

Title: Spanish SDC-Aurora Database for ETSI STQ Aurora WI008 Advanced DSR Front-End Evaluation: Description and Baseline Results
Source: Dušan Macho, UPC
Date: December 13, 2000
Version: 1.0

1. Introduction

This document describes the subset of Spanish SDC-Aurora database provided to the Aurora consortium for the evaluation of noise robust DSR front-ends. Both, database statistics and baseline results are reported. Similar reports on Finish and Italian SDC-Aurora databases can be found in [2]-[4].

2. Database Description and Statistics

The following processing was applied to the original SDC (SpeechDat Car) data [1]:

- a) DC offset removal (unsigned to signed conversion)
- b) down-sampling from 16 to 8 kHz by using the ITU-T software tools library
- c) cutting off the speaker synchronization ring by an automated process

The following utterances from the original SDC database were included in Spanish SDC-Aurora subset:

- 1) 1 sequence of 10 isolated digits (b1)
- 2) 1 sheet number, 4+ digits (c1)
- 3) 1 credit card number, 16 digits (c3)
- 4) 1 PIN code, 6 digits (c4)
- 5) 4 utterances of isolated digit, 1 digit per utterance (i1, i2, i3, i4)

The two characters in parenthesis are used in filenames to identify each utterance (they are derived from the original SDC item identification).

Spanish SDC-Aurora database contains 4914 recordings (files) and more than 160 speakers. Recordings from close-talking microphone (channel 0 – ch0) and one of the hands-free microphones (channel 1 – ch1) are included.

Like in other SDC-Aurora databases, all files were divided into three noisy conditions – **quiet**, **low noisy** and **high noisy** – depending on the driving conditions. Table 1 shows number of files in each of these three conditions.

Noise conditions (DRIVING_CONDITIONS)	# of files	Percentage
quiet (STOP_MOTOR_RUNNING)	792	16.12%
low noisy (TOWN_TRAFFIC + LOW_SPEED_ROUGH_ROAD)	2422	49.28%
high noisy (HIGH_SPEED_GOOD_ROAD)	1700	34.60%
Total	4914	100%

Table 1 Number of files in each noisy condition and distribution of driving conditions into the three noisy conditions. Note that the same utterance recorded by close-talking and hands-free microphone is considered as two files.

In the following paragraphs, we describe the statistical properties of Spanish SDC-Aurora database in well-matched, medium mismatch and high mismatch experiments. For each experiment, noise condition distribution, gender distribution and number of repetitions per digit are reported.

2.1 Well-Matched (WM) Experiment

In WM experiment, 70% of all files (including both, close-talking and hands-free microphone versions) have been used for training and 30% for testing.

Noise condition	Train (ch0+ch1)		Test (ch0+ch1)	
	# of files	Percentage	# of files	Percentage
quiet	532	15.68%	260	17.08%
low	1668	49.17%	754	49.54%
high	1192	35.14%	508	33.38%
Total	3392	100%	1522	100%

Table 2 Noise condition distribution in WM experiment.

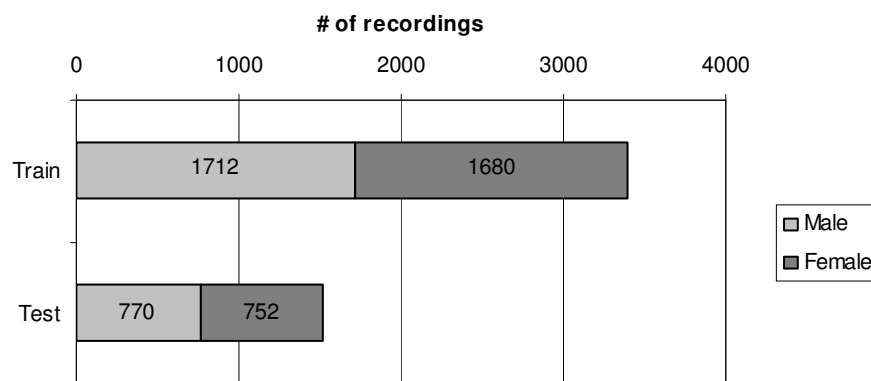


Figure 1 Gender distribution in WM experiment.

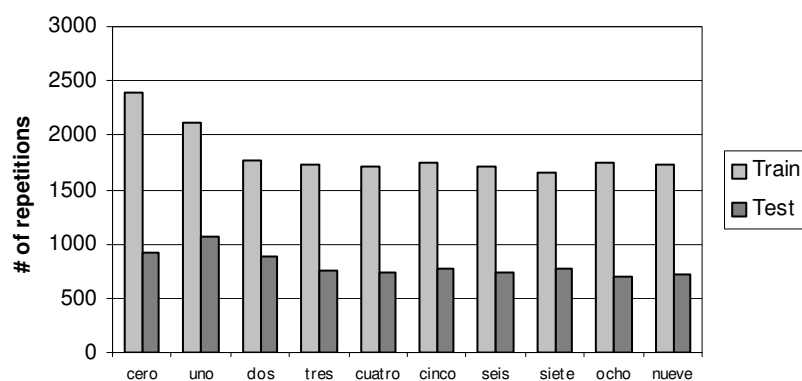


Figure 2 Number of repetitions per each digit in WM experiment.

2.2 Medium-Mismatch (MM) Experiment

In MM experiment, all hands-free microphone recordings from quiet and low noisy conditions were used for training and all hands-free microphone recordings from high noisy conditions were used for testing. By using this configuration, 65% of all data used in this experiment were employed for training and 35% for testing.

Noise condition	Train (ch1)		Test (ch1)	
	# of files	Percentage	# of files	Percentage
quiet	396	24.64%	0	0%
low	1211	75.36%	0	0%
high	0	0%	850	100%
Total	1607	100%	850	100%

Table 3 Noise condition distribution in MM experiment.

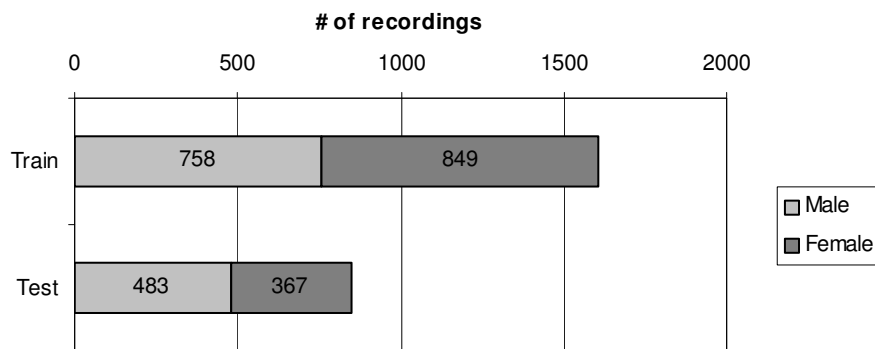


Figure 3 Gender distribution in MM experiment.

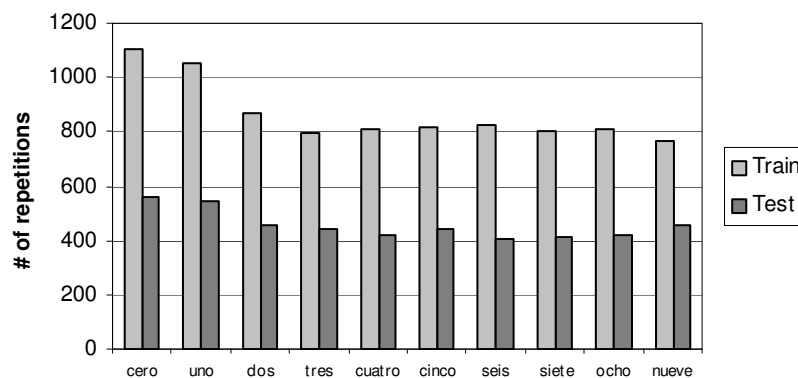


Figure 4 Number of repetitions per each digit in MM experiment.

2.3 High-Mismatch (HM) Experiment

In HM experiment, we used 70% of close-talking microphone recordings from all conditions for training and 30% of hands-free microphone recordings from low and high noisy conditions for testing.

Noise condition	Train (ch0)		Test (ch1)	
	# of files	Percentage	# of files	Percentage
quiet	266	15.68%	0	0%
low	834	49.17%	377	59.75%
high	596	35.14%	254	40.25%
Total	1696	100%	631	100%

Table 4 Noise condition distribution in HM experiment.

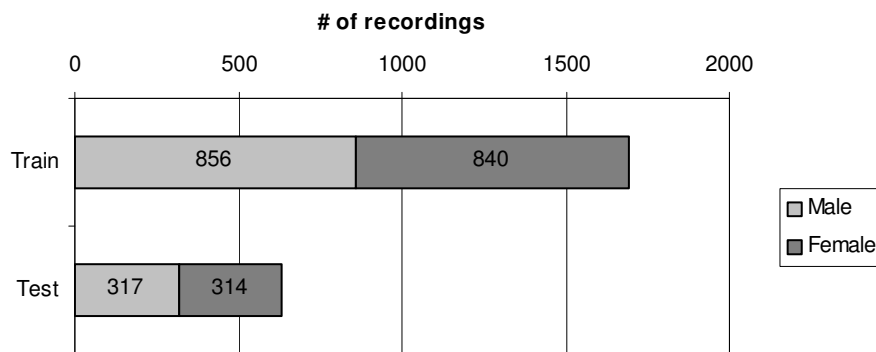


Figure 5 Gender distribution in HM experiment.

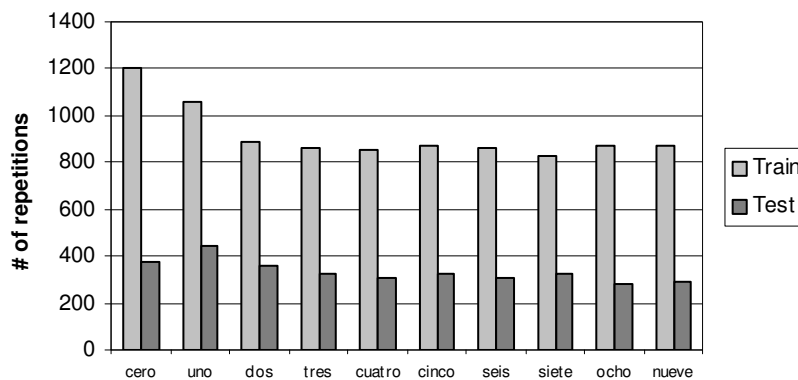


Figure 6 Number of repetitions per each digit in HM experiment.

3. Baseline Results

To generate the baseline results with Spanish SDC-Aurora database, the W007 standard front-end has been used and the back-end had the same setup as in [3] or [4]. Table 5 shows performances in each of the three mismatch conditions by using signals with both 8kHz and 16kHz sampling frequencies (F_s). Tests have been performed on Sun4u Sparc, SunOS 5.8 (Solaris 8) operating system.

Condition	Word Accuracy	
	$F_s=8\text{kHz}$	$F_s=16\text{kHz}$
WM	86.85%	87.44%
MM	73.74%	67.31%
HM	42.23%	35.25%
Average	67.61%	63.33%

Table 5 Baseline recognition percentages in the all three mismatch conditions for Spanish SDC-Aurora database.

References

- [1] A. Moreno, SpeechDat Car: Spanish - Selected Subset, Ver.1, Universitat Politècnica de Catalunya, Spain, April 2000.
- [2] Availability of Finnish SpeechDat-Car Database for ETSI STQ WI008 Front-End Standardization, STQ Aurora DSR Working Group input document AU/217/99.
- [3] Baseline Results for Subset of SpeechDat Car Finnish Database for ETSI STQ WI008 Advanced Front-End Evaluation, STQ Aurora DSR Working Group input document AU/225/00.
- [4] Description and Baseline Results for the Subset of the SpeechDat-Car Italian Database used for ETSI STQ Aurora WI008 Advanced DSR Front-End Evaluation, STQ Aurora DSR Working Group input document AU/237/00.