**59.Spoken Arabic Digit**

1. 数据库网址

https://archive.ics.uci.edu/ml/datasets/Spoken+Arabic+Digit

2. 数据库描述

【1.[数据集名称]数据集由[机构名或人名]采集；】The data used in our experiments were collected by the Laboratory of Automatic and Signals, University of Badji-Mokhtar.【2.用于[什么实验目的]】We used to classify the spoken Arabic digit from males and females.【3】Dataset from 8800(10 digits x 10 repetitions x 88 speakers) time series of 13 Frequency Cepstral Coefficients (MFCCs) had taken from 44 males and 44 females Arabic native speakers between the ages 18 and 40 to represent ten spoken Arabic digit. Each line on the data base represents 13 MFCCs coefficients in the increasing order separated by spaces. This corresponds to one analysis frame. The 13 Mel Frequency Cepstral Coefficients (MFCCs) are computed with the following conditions; Sampling rate is 11025 Hz, 16 bits; Window applied: hamming, Filter pre-emphasized: 1-0.97Z^(-1).【4】The database has 8800 samples, respectively belong to males with 4400 samples and females with 4400 samples. The categories can be shown in Table 1.

Table 1 Category Distribution of dataset [根据数据库绘制]

|  |  |  |  |
| --- | --- | --- | --- |
| Categories | Traning | Testing | Total Number of Samples |
| Males | 3300 | 1100 | 4400 |
| females | 3300 | 1100 | 4400 |
| Total number of samples in total | 6600 | 2200 | 8800 |

3. 精简描述

The Data in our experiment were collected by the Laboratory of Automatic and Signals, University of Badji-Mokhtar. The dataset includes have 8800 samples, which used to classify the spoken Arabic digit from males and females. Through which, we divided the dataset into two part, training data set with 6600 samples and forecasting data set with 2200 samples.