NORWEGIAN

Contents

Summary

The Norwegian EUROM_1 corpus is distributed on four CD-ROMs. The corpus consists of speech recorded from 60 informants, 30 male and 30 female. The recordings were made in an anechoic chamber, directly digitized at a sampling rate of 20 kHz. The textual content of the corpus are 100 numbers in the range 0-9999, CVC-words, CVC-words in left and right context (i.e. word triplets), 5-sentence mini-passages and filler sentences. The corpus is divided in three sub-corpora. The "Many-talker" set consists of all 60 speakers. All speakers read all 100 numbers, 3 mini passages and 5 filler sentences. The "Few-talker" set consists of 10 of the informants, 5 males and 5 females. These speakers read all 100 numbers and the CVC words 5 times, plus 15 mini-passages and 25 filler sentences in addition to the recordings they

produced for the "Many-talker" set. The "Very few talker" set consists of 2 male and 2 female speakers. These speakers (who were also part of the two other subsets of the corpus) read the CVC-words within five contexts once and the (isolated) context words five times.

The data: The speech and laryngograph samples are 16 bit linear stored in binary, headerless data files. The laryngograph recordings were only made for the "Few talker" and "Very few talker" sets. Label files in the SAM standard containing the orthographic transcription of the speech files are in the same directories as the speech (and laryngograph) files.

The collection of speech material for EUROM_1 took place at the Norwegian Institute of Technology (NTH) during the summer 1991 with financial support from the Royal Norwegian Council for Scientific and Industrial Research (NTNF) and Norwegian Telecom Research (TF). The recording was designed to be used within the ESPRIT-SAM project as the Norwegian part of the multi-lingual EUROM_1 speech database. The aims of this report are to enable speech researchers to make full use of the Norwegian EUROM_1 speech database.

Background

In the ESPRIT project 2589, Multi-lingual Speech Input/Output Assessment, Methodology and Standardisation (SAM), the main objective was to define and develop standards for manufacturing and assessment of speech technology products across languages within the European Community and the EFTA-countries Sweden and Norway. Thus, within the SAM-project the multi-lingual EUROM0 [Grice'89] and EUROM1 [Sherwood'92] speech databases have been compiled using common hardware [Lindberg'89] and software for speech recording e.g. EUROPEC [Zeilinger'91], and speech analysis, PTS [Caerou'91]. The software packages and the computer readable SAM Phonetic Alphabet, SAMPA [Wells'92], were developed within the SAM project and are now increasingly used as de facto standards in Europe.

For EUROM0, recordings of Danish, Dutch, English, French and Italian were compiled on a CD-ROM. The different languages were recorded in different sites using common recording protocols. However, the recordings got different levels of background noise because some sites recorded in anechoic chambers whereas others recorded in quiet office rooms. This may have influenced the succeeding annotations of some recordings. E.g. the high-frequency noise in the Danish EUROM0 recording has been claimed to be difficult to distinguish from word final friction [Grice'89,

p.43]. The EUROM0 CD-ROM contains digits, digit-triples and a continuous passage. The corresponding Norwegian and Swedish continuous passage recordings have also been compiled, but these are not stored on a CD-ROM. The Norwegian continuous passage was manually annotated according to the description in [Kvale'91] and [Kvale'93].

The EUROM0 speech corpora were rather small, since only four informants (2 males and 2 females) for each language read the continuous passage which took approximately 2 minutes for each one. This passage was not phonemically balanced (for more details, see [Kvale'93]). Therefore, we decided within the SAM project to compile a larger and more carefully defined speech database, called EUROM1. These recordings were carried out in accordance with [Tomlinson'90] for Danish, Dutch, English, French, German, Italian, Norwegian and Swedish. Descriptions of these recordings have been collected and edited by Tim Sherwood and Hillary Fuller at NPL (National Physical Laboratory), United Kingdom [Sherwood'92]. For Norwegian, the paper: "Orthographic prompting text and phonotypical transcription of Norwegian EUROM.1 recordings" by [Kvale'92], lists all the orthographic prompting texts and the corresponding phonotypical transcriptions of the 5-sentence blocks for the Norwegian EUROM.1 (see below). The phonemic SAMPA symbol set [Wells'92] was used in the phonotypical transcription.

Database Production

1 Recording Environment

The recordings were performed in the anechoic chamber at NTH, division for Telecommunication, using the SESAM workstation [Lindberg'89] with the OROS AU22 signal processing card and the EUROPEC v. 4.0 recording software. Calibrations of the recording chain were performed according to [Tomlinson'90]. A calibration with a 1 kHz sine wave was recorded and stored for each speaker (see file names below).

The recording session was triggered by the signal level. In order to be certain to include voiceless fricatives in the beginning of an utterance, the signal was buffered and the recording included 200ms before the signal level exceeded -30 dB. The recording session stopped when the signal level had been below -40dB for 1 second. The informants were sitting on a soft chair with their head position fixed. The prompting texts to be read aloud were given on a sheet of paper which the informants conveniently held. The AKG CK22 microphone was placed 50 centimetres from the speaker's mouth, at the mouth level, but offset by 15 degrees from the axis. The sampling-frequency was 20kHz.

Other technical details regarding e.g. the recording equipment and calibrations performed, are described in [Parelius'91].

2 Database Text Corpus

The prompting text consists of CVC-words, numbers, 5-sentence blocks, and filler sentences. Here, an overview of the texts is given. The phoneme codes are given in SAMPA symbols [Wells'92] and enclosed in backslashes / /.

2.1 CVC-words

CVC-word is an abbreviation for words with the structure Consonant-Vowel-Consonant, such as the word "man". The CVC-words were systematically structured to cover the following five cases (named with the textblock identifiers used in the recordings and in the corresponding file names):

S1 All initial consonants in fixed _VC context.
S2 Selected initial consonants in corner vowel (/i/, /A/, /u/) + C-context.
S3 Selected initial consonant clusters in fixed _VC context.
S4 All final consonants in fixed CV_ context.
S5 All vowels in fixed C_C context.

In case S1, the _VC-context was selected to be /i:l/, giving 17 words of the type /pi:l/, /bi:l/ and so on. Note that the nasal /N/ and the retroflex consonants never occur in the beginning of words in Norwegian.

In case S2, the selected initial consonants were the plosives /p b t d k g/, the fricatives /f v C s S/ and the nasal /m/. Since we had selected the final consonant to be /l/ we only needed to construct words with /A/ and /u/ in addition to the words in case S1. Thus, textblock S2 consists of 24 CVC-words.

In case S3, we selected three types of initial consonant clusters: (1) all the plosives + /r/, (2) bilabial and velar plosives + /l/, and (3) velar plosives + /m/. Note that the words in case S3 are of the form CCVC, i.e. with two consonants before the vowel. For simplicity, all the monosyllabic words defined in this section are called CVC-words in this paper.

In case S4, the CV_ -context was selected to be /sA:/, giving 22 words in this textblock.

In case S5, the C_C-context is selected to be /s/_/k/, giving 18 words in this textblock.

All the CVC-words were pronounced both singly and embedded in carrier phrases. The carrier phrases were five well defined contexts consisting of one word preceding and one word following the actual CVC-word. The CVC-word surrounded by the

two words provide controlled data for coarticulation research. In addition, reading such small sentences reduces certain prosodic phenomena which occur when reading single word lists. We selected the same context-structures as were recommended for the other languages participating in the SAM-project. I selected the context words for Norwegian as:

Context 1		Context 2	Context description
trykk	CVC-word	snart	(final /k/ --- initial /s/)
press	CVC-word	litt	(final /s/ --- initial /l/)
mal	CVC-word	kult	(final /l/ --- initial /k/)
få	CVC-word	over	(final and initial back rounded vowel)
si	CVC-word	is	(final and initial front unrounded vowel)

For instance with the CVC-word "pil" in textblock S1, the carrier phrases become "trykk pil snart", "press pil litt", "mal pil kult", "få pil over", and "si pil is". Thus, most of these sentences were on the form VERB NOUN ADVERB. For reading, we constructed one text block of carrier phrases corresponding to each textblock of single CVC-words. The textblocks are called T1-T5, U1-U5,V1-V5, W1-W5 and X1-X5. Hence, textblock T1 is the carrier "trykk - snart" applied on the single CVC-words in S1, while textblock T2 is the same carrier phrase applied on the single CVC-words in S2. Similarly, textblock U4 is the carrier "press - litt" applied on the single CVC-words in S4.

2.2 Numbers

A selection of 100 numbers in the range of 0 to 9999 were specified in such a way that all the phonotactic possibilities of the number system were covered. We used the same numbers as proposed by University College London (UCL). In Norwegian there are several ways of pronouncing the same number. Thus, to ensure uniform pronunciation, all numbers were presented both as a string of digits and orthographically. In addition, for the number 16 the informants were told how to say it, i.e. the prompting text was in this case 16 seksten /s{isten/. (In practice, I expect that most realisations of this word will be /s{ist=n/. For reading, the numbers were grouped into 5 blocks of 20 numbers each, named N1-N5.

2.3 5-Sentence Blocks or Mini-Passages

In order to provide prosodic structure above the single-sentence level, 40 mini-passages each consisting of 5 sentences with coherent semantic structure, were read. The mini-passages were translated from the English original produced by UCL, to resemble each other in syntactic and semantic complexity for the different languages.

However, some modifications and adaptations for Norwegian were made, but we tried to preserve the same structure of the sentences.

The mini-passages are divided into two main groups, called Inquiry situations and Descriptive/ Conversational passages. Most of the inquiry passages are situations where some information or help via phone is asked for. Descriptive passages are mini-passages where a person explains something to another or they are examples of broadcasted information.

The mini-passages were named O0-O9, P0-P9, Q0-Q9 and R0-R9. Of these, O0-O6, Q6-Q9 and R0-R1 were inquiry situations, and the rest, O7-O9, P0-P9, Q0-Q5 and R2-R9 were descriptive or conversational passages.

Based on the phonotypical transcription of the 40 mini-passages [Kvale'92] an analysis of the phoneme and diphone distributions was performed by the software package SAMTRA by the University of Bielefeld, Germany. This investigation showed that some phonemes and diphones occurred rarely. Our own investigation of the phonotypical transcription is shown in chapter 5 below. The following phonemes occurred less than 50 times: the long vowels /2:/ and /y:/, the short vowels /{/, /2/, /y/, the fricative /C/, the retroflexed consonants /rt/, /rd/, /rn/, /rl/, and /rL/, and the diphthongs /{i/, /2y/, /A}/, /Oy/, /}i/. In contrast the schwa @ occurred more than 700 times. Hence, to make the phoneme coverage of the recordings more comprehensive, a set of filler sentences was necessary.

2.4 Filler Sentences

Based on the investigation of the phonotypical transcription of the 40 mini-passages described above, we designed 50 filler sentences which were especially rich in the phonemes that were represented less than 50 times in the mini-passages. The sentences were not semantically connected, but each filler sentence was meaningful.

For the reading session, the filler sentences were grouped into 10 blocks of 5 sentences each, named F0-F9.

3 Subjects

The Norwegian EUROM.1 database contains 30 female and 30 male native speakers. Some phonetically relevant personal information on each subject, such as age, sex, weight, height, smoking habits, and dialectal background, is listed in Appendix A.

Following the Recording Protocols, [Barry'91], the informants should vary as much as possible, especially a wide range of accents should be obtained for the database. Our goal was that the informants for the Norwegian EUROM1 should be selected so that at least the dialects in the four biggest cities in Norway; Oslo, Bergen, Trondheim and Stavanger, were represented.

However, due to limited resources the informants were primarily selected among our friends and students and staff at NTH. They may thus not represent an average of the Norwegian population. Some personal data for the 60 informants may underline this fact - 23 had their origin in Mid-Norway (i.e. Tr¢ndelag), 17 South Eastern Norway, 8 Northern Norway, 6 North-Western Norway, and only 2 from Southern-Norway, Stavanger and Bergen respectively. Only 4 smoked. 43 were between 20 and 30 years of age, none was older than 60 (see table 1 below). There were 23 students from NTH, 14 with M.Sc degree or higher, 10 engineers, 8 secretaries, 2 economists, and 3 others.

Age	**N**	**Talker ID (male)**	**N**	**Talker ID (female)**
20 - 30	23	AJ(*), BB, BE, EC, GB, GO HO, KE, KK, KR, LS(*), NK, RH(*), SD, SE, SG(+) SJ, SS, TE, TH, TS, WP, WS	20	BC, BR, EA, GK(*), HH, HL, HS, JI, JM, KT, LN, LE, OH, PM, RE(*), RS(*), SB, SI, ST, TA
31 - 40	3	LJ, NT, OG	3	AI(+), AR, WZ
41 - 50	2	AT, JT	6	BA, BH(+), GL, JA, LI, SA
51 - 60	2	FA(+), OK	1	BS

Table 1: The distribution of speakers across various age categories

The Few Talkers are marked with (*), while the Very Few Talkers informants are marked with (+) (see section 4).

The informants practised reading the given text before the actual recording took place. Although there is no standard pronunciation of Norwegian, the informants were asked to read aloud the prompting text in "standard South-Eastern Norwegian". However, tones and phonemes such as the /r/, may be very difficult to "normalise" for people with a different dialect background than the South Eastern Norwegian.

To ensure a common reading of the numbers, the numbers were written out in digits and words. If the CVC-words were uncommon or meaningless, the informants were instructed how to pronounce the words (according the phonotypical transcription).

4 Organization of the Reading Session

The Norwegian EUROM.1 database consists of three speech corpora:

1. Many talker set
2. Few talker set
3. Very few talker set

For the Many Talkers Set, all 60 informants read aloud all 100 numbers, 3 mini-passages, and 5 filler sentences.

For the Few Talkers Set, 5 females and 5 males were selected from the Many Talker set. These informants read the 100 numbers five times and the CVC-words five times, plus 15 mini- passages and 25 filler sentences (in addition to the mini-passages and filler sentences read for the Many Talker Set).

Four of the informants (2 females and 2 males) in the Few Talker Set formed the Very Few Talkers Set. In addition they read the CVC-words within the 5 contexts once and the context words themselves five times.

For the informants in the Few and Very Few Talkers Set the laryngograph signal was also recorded.

The mini-passages and the filler sentences prompting texts were used in strict rotation. For example, for the Many Talker Set, informant 1 read mini-passages 1-3 and filler sentences 1-5. Informant 2 then read mini-passages 4-6 and filler sentences 6-10, and so on as shown in Table 2 for the Many Talker Set.

The same rotation principle applies to the Few Talkers Set as shown in Table 3.

Note that the Mini-Passages were not all read an equal number of times each. For the Many Talkers Set all the Mini-Passages were read four times, and the 20 first were read once more. For the Few Talkers Set, all the Mini-Passages were read three times, and the 30 first were read once more.

All the Filler sentences were read six times in the Many Talkers Set and five times in the Few Talkers Set.

Many Talker Set:

Subject	**Mini-passage**	**Filler sentence**
1	1 - 3	1 - 5
2	4 - 6	6 - 10
3	7 - 9	11 – 15
4	10 - 12	16 - 20
5	13 - 15	21 - 25
6	16 - 18	26 - 30
7	19 - 21	31 - 35
8	22 - 24	36 - 40
9	25 - 27	41 – 45
10	28 - 30	46 - 50
11	31 - 33	1 - 5

12	34 - 36	6 - 10
13	37 - 39	11 – 15
14	40 - 2	16 - 20
15	3 - 5	21 - 25
16	6 - 8	26 - 30
.	. .	. .
.	. .	. .
60	18 - 20	46 - 50

Table 2 Organisation of the reading session for the Many Talker Set

Few Talker Set:

Subject	**Mini-passage**	**Filler sentence**
1	1 - 15	1 - 25
2	16 - 30	26 - 50
3	31 - 5	1 – 25
4	6 - 20	26 - 50
5	21 - 35	1 - 25
6	36 - 10	26 - 50
7	11 - 25	1 - 25
8	26 - 40	26 - 50
9	1 - 15	1 – 25
10	16 - 30	26 - 50

Table 3 Organisation of the reading session for the Few Talker Set

File names

The files were named by the EUROPEC recording software according to the following naming scheme:

T T P P X X X X . C N F

where

T T is the informant's ID, i.e. the initials of his name

P P is the name of the text block (a letter and a number), where

C	= Calibration signal
F	= Filler sentences
N	= Numbers
O-R	= Mini-passages
S	= CVC-words
T-X	= CVC-words in context
Y	= Context words

X X X X is the serial number for the recording of the actual subject

C is the Corpus index, where

C	= Calibration signal
F	= Filler sentences
N	= Numbers
P	= Mini-passages
S	= CVC-words in context
W	= Isolated words (CVC, context words)

N is the Nationality index, e.g. N= Norwegian

F is the Filetype, where

S	= Sampled speech
O	= Orthographic file
L	= Laryngograph signal

An orthographic file contains the text block and phonetically relevant information about the subject

The serial number for the recording of actual informant had the following values:

	File number	**File type**
Many Talker set:	1-5	Numbers
	6-8	Mini-passages
	9	Filler sentences
Few Talker set:	10-34	Numbers
	35-59	CVC-words
	60-74	Mini-passages
	75-79	Filler sentences
Very Few Talker set:	80-104	CVC-words in context
	105-109	Context words

As an example, speaker AR was in the Many Talker set and read all the numbers, the filler sentence block F1, and the mini-passages O3, O4 and O5. Since the numbers were divided into 5 blocks with 20 numbers in each, we got nine recordings:

Numbers	**Mini-passages**	**Filler sentences**
arn10001.nno	aro30006.pno	arf10009.fno
arn10001.nns	aro30006.pns	arf10009.fns
arn20002.nno	aro40007.pno	
arn20002.nns	aro40007.pns	
arn30003.nno	aro50008.pno	
arn30003.nns	aro50008.pns	
arn40004.nno		
arn40004.nns		
arn50005.nno		
arn50005.nns		

Phonotypical Transcription

By Phonotypical Transcription we mean a transcription that is based on our prediction of how Norwegian orthographic text would be read out aloud. Hence, for continuous speech the phonotypical transcription allows for assimilations and elisions across word boundaries.

The computer readable phonemic symbol set SAMPA [Wells'92] was applied according to the Norwegian phoneme inventory defined in [Kvale'92].

As mentioned in section 2.3, the informants were asked to read aloud the prompting text in "standard South-Eastern Norwegian". However, there is no defined standard pronunciation of Norwegian, so many words in the EUROM.1 database were probably pronounced differently, e.g. the capital Oslo pronounced /uslu/ or /uSlu/. In addition, certain phonemes such as the /r/, may be very difficult to "normalise" for people with different dialect backgrounds.

In the phonotypical transcription we had to select one out of several alternative pronunciations. For the particular word "Oslo", or more generally the sl-transition, we decided to use /sl/ in our phonotypical transcription. Words like "blå" ('blue'), were predicted to be realised with the retroflex flap, i.e. /brLO:/.

In some cases it is difficult to decide which symbol to use, e.g. when transcribing the short e- sound which has got several allophones but which has to be transcribed by one of the two symbols /e/ and @ (note that @ is not defined as a phoneme in the Norwegian Phoneme Inventory [Kvale'92]). A similar problem is encountered when transcribing vowels in the /u/-/O/ area.

Experience with manual segmentation and labelling of the Norwegian continuous passage corresponding to the multi-lingual EUROM0-recordings [Kvale'91], [Kvale'93], showed that when a homorganic nasal or lateral follows a plosive, as in "atten" /Atn/ ('eighteen') and "b¢ddel" /b2dl/ ('executioner'), there is a nasal and lateral release respectively of the closure phase. The syllabification of e.g. the /n/ in "atten" is marked as /=n/.

Short and long vowel are phonemic in Norwegian. When marking length we have made these assumptions: All stressed Norwegian syllables contain an element of length which is realized by a long vowel e.g. /e:/, a diphthong e.g. /{i/, or a short vowel followed by one or more consonant(s) e.g. oss - /Os/, ost - /ust/. Thus only long single vowels are marked for length.

In the Mini-Passages and the Filler Sentences stress and tone have been marked on multisyllabic words according to our prediction of how the texts would be read out. The SAMPA-symbol for primary stress and word tone I is " and for word tone II the SAMPA- symbol is "" [Wells'92, p.6]. As an example, the word rota (n) ('the root') is transcribed /"ru:tA/ and the word rota (v) ('messed') is transcribed /""ru:tA/. For monosyllabic words stress has been omitted since stress in those cases have no phonemic value in Norwegian.

All words pronounced in isolation are assumed to be stressed, and thus transcribed in the citation form. Normally, a vowel preceding a single consonant is realised as a long vowel, e.g. "vin" /vi:n/, whereas a vowel preceding two or more orthographic consonants is realised with a short vowel, e.g. "vinn" /vin/. Exceptions have to be made for words like "pram" /prAm/, where a doubling of the consonant is not allowed in word final position in Norwegian orthography.

The phonotypical transcriptions are enclosed in backslashes / /. In order to increase the readability of the phonotypical transcription, the transcribed words are separated with a blank when the coarticulation across word boundaries is not expected to result in phonemic assimilations. That is, the blanks do not indicate pauses in the reading.