the optional mnemonics SYS, EXP and DAT are used.

– All files must contain the same mnemonics. This holds as well for the optional mnemonics.

OK

– Order restrictions:
LHD and TYP are first
LBR and LBO come after LBD
ELF is end of file keyword

OK

– Neither illegal attributes nor illegal values should appear

⇒ *SNQ is 0 in at least one of the channels in 21 files.*

– For MIP and MIT the following arrangements are respected:

SCENARIO	CLOSE DISTANCE		MEDIUM DISTANCE		FAR DISTANCE
office,	Sennheiser ME	Nokia	Sennheiser ME	-	Mikrofonbau

entertainment	104	Lavalier HDC-6D	64		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	-
impulse response (_01-_03)	Mikrofonbau Haun MBNM-550 E-L		Mikrofonbau Haun MBNM-550 E-L		
impulse response (_04-_06)	Mikrofonbau Haun MBNM-550 E-L				Mikrofonbau Haun MBNM-550 E-L

OK

6 LEXICON

A. Formal check

– Check lexicon existence (\<database>\TABLE\LEXICON.TBL)

OK

– The entries should be alphabetically ordered

OK

– Used SAMPA symbols are provided in \<database>\DOC\SAMPALEX.PS

OK

– In transcriptions only SAMPA symbols are allowed

OK

– All SAMPA phoneme symbols should be covered

OK

– Phoneme symbols must be separated by blanks

OK

– A line in the lexicon should have the following format
<grapheme form> <TAB> [<frequency> <TAB>] <phoneme transcription> [<altern.>]
[TAB] is ASCII 9.

OK

– Each line is delimited by <CR><LF>

OK

– All entries should have at least one phone transcription

OK

– Alternative transcriptions are optional.
They may follow the first transcription, separated by [TAB] or have a separate entry (only in case also frequency information is supplied)

OK

– Orthographic entries are as a rule split by spaces only, not by apostrophes, and not by hyphens.

OK

– Words with *, or ~ should not appear in the lexicon

OK

– The lexicon should be complete
  - Check for undercompleteness (are all words in lexicon)

OK

  - Check for overcompleteness
(Undercompleteness is worse than overcompleteness. Overcompleteness cannot be a reason for rejection)

49 extra words were found

– Lexicon contents should be taken from actual utterances, so the entries should exactly match the transcriptions.

OK

– Optional information: stress, word/morphological/syllabic boundaries.
But, if provided, then it should follow the Speecon conventions.

Not provided

B. Content check by expert phonetician (carried out at pre-validation)

1000 lexicon entries should be checked for phonetic correctness by native speaker phoneticians that were not involved in the original transcription process, or by comparing with other available pronunciation lexicons.

The validation of the phonemic correctness of the lexicon entries is organised as follows:
- 1000 entries are randomly extracted from the lexicon;
- Of phonemic transcriptions only the first one is kept;
- The check is carried out at the segmental level only (not on syllable boundaries or stress marks, if provided)
- The check is carried out by a phonetically educated person who is a native speaker of the language
- The given transcription receives the benefit of the doubt
- The given transcription is correct if it represents a possible pronunciation of the word (which is not necessarily the most common)
- Each transcription is rated on a 3-point scale: OK; Minor error; Severe error
- A minor error occurs if only one symbol in the transcription is wrong
- A severe error occurs if more than one symbol is wrong

Criteria:
- A maximum of 10% minor errors is allowed. Minor means only one erroneous symbol in the transcription.

  OK, 67 minor errors were found (6.7%)

- A maximum of 5% severe errors is allowed. Severe means more than one erroneous symbol in the transcription.

  OK, 4 severe errors were found (0.4%)

7 SPEAKERS

– Check existence speaker database files (SPEAKER.TBL)

OK

– Obligatory information in SPEAKER.TBL:
- unique number (speaker/caller) SCD
- sex SEX
- age AGE
- accent ACC

OK

– Each line is delimited by <CR><LF>

OK

– Each field is separated by [TAB] (ASCII 9)

OK

– A minimum of 550 adult speakers is recorded

OK, 572 speakers were found in 590 sessions

– A database contains between 4 and 6 accent regions or dialects

OK, 4 accents were found

B. Checks for adult speakers

– Each gender must be represented between 45-55%; for 550 speakers 248-303 speakers per gender

OK, 292 sessions with female and 298 with male speakers

– For each of the four environments (office, entertainment, car, public places) the repartition of one gender must not fall below 30% or exceed 70%

OK, 32% sessions with female speakers were found in the car environment, 57% in entertainment, 53% in office and 51% in public place.

- Each accent region is represented by at least 0.70 * 550/D speakers, D being the number of dialects distinguished in the database.

  OK, 116-169 speakers per dialect

- For each dialectal region the repartition of one gender must not fall below 30% or exceed 70%

  OK

- In the office and public places environments (200 speakers each), each dialect region is represented by at least 0.5 * 200 / D speakers, D being the number of dialects distinguished in the database.

  OK, at least 31 speakers were found per dialect

- Ages: for adult speakers the following criteria are valid:

Age interval:	**Proportion of speakers**	**Requirement**
15-30	≥ 30%	45-55% male
31-45	≥ 30%	45-55% male
46+	≥ 10%	45-55% male

OK, per category 48% (including 1 speaker aged 15), 35% and 17% sessions were found with correct gender distribution.

8 RECORDING CONDITIONS

– Check existence and format of recording conditions tables (\<database>\TABLE\REC_COND.TBL) Required attributes:
  - Session number SES
  - Microphone position(s) MIP
  - Microphone type(s) MIT
  - Scenario code SCC

  OK

– Check existence and format session tables (\<database>\TABLE\SESSION.TBL). Required attributes:
  - Session number SES
  - Speaker code SCD
  - Recording place REP
  - Recording date RED
  - Recording time (of first item) RET

  OK

– Optional attributes SESSION.TBL:
  - Prompt sheet text file TXF
  - Sheet number SHT

  Not used

– Sessions should be distributed over the environments as follows

	Environment	**#Sessions**
Home	Office	190-210
	Entertainment	71-79
Mobile	Car	71-79
	Public Places	190-210

OK, per environment 211, 76, 210 and 93 sessions were found respectively

– For the OFFICE environment the following restrictions apply

Size (for documentation purposes only)	Place	Number of places	Position	Number of positions per place	Number of speakers per place and position	Number of speakers per position category

SQM_00_10, SQM_10_20, SQM_20_30, SQM_30+	OFFICE_01, OFFICE_02, ...	at least 4	CLOSE_WALL_01,..., CLOSE_WALL_04	1-4	1-10	80 – 120
			FAR_WALL_01,..., FAR_WALL_04	1-4	1-10	80 – 120

OK, 15 places were found with 1-4 positions per place with 110 and 101 sessions per position category.
In OFFICE_09, CLOSE_WALL_01 14 sessions were found (max is 10), but 5 come from the additional sessions.
In OFFICE_09, FAR_WALL_01 11 sessions were found (max is 10), but 1 comes from the additional sessions.

– For the ENTERTAINMENT environment the following restrictions apply

Size (for documentation purposes only)	Place	Number of places	Position	Number of positions per place	Audio	Number of speakers per place and position	Number of speakers per position category	Number of speakers per audio category
SQM_00_10, SQM_10_20, SQM_20_30, SQM_30+	ENTERTAIN_01, ENTERTAIN_02, ...	at least 3	CLOSE_WALL_01, ..., CLOSE_WALL_04	1-4	ON, OFF	0- 10	25-50	25-50
			FAR_WALL_01, ..., FAR_WALL_04	1-4	ON, OFF	0- 10	25-50	25-50

OK, 16 places were found with max 1 position per place, 0-9 sessions per place and position, 41 and 35 sessions per position category and 42 and 34 sessions per audio category.

– For the PUBLIC place environment the following restrictions apply

Size (for documentation purposes only)	Place	Number of places	Position	Number of positions per place	Number of speakers per place and position	Number of speakers per place category	Number of speakers per position category
SQM_100_200, SQM_200+	PUBHALL_01, PUBHALL_02, ...	At least 2	CLOSE_WALL_01, ..., CLOSE_WALL_04	0 – 4	0 – 10	90 - 110	40 – 70
			FAR_WALL_01, ..., FAR_WALL_04	0 – 4	0 – 10		40 – 70
	PUBOPEN_01, PUBOPEN_02, ...	At least 2	CLOSE_WALL_01, ..., CLOSE_WALL_03	0 – 3	0 – 10	90 - 110	20 – 45
			FAR_WALL_01, ..., FAR_WALL_03	0 – 3	0 – 10		20 – 45
			NO_WALL_01, ..., NO_WALL_03	0 – 3	0 – 10		20 – 45

OK, 10 places in PUBHALL and 13 in PUBOPEN were found with 0-3 positions per place, 55 and 45 sessions per position category and 100 sessions per place

category in PUBHALL; 35, 44 and 31 sessions per position category and 110 sessions per place category in PUBOPEN.

In PUBHALL_07, CLOSE_WALL_01 11 speakers were found, but 5 come from the additional sessions.

– For the CAR environment the following restrictions apply

Position	Place	Number of places, i.e. cars	Driving	Number of speakers per car category and driving condition	Number of speakers per car category
CODRIVER	CARMIDDLE_01, CARMIDDLE_02, ...	at least 1	ENGINE_OFF	3-9	30- 45
			ENGINE_ON	3-9	
			CITY_30_70	3-9	
			COUNTRY_60_100	3-9	
			HIGHWAY_90_130	3-9	
	CARUPPER_01, CARUPPER_02, ...	at least 1	ENGINE_OFF	3-9	30- 45
			ENGINE_ON	3-9	
			CITY_30_70	3-9	
			COUNTRY_60_100	3-9	
			HIGHWAY_90_130	3-9	

OK, 7 cars in CARMIDDLE and 3 in CARUPPER were found with 55 and 38 speakers per car category.

In CARMIDDLE, COUNTRY_60_100 15 sessions were found (max is 9) and 7 come from the additional sessions.

⇒ 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.

9 TRANSCRIPTION

A. Validation by software tools

– Transliteration is case-sensitive unless specified otherwise.
(In general lower case is used also at sentence beginning. Only exception: proper names and spelled words, ZIP codes, acronyms and abbreviations.
In the latter case blanks should be used in between the letters.)

OK, case-insensitive except for spelt letters.

– Punctuation marks should not be used in the transliterations

OK

– Digits must appear in full orthographic form

OK

– In principle only the following symbols are allowed to indicate non-speech acoustic events:
[fil] [spk] [sta] int]
Other symbols (and language equivalents) must be mentioned in the documentation

OK

– Asterisks should be used to indicate mispronunciations

OK

– Double asterisks should be used for not understandable parts

OK

– Tildes should be used to indicate truncations

OK

B. Validation by human experts

This validation involves 1000 short items and 1000 long items. 20% of the long items stemmed from the spontaneous speech. The items are proportionally selected from the

adults' database and the children's database. The results of the combined transcription validation for adult and child speakers are presented here.

The following corpus items are considered as short items: single word utterances (application words, single digits, Y/N questions, names, phonetically rich words). All other items are considered as long items.

A native speaker of the language performed the check on the speech part of each utterance. The transcription validation of the non-speech symbols (everything between squared brackets) was not necessarily done by a native speaker of the language, but by someone experienced in listening to background noises and capable to decide which noises should be transcribed or not. The transcriptions in the label files were checked by listening to the corresponding speech files and by correcting the transcriptions if necessary. As a general rule, the delivered transcription should always have the benefit of the doubt; only overt errors should be corrected.

Three types of errors are distinguished:

1. Errors in the transcription of speech
2. Errors in the transcription of non-speech (background noises)
3. Channel mismatch

A channel mismatch means that recordings that are supposedly simultaneous do not contain the same utterance or contain only part of the same utterance. One file of the other channels is linked to each tested file of the close-talk channel in order to test this.

The following error criteria are used:

1. For speech a maximum of 5% of the validated utterances (=files) may contain a transcription error.

2. For non-speech a maximum of 20% of the validated utterances (=files) may contain a transcription error.

3. A maximum of 5% channel mismatches may be found

RESULTS

1. Long items

Transcription errors with respect to speech were found in **36** items. This amounts to **1.8%**, which is below the criterion of 5%.

Errors in the transcription of non-speech were found in **22** items. This amounts to **1.1%** of the items, which is below the criterion of 20%.

Channel mismatches were found in 0 items. This amounts to 0% of the items, which is below the criterion of 5%.

2. Short items

Transcription errors with respect to speech were found in **9** items. This amounts to **0.5%**, which is below the criterion of 5%.

Errors in the transcription of non-speech were found in **23** items. This amounts to **1.2%** of the items, which is below the criterion of 20%.

Channel mismatches were found in 0 items. This amounts to 0% of the items, which is below the criterion of 5%.

3. Overall result

*When long and short item sets were put together, errors were found with respect to the transcription of speech in **45** items. This amounts to **2.3%**, which is below the 5% criterion.*

*Errors in the transcription of non-speech were found in **45** items. This amounts to **2.3%** which is below the 20% criterion.*

Channel mismatches were found in 0 items. This amounts to 0% which is below the 5% criterion.

10 SUMMARY

This database is approved by the SPEECON consortium.

1. Documentation

⇒ *README.TXT should only be found on the documentation disk (section 2.2.1).*
⇒ *The labels MIP and MIT are missing from Table 12.*
⇒ *Not clear which corpus id is used for city and which for street names (section 5.3.13).*
⇒ *The symbols for rare phonemes in section 10.2 do not correspond to the ones used in the database.*

2. Database structure, formats and file names

⇒ *README.TXT: In the text CDs are referred to whereas the list contains a DVD distribution.*
⇒ *The copyright statement includes a reference to CD-ROM.*
⇒ *SUMMAR0.TXT should contain spaces instead of tabs between the different fields.*

3. Corpus items: design and completeness

⇒ *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.*
⇒ *CB1-2: For the digit form "osm" only 398 repetitions were found (min is 467). This deficiency is recognised in DESIGN.DOC.*
⇒ *CC1-4: For the following digit forms the number of repetitions is < min 584: čtyry (569), sedm (543) and osm (488). This deficiency is recognised in DESIGN.DOC.*
⇒ *CE1: 2 numbers are 14-digit long.*
⇒ *CW2: 548 instead of 550 email addresses were found.*
⇒ *24 instead of 20 different keyboard characters were found.*

Application words:

⇒ *There is a discrepancy between the spelling in DESIGN.DOC and in the database: výstaviště vs výstaviště.*
⇒ *There are some inconsistencies in the spelling at the transcription level: audio video vs audio-video, rádio vs radio.*

4. Speech data files

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

5. Annotation files

⇒ *SNQ is 0 in at least one of the channels in 21 files.*

6. Lexicon

OK

7. Speakers

OK

8. Recording conditions

⇒ *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.*

9. Transcription

OK