

Swedish Project

Contents

Technical information

General information

Corpora

Subjects

Experimental recording conditions

Explanation of terms

References

Technical information

- Recordings were carried out at KTH, Stockholm during the year 1992, using the EUROPEC (V 4.11) software developed in the framework of the ESPRIT n°2589 SAM Project.
- Acquisition equipment was according to SAM recommendations:
 - B&K 4165 condenser measurement microphone
 - B&K 2603 amplifier
 - OROS AU22 I/O board
 - Sampling frequency: 20000 Hz.
 - samples coded on 16 bits.

General information

The EUROM1 database files are in accordance with the European ESPRIT N°2589-SAM recommendation, the main characteristics of which are listed below:

- digitized speech samples are gathered in a simple binary file.
- every speech sample file is associated an ASCII description file; this association relying on the file names.

The multilingual speech database EUROM_1 has been recorded in the framework of the ESPRIT n°2589 - SAM (Multilingual Speech Input/Output Assessment Methodology and Standardization) [1][4]. The recordings that form the EUROM1_S (Swedish part of EUROM1) were carried out at KTH, Stockholm, during the year 1992 [6]. The total amount of data is approximately 1.9 Gigabytes, that are stored on four CD-ROMs.

Corpora

Corpora consist of a Swedish adaptation (and not a strict translation) of the typical corpus commonly agreed.

- Continuous speech

- 40 passages made of five task related sentences.
- 50 patching sentences, designed to compensate for uneven phoneme distribution in the passage material. These sentences are not task related.

- CVC Words

The CVC word lists were built up according to the British material with five list types and also carrier phrases of the suggested type. Eighty-two isolated words were used. There were no problems in the reading of the items although some of the speakers did not have a clear difference in the pronunciation of some of the vowels, in accordance with what is known from Swedish dialects. All the Few Speakers recorded the CVC word material while the Very Few Speakers also recorded the CVC words in the carrier phrases as well as the carrier phrase words spoken in isolation.

- Numbers

The numbers were divided into five blocks, each containing twenty numbers according to the specification of the British material. Each block was recorded as one single take. The pronunciation of Swedish numbers can vary, and a short rehearsal was performed before the reading of this material. An important detail was that the numbers beginning with the digit "1" as in "133 (one hundred and thirty-three)" should be pronounced in this database recording. It is quite common in Swedish to delete the "one" and start with "hundred". This was a problem for some of the readers, who had to repeat these lists. The words for hundreds or thousands are pronounced with word accent I and II, sometimes as a free distribution for some speakers. We accepted both word accents in the readings.

Table of corpus

CORP. CODE	CORPUS NAME	N. of ITEMS	ITEMS TYPE
F0	Sentences	5	S
F1	Sentences	5	S
F2	Sentences	5	S
F3	Sentences	5	S
F4	Sentences	5	S
F5	Sentences	5	S
F6	Sentences	5	S
F7	Sentences	5	S
F8	Sentences	5	S
F9	Sentences	5	S
N1	Numbers	20	N
N2	Numbers	20	N
N3	Numbers	20	N
N4	Numbers	20	N
N5	Numbers	20	N
O0	Passage	1	P
...			
O9	Passage	1	P
P1	Passage	1	P
...			
P9	Passage	1	P
Q0	Passage	1	P
...			
Q9	Passage	1	P
R0	Passage	1	P
...			
R9	Passage	1	P
S1	CVC_1 with initial consonant variation	20	W
S2	CVC_2 with initial consonant variation	23	W
S3	CVC_3 with initial consonant variation	5	W
S4	CVC_4 with final consonant variation	19	W
S5	CVC_5 with vowels variation in a context	15	W
T1	CVC_1 in context 1	20	S
T2	CVC_2 in context 1	23	S
T3	CVC_3 in context 1	5	S
T4	CVC_4 in context 1	19	S
T5	CVC_5 in context 1	15	S

U1	CVC_1 in context 2	20	S
U2	CVC_2 in context 2	23	S
U3	CVC_3 in context 2	5	S
U4	CVC_4 in context 2	19	S
U5	CVC_5 in context 2	15	S
V1	CVC_1 in context 3	20	S
V2	CVC_2 in context 3	23	S
V3	CVC_3 in context 3	5	S
V4	CVC_4 in context 3	19	S
V5	CVC_5 in context 3	15	S
W1	CVC_1 in context 4	20	S
W2	CVC_2 in context 4	23	S
W3	CVC_3 in context 4	5	S
W4	CVC_4 in context 4	19	S
W5	CVC_5 in context 4	15	S
X1	CVC_1 in context 5	20	S
X2	CVC_2 in context 5	23	S
X3	CVC_3 in context 5	5	S
X4	CVC_4 in context 5	19	S
X5	CVC_5 in context 5	15	S
Y1	Words of the contexts in isolation	10	S

Legend (also appears in the filenames)

D	Digits
L	Letters
N	Numbers
S	Sentences
P	Passage
W	Words

Subjects

Sixty speakers of various ages and voices were chosen to comprise the speaker set. A subset of the Many Speaker Set consisting of ten speakers were chosen as the Few Speaker Set. From this group, another subset of four speakers (2M, 2F) were chosen as the Very Few Speaker Set, according to the recommendations in the SAM Document describing the database recording procedure. Dialectal differences were restricted in that we did not use speakers of dialects with distinctly other allophones, such as the southern Swedish dialect with back fricative /r/, but kept to a standard middle Swedish, Stockholm dialect, with minor deviations due to upbringing in other parts of Sweden. The Few Talker Set (and the Very Few Talker Set) were subjects that were used to be recorded in an anechoic chamber and to read diverse

type of material in a consistent way. We used an equal number of male and female speakers.

1st Group (MANY)

Sp. Code	Sex	DoB	Height	Weight	Educ. Level	Smoking	Accent	Pathol.
AK	M	1949	183	85	DSc	no	Stockholm	no
AM	F	1945	173	64	BA	medium	Stockholm	no
BA	F	1953	165	65	Student	no	Stockholm	no
BM	M	1948	176	74	Eng	no	Stockholm	no
BV	M	1965	186	85	Eng	no	Stockholm	no
CA	M	1968	176	75	Eng	no	Stockholm	no
CB	M	1933	170	64	MA	no	Stockholm	no
CR	M	1945	188	95	DSc	no	Stockholm	no
DM	M	1954	175	69	Eng	no	Stockholm	no
EB	M	1946	183	82	MSc	no	Stockholm	no
EE	F	1961	161	51	BA	has	Stockholm	no
EK	M	1946	173	71	Eng	no	Stockholm	no
EN	M	1964	170	65	Eng	no	Stockholm	no
FA	M	1958	192	73	Eng	no	Stockholm	no
FG	M	1919	184	65	Prof	no	Stockholm	no
GB	M	1946	187	79	Prof	occasional	Stockholm	no
GJ	M	1966	175	60	Eng	no	Stockholm	no
GP	F	1949	173	63	MD	no	Stockholm	no
HA	F	1963	173	68	Eng	no	Stockholm	no
HB	F	1941	164	56	DMSc	no	Stockholm	no
HJ	M	1966	180	77	Eng	no	Stockholm	no
HL	F	1948	167	64	BA	no	Stockholm	no
HT	F	1968	161	50	BA	no	Stockholm	no
IE	F	1950	175	58	Sp. Path.	no	Stockholm	no
IJ	F	1969	180	72	Student	no	Stockholm	no
JB	F	1969	171	52	High School	no	Stockholm	no
JE	M	1941	185	75	DSc	no	Uppsala	no
JM	F	1967	167	53	Eng	no	Stockholm	no
KA	F	1935	172	55	PhD	no	Stockholm	no
KB	M	1946	180	74	Eng	no	Stockholm	no
KC	F	1961	165	50	BA	medium	Stockholm	no
KI	F	1945	174	68	Eng	no	Stockholm	no
KJ	M	1944	178	75	Logoped	no	Stockholm	no
LE	F	1965	168	65	Logoped	no	Stockholm	no
LG	M	1948	174	87	BA	no	Stockholm	no
LI	F	1926	166	82	Speech	no	Stockholm	no

					Therapist			
LJ	M	1936	176	70	DSc	yes	Stockholm	no
LP	M	1955	183	108	Composer	no	Stockholm	no
LT	F	1972	177	68	High	medium	Stockholm	no
MA	F	1955	160	57	Sp. Path	light	Stockholm	no
MB	F	1964	170	64	Musician	medium	Stockholm	no
ML	M	1957	181	77	DSc	no	Stockholm	no
MS	F	1952	162	60	Logoped	no	Stockholm	no
MT	F	1952	168	54	Logoped	no	Stockholm	no
NC	F	1962	172	68	BA	no	Stockholm	no
NL	M	1947	181	90	DSc	no	Stockholm	no
NN	M	1956	196	100	Eng	no	Stockholm	no
OA	M	1947	171	60	Eng	occasional	Stockholm	no
OC	F	1966	176	65	Student	no	Stockholm	no
PE	M	1930	188	68	Eng	no	Stockholm	no
RA	F	1957	162	55	BA	no	Stockholm	allergy
RE	F	1949	175	70	Sp. Path	no	Stockholm	no
SG	F	1923	172	65	Actress	no	Stockholm	no
SJ	M	1936	201	95	Prof	yes	Stockholm	no
SK	M	1940	192	84	DSc	no	Stockholm	no
SL	F	1969	176	64	Student	no	Stockholm	no
SN	M	1966	182	75	Eng	no	Stockholm	no
SU	F	1947	180	77	MA	no	Stockholm	no
TS	M	1956	177	71	DSc	no	Stockholm	no
WM	F	1959	175	67	Linguist	yes	Stockholm	no

Number of speakers: 60

2nd Group (FEW)

Speaker Code	Sex	DoB	Height	Weight	Educ. Level	Smoking	Accent	Pathol.
AM	F	1945	173	64	BA	medium	Stockholm	no
BM	M	1948	176	74	Eng	no	Stockholm	no
CR	M	1945	188	95	DSc	no	Stockholm	no
GB	M	1946	187	79	Prof	occasional	Stockholm	no
HB	F	1941	164	56	DMSc	no	Stockholm	no
KI	F	1945	174	68	Eng	no	Stockholm	no
LE	F	1965	168	65	Logoped	no	Stockholm	no
MT	F	1952	168	54	Logoped	no	Stockholm	no
NL	M	1947	181	90	DSc	no	Stockholm	no
TS	M	1956	177	71	DSc	no	Stockholm	no

Number of speakers: 10

3rd Group (Very Few)

Speaker Code	Sex	DoB	Height	Weight	Educ. Level	Smoking	Accent	Pathol.
AM	F	1945	173	64	BA	medium	Stockholm	no
BM	M	1948	176	74	Eng	no	Stockholm	no
KI	F	1945	174	68	Eng	no	Stockholm	no
NL	M	1947	181	90	DSc	no	Stockholm	no

Number of speakers: 4

Legend:

DoB	Date of Birth
Educ. Level	Level of Education
Smoking	Smoking habits
Pathol.	Pathology

Experimental recording conditions

The recordings took place in an anechoic chamber at the Dept of Speech Communication and Music Acoustics, KTH. Outside the chamber there was a monitoring room with computer and tape recorder. In this room the introduction to the test took place and the subjects familiarized themselves with the task. During the recording session a microphone/loudspeaker system served as a communication link between subject and recorder.

The recording equipment consisted of

- a B&K half-inch microphone 4165
- a B&K sound level meter 2215
- OROS board AU21 in an Intel AT computer
- Tape back-up system TECMAR THS 2200
- a safety back-up DAT recorder: SONY Digital Audio Tape Deck DTC-1000 ES

The sound level meter with the microphone was placed in the recording room while the computer was outside in the monitoring room. The noise from the prompting text monitor was negligible. No laryngograph was used.

The setting of the OROS board was done according to the recommendations in [3], i.e. 20 kHz sampling frequency, variable input gain, line input). With the chosen sampling frequency the oversampling was set to 2. During the longer sessions (Few and Very Few Speaker Set) dumping of the hard disk content on data tapes had to be performed. This was done with a software program QT. For the Many Talker Set Corpora back-ups were done between sessions, thereby not interfering with the sessions.

A Continuous Recording Mode was chosen for all material except the Passage material for which the Signal Mode ("Automatic") was chosen. The calibration signals were recorded in Manual Mode. The choice of the continuous mode implied that the takes should contain no discontinuities and be error free. This made demands on subjects and recording staff as it is very easy to neglect small mispronunciations and slips in the reading. However, the recording sessions were successful in that the speakers were able to pronounce the often tricky sound strings quite consistently, especially the sentence material. The set-up of the recordings was as follows:

	Continuous Mode	Signal Mode
Triggering Level	0 dB	-40 dB
Extinction Level	-40 dB	-40 dB
End Silence	3000 ms	2000 ms
Signal Head	0	100 ms
Signal Queue	0	100 ms

Explanation of terms

Triggering Level	A level the signal must reach in order to trigger the recording process.
Extinction Level	A level the signal must cross down to be considered to be silence.
End Silence	The duration of silence that determines the end of the recording.
Signal Head	Part of the signal before the triggering moment which is kept on disk.
Signal Queue	Part of signal after the extinction moment that is kept on disk.

References

- [1] ESPRIT n°2589 - SAM Final Report, Year three. SAM-UCL-G004.
- [2] Barry W.J., Fourcin A.J., "Selection of speakers", SAM report: SAM-UCL-030, 20 May 1991.
- [3] Tomlinson M.J.: "Guide to Database Generation - Recording Protocol", Final

version - SAM-RSRE-015 - January 91. SAM internal document, source: RSRE, Malvern, UK.

[4] Sherwood T., Fuller H. "Guide to EUROM_1 speech database", SAM report SAM-NPL-102, 20 April 1992.

[5] Zeiliger J., "Publishing CD-ROMs from EUROM_1", SAM-A report: SAM-A/ICP/004/V1, 17 March 1993.

[6] Nord L., "Documentation of the Swedish EUROM-1 Database", SAM-KTH-1992-02, 21 April 1992