Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 3159-0040
Vol. 2 Issue 10, October - 2015

Evaluating ASR Systems Under Noisy Condition Using Aurora Database

Saimir Tola[1)]
Department of Mathematical Engineering
Polytechnic University of Tirana, Albania
E-mail:saimir_tola@yahoo.com

Ligor Nikolla[2)]
Department of Mathematical Engineering
Polytechnic University of Tirana, Albania

***Abstract*—This paper shows/describes the Aurora database and the evaluation of the Automatic speech recognition (ASR) in noisy condition. This database will be used for the evaluation of a complete recognition system. In this experiment we have added 8 different real-world noise over a range of signal to noise rotations. This database is public so the researchers can use freely this data for they work.**

Keywords—Automatic Speech recognition

Introduction

We define the robustness of a recognition system by two features:

- to handle the presence of a background noise
- to cope with the distortion by the frequency characteristic of the transmission channel

so to make a compare for the we need to create training tests. We use Noisex-92 database [1]. A male and a female are putted under a recorder in noisy conditions. The vocabulary used by them is American English.

The noisy database and the training and the test sets can determine the performance of a recognition system. Using the artificially distorted TIDigits data the Aurora, evaluation will be based on recognition experiments. The noise is taken from a noisy car environment. The data of speechDat-car [2] are going to be used.

Noisy speech database

The data used are taken from a male and a female adult. The 20 KHz data have been downsampled at 8 KHz with a low-pass filter extracting the spectrum from 0 to 4 KHz. After this we add noise to the clean data.

We add noise to the filtered TIDigits , to determine the speech energy we apply the ITU[3]. After that we calculate the noise energy assuming that the duration of the noise is larger than the duration of the signal.

The noise signals used for the background signal is:

1. Suburban train
2. Crowd of people
3. Car
4. Exhibition hall
5. Restaurant
6. Street
7. Airport
8. Train station

The noise signals are added to the TIDigits at : 20 dB, 15 dB, 10 dB, 5 dB, 0 dB, -5 dB.

Definition of training and tests sets

We have conducted two type of experiments :

1. On clean data only
2. Training on multi-condition data

In the first case to evaluate the performance of the system is easier and the recognition performance is better. In the multi-condition case in the first mode 84440 utterances are selected from the training part of the TIDigits containing the recording of 20 subsets with 422 utterances in each subset. The 20 subsets present 4 different noise scenario at 5 different SNRs. The noises are : train, crowd of people, car, exhibition hall.

We conduct three tests :

Test A

This test contains 28028 utterances. It contains the same noises for the multi condition training which lead to a high match of training and test data.

Test B

We conduct the same experiment but we use the different noises (Restaurant, street, airport, train station).

Test C

it contain two of the 4 subsets with 1001 utterances in each. This set is intended to show the influence of a recognition performance when a different frequency characteristic is presented at the input of the recognizers.

The word accuracy is listed in table 1 for test set A in multi condition training. We see that the performance detoriate for decreasing SNR. A measurement performance for the hole test is shown below. This average performance between 0 and 20 dB takes a value of 87.81% for test set A.

The results for test set B is listed in table 2. The average performance is 86.27 % for the SNR between 0 and 20 dB.

For the test C the results are listed in table 3. The average word accuracy is 83.77%.

Table 1

SNR/dB	Subway	Bubble	Car	Exhibition	average
clean	98.68	98.52	98.39	98.49	98.52
20	97.61	97.73	98.03	97.41	97.69
15	69.47	97.04	97.61	96.67	96.94
10	94.44	95.28	95.74	94.11	94.89
5	88.36	87.55	87.80	87.60	87.82
0	66.90	62.15	53.44	64.36	61.71
-5	26.13	27.18	20.58	24.34	24.55
Average between 0 and 20 dB	88.75	87.95	86.52	88.03	87.81

table 2

SNR/dB	Restaurant	Street	Airport	Train-station	average
Clean	98.68	98.52	98.39	98.49	98.52
20	96.87	97.58	97.44	97.01	97.22
15	95.30	96.31	96.12	95.53	95.81
10	91.96	94.35	93.29	92.87	93.11
5	83.54	85.61	86.25	83.52	84.73
0	59.29	61.34	65.11	56.12	60.46
-5	25.51	27.60	29.41	21.07	25.89
Average between 0 and 20 dB	85.39	87.03	87.64	85.01	86.27

Table 3

SNR/dB	Subway	Street	average
Clean	98.50	98.58	98.54
20	97.30	96.55	96.92
15	96.35	95.53	95.94
10	93.34	92.50	92.92
5	82.41	82.53	82.47
0	46.82	54.44	50.63
-5	18.91	24.24	21.57
Average between 0 and 20 dB	83.24	84.31	83.77

Table 4

SNR/dB	Subway	Bubble	Car	Exhibition	average
clean	98.93	99.00	98.98	99.20	99.02
20	97.05	90.15	97.41	96.39	95.25
15	93.49	73.76	90.04	92.04	87.33
10	78.72	49.43	67.01	75.66	67.70
5	52.16	26.81	34.09	44.83	39.47
0	26.01	9.28	14.46	18.05	16.95
-5	11.18	1.57	9.39	9.60	7.93
Average between 0 and 20 dB	69.48	49.88	60.60	65.39	61.34

In table 4,5,6 the training recognizer is on clear data only.

Table 5

SNR/dB	Restaurant	Street	Airport	Train-station	average
clean	98.93	99.00	98.96	99.20	99.02
20	89.99	95.74	90.64	94.72	92.77
15	**76.24**	**88.45**	**77.01**	**83.65**	**81.33**
10	54.77	67.11	53.86	60.29	59.00
5	31.01	38.45	30.33	27.92	31.92
0	26.01	9.28	14.46	18.05	13.69
-5	10.96	10.46	8.23	8.45	7.65
Average between 0 and 20 dB	52.59	61.51	53.25	55.63	55.74

Table 6

SNR/dB	Subway	street	average
clean	99.14	98.97	99.05
20	93.46	95.13	94.29
15	86.77	88.91	87.8/4
10	73.90	74.43	74.16
5	51.27	49.21	50.24
0	25.42	22.91	24.16
-5	11.82	11.15	11.48
Average between 0 and 20 dB	66.16	66.11	66.14

References

[1] A. Varga, H.J.M Steeneken, ‘’assessment for automatic speech recognition: NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems’’, Speech communication Vol 12 No 3 pp.247-252 1993

[2] http ://www.speechdat.org/SP-CAR

[3] ITU recommendation P.56, ‘’Objective measurement of active speech level’’, Mar. 1993