Title: **Availability of Finnish SpeechDat-Car database for ETSI STQ WI008 front-end standardisation**

Source: **Nokia**

Date: **16-Nov-1999**

Version: **1.2**

1. BACKGROUND

ETSI DSR TRACK 2 standardisation initiative tries to find out a significantly more noise robust speech recognition front-end than the one standardised in TRACK 1. For performance verification, realistic test data is needed. However, commercial speech databases for this purpose either do not exist, or they are rather expensive.

Currently, it has been planned that ***isolated digits*** and ***digit strings*** (excluding spontaneous material) from SpeechDat-Car databases would be utilised for this purpose. Since SpeechDat-Car databases are still in the collection phase, utterances from the first 133 speakers have been proposed to be included in so called *SDC-Aurora* database.

Remaining questions are the following:

- What languages are included in SDC-Aurora database?
- When are these languages available?
- Is there enough material in the database (matched/mismatched)?
- What is the pricing policy?

This document aims at answering these questions from Finnish language point of view.

2. FINNISH SDC-AURORA DATABASE

Finnish SpeechDat-Car database collection has progressed according to the plans, and currently 400 sessions have been recorded, out of which 300+ have been annotated. Corresponding amount of speakers are 200 and 150+. That is, SDC-Aurora specifications are already exceeded.

So far, recordings have been carried our in five different driving conditions, listed below:

Driving condition	**Condition**
0 km/h, engine on	Quiet
40-60 km/h	Low noise
40-60 km/h, window open	Low noise
100-120 km/h	High noise
100-120 km/h, music on	High noise

Table 1: Finnish SpeechDat-Car recording status

All sessions have been recorded with a close-talking (CT) microphone. In addition, hands-free (HF) microphone recordings are also available.

In Finnish language case, we propose that the sessions are extracted from the database in the following way:

Condition	Quiet	Low noise	High noise
Sessions (% of total)	25	37.5	37.5

Table 2: Finnish SDC-Aurora database division into different conditions

In ETSI DSR TRACK 2 tests, so called well-matched, medium-mismatch and high-mismatch conditions are used. In Finnish SDC-Aurora database, we propose that training and testing for these conditions are carried out in the following way:

Well-matched condition:

- Training with 70% of all conditions (HF and CT microphones)
- Testing with 30% of all conditions (HF and CT microphones)

Medium-mismatch condition:

- Training with 70% of *quiet* and *low noise* (HF microphone)
- Testing with 30% of *high noise* (HF microphone)

High-mismatch condition:

- Training with 70% of all conditions (CT microphone)
- Testing with 30% of *low noise* and *high noise* (HF microphone)

For each test, speakers will be split into separate training and test sets so that 70% of the speakers are in the training set and the rest in the test set.

The well-matched condition would be a so called multi-environment test, in which training and testing material consist of recordings made in many different driving conditions. Both microphones, close-talking and hands-free, are be used for training and testing.

For medium-mismatch condition, training is done using quiet and low noise material, and testing is done with high noise material. Medium-mismatch condition test is done with hands-free microphone only.

In high-mismatch condition case training is done using close-talking microphone material from all driving conditions. Testing is done using hands-free microphone material taken for low noise and high noise driving conditions.

3. CURRENT FINNISH SDC_AURORA DATABASE STATISTICS

In terms of number of repetitions per each digit, this has resulted in the following data sets:

	Well-matched	**Medium-mismatch**	**High-mismatch**
Training	1778*	561	889
Testing	770*	146	283

Table 3: Data set sizes (in terms of utterances per digit)

*Notice that the same utterance recorded with two microphones is considered as two separate repetitions.

The gender distribution for the utterances in various noise conditions is given below in Table 4.

Condition	Driving condition	Males	Females
Quiet	0 km/h, engine on	259	291
Low noise	40-60 km/h	207	206
Low noise	40-60 km/h, window open	206	206
High noise	100-120 km/h	207	206
High noise	100-120 km/h, music on	206	206

Table 4 Gender Distribution for the utterances in different driving conditions

Table 5 outlines the total number of male and female speakers in each noise condition. Utterances of 133 speakers have been used in this database.

	TRAIN		TEST	
Condition	Males	Females	Males	Females
Quiet	27	31	27	30
Low noise	44	42	43	41
High noise	42	42	42	40

Table 5 Number of speakers in different driving conditions

Finally, Table 6 provides the pronunciation variations that are present in the current subset of the SpeechDat-Car database made available for AURORA evaluation.

One	Two	Three	Four	Five	Six	Seven	Eight	Nine	zero
yks	kaks	kolme	neljä	Viis	kuus	seitsemän	kahadeksan	yhdeksän	Nolla
yksi	kaksi	kolome	nelijä	Viisi	kuusi	seittemä	kahdeksa	yheeksän	
			nelejä			seittemän	kahdeksan	yheksä	
							kaheksan	yheksän	
							kaheeksa	yhdeksä	
							kaheeksan	yheeksä	
							kaheksa		

Table 1 List of pronunciation variations included in the AURORA subset of database

4. AVAILABILITY OF FINNISH SDC-AURORA DATABASE

The Finnish SDC-Aurora database will be available as soon as the SDC-Aurora database for other languages are available. Delivery time is 10 working days. Database is delivered as CD-ROMs.

Database includes two-channel recordings: one close-talking microphone and one hands-free microphone channel. Database includes label files. Database is not validated.

The price will be 1.000 EURO, excluding VAT (22%).

Database is made available to be used for research purposes only.

5. HTK SCRIPTS AND RECOGNISER CONFIGURATION

The recogniser configuration will be as described in documents [1], [2]. The HTK scripts will be provided in the database CD-ROM.

6. REFERENCES

[1] Preliminary description of a subset of SpeechDat-Car to be used in ETSI STQ WI008 front-end standardisation, STQ Aurora DSR Working Group input document AU/189/99

[2] Experimental framework for the Performance Evaluation of Distributed Speech Recognition Front-ends, Motorola, 30th Sept 1998, Aurora project input document.