Danish SpeechDat-Car Digits Database for ETSI STQ Aurora Advanced DSR

Børge Lindberg,
CPK, Aalborg University
January, 2001
Version 1.0

1. Introduction

This document describes the subset of the Danish SpeechDat-Car database provided to the Aurora consortium for the evaluation of noise robust DSR front-ends. The selection of utterances as well as some baseline results are reported. Similar results are available for Finnish, Italian and Spanish.

2. Selection of utterances

The utterances were selected according to the following criteria:

a) Only isolated and connected digit utterances were selected and only from the sets B1, I1, I2, I3, I4, C1, C3 and C4
b) Utterances containing other words than the Danish digits were discarded
c) Utterances containing mispronounced or truncated digits were discarded
d) Utterances containing the noise labels [dit], [int], [sta], [spk] and [fil] were kept
e) Utterances with lengthening marks were kept (utterances containing remarkable lengthening of a sound (phoneme)); lengthening mark (“:”) however ignored
f) Utterances with long pause marks were kept (utterances with unnatural pauses); pause mark (“.”) however ignored
g) Speakers having less than five (out of a maximum of eight) utterances in each of the two recording conditions were removed
h) Male speakers were randomly removed until equal sex balance was achieved

As a result of this selection and exclusion procedure, 2457 files remained. These are recordings from 104 male and 104 female speakers over two channels (close talking (CT) and hands free (HF)); i.e. a total of 4914 speech utterance files.

The distribution over lists is as follows:

Lists	**B1**	**C1,C3,C4**	**I1,I2,I3,I4**	**Total**
Number of recordings	362	448	1647	2457

The distribution of noise conditions are as follows:

Condition	**# of recordings**	**Percentage (%)**
Quiet (STOP_MOTOR_RUNNING)	265	10.8
Low (TOWN_TRAFFIC + LOW_SPEED_ROUGH_ROAD)	1513	61.6
High (HIGH_SPEED_ROUGH_ROAD)	679	27.6
Number of recordings	2457	100.0

3. Baseline results

The utterance files were processed as follows:

a) conversion to signed integers
b) DC offset removal
c) removal of leading DTMF's (2.5 seconds were removed from all files)
d) for the case of 8 kHz data, the files were down sampled using the ITU-tool:
```
filter -down -HQ2 infile outfile
```
e) feature processing, using the standard feature processing tool distributed by the Aurora consortium (assuming 8 kHz sampling rate):
```
frontend -q -swap -F RAW -fs 8 -noc0
```

Training and testing were done for the following conditions:

Well matched (WM):

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

Mid-mismatch (MM):

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

High-mismatch (HM):

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

The following base line results were achieved (using the public domain version 3.0 of the HTK software tool):

Condition	Word Accuracy (after 24 iterations)
WM	77.8%
MM	47.4 %
HM	31.9 %
Average	52.3%

Table 1 Baseline results for the Danish SDC-Aurora database (8 kHz)

4. Acknowledgements

Ramalingam Hariharan, Nokia, Dusan Macho, UPC and Ulf Knoblich, Alcatel are acknowledged for their contributions to the scripts required for the experiments reported.