Speech Validation System using MFCC and DTW in Python.

عبدالرحمن مصطفى موسى محمد رباح، احمد صبري علي، عبدالله عادل ابراهيم، عمرو محمد السيد

4th Year, Electronics and Communication Department, Faculty of Engineering, Cairo University

Giza, 12613, Egypt

Abdelrahman.Rabah00@eng-st.cu.edu.eg

ahmed.omer99@eng-st.cu.edu.eg

Abdullah.Fattah99@eng-st.cu.edu.eg

amr.hassan99@eng-st.cu.edu.eg

***Abstract*—** In this paper describe an implementation of speechs validation to validate the pronunciation of phonemes in Arabic language. The words dataset is split into pairs, each pair has 2 words which almost have the same pronunciation. To extract the audio files features, Mel frequency cepstral coefficients (MFCC) is used to convert audio in time domain into frequency domain. A reference speaker is used to calculate the Distance Time Warping (DTW) between this reference and other test speakers to differentiate between each word in the same pairs if they belong to the pair whether was correct or wrong, or they don't belong to this pair.

Keywords— Speech Validation, MFCC, DTW

I. INTRODUCTION

It's desired in this project to create a Speech Validation system which is able to predict if a pronounced word belongs to a specific pair or not, and if it belongs to this pair, the system should be able to validate this word if it's correct or not.

II. METHODOLOGY

A. System's Model

I. In a quiet noise-free environment, 123 words are recorded following a set of specifications; mono stereo with 16 kHz sampling frequency processed using Cool Edit Pro.
II. Then running a segmentation tool to name the audio files as desired.
III. Running the MFCC.exe which will be explained later.
IV. Then download the rest of test speakers and reference speakers which are valid and follow the specifications.
V. Creating a working directory for the data.
VI. Creating a function collectWords which reads and combines the mat extension file along with their names to avoid any confusion.
VII. Scaling the data so the upcoming DTW values won't be too big. Then comparing each speaker to the references using the last word to decide which of the 3 references will be this speaker's reference.
VIII. To calculate the DTW, 2 functions were created to make it easier to search for desired words. find_Pair_Word_Gender and find_Group_Student, their names are self-explanatory.
IX. calc_dtw function returns a 1*n vector (1D), this n is the number of records for the same word.
X. Removing the non-valid records which doesn't follow the specifications, like the MFCC length is double or half the reference's MFCC length.
XI. Removing the outliers which are a data value that is numerically distant from other data.
XII. Then taking 20 random samples to calculate the threshold.
XIII. Judgement of these records

The following Figure 1 is modeling the system.

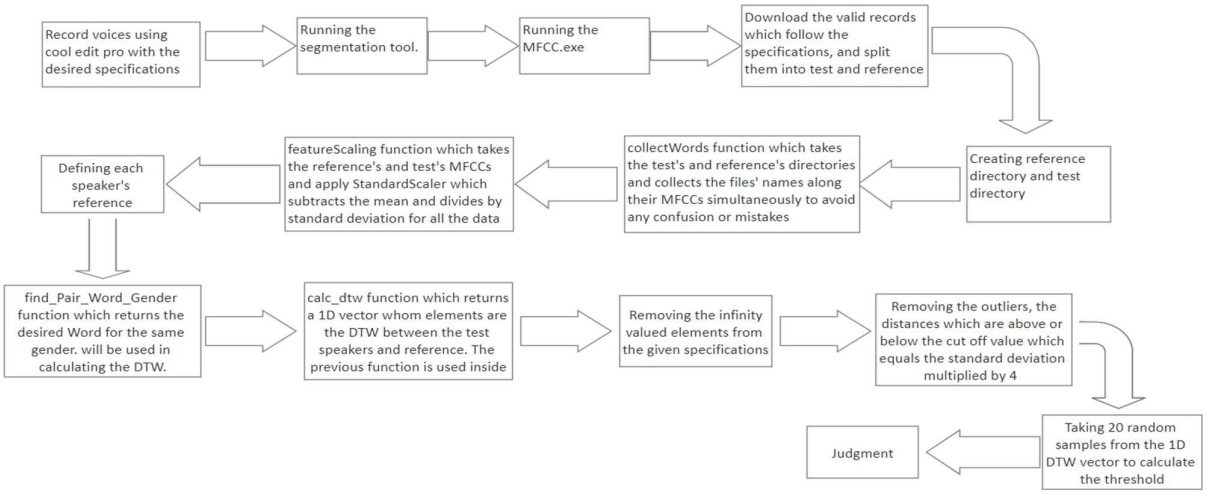

Figure 1 System Model

B. *Feature extraction using Mel Frequency Cepstrum Coefficient (MFCC) method:*

MFCC is a method of feature extraction of voice signals. Our ear has cochlea which basically has more filters at low frequency and very few filters at higher frequency. Convert time domain signals into frequency domain signal by mimicking cochlea function using Mel filters. Feature extraction is the process of determining a value or vector that can be used as an object or an individual identity. MFCC is the most used method in various areas of voice processing field, because it is considered quite good in representing signal. Feature is the coefficient of cepstral, the coefficient of cepstral used still considering the perception of the human hearing system. The workings of MFCC are based on the different frequencies that can be captured by the human ear to represent the sound signals as humans represent them. The process extracting MFCCs for a given voice sample is shown in the following Figure. [1]

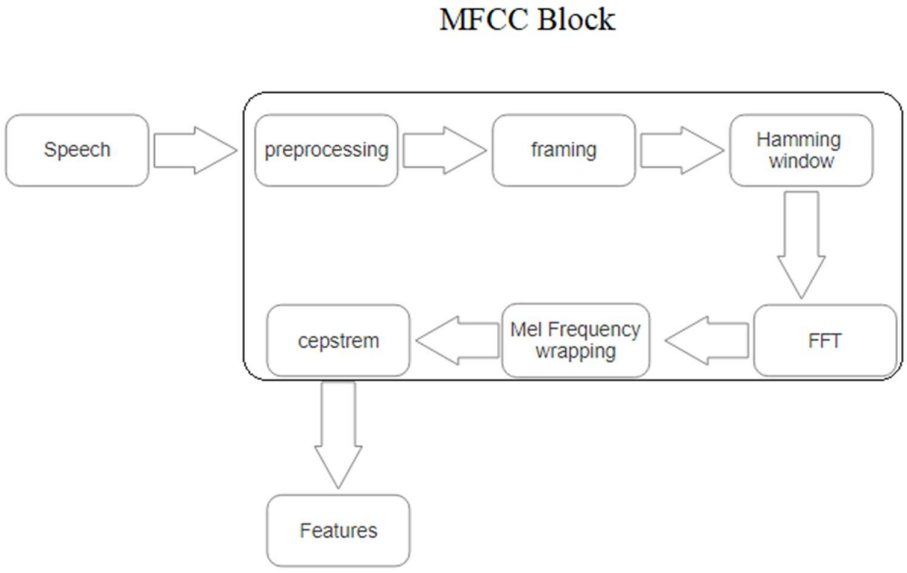

Figure 2 MFCC Block

Before features extraction, the voice sample must be converted by Analog to Digital Converter (ADC), followed by Pre-Emphasis and Filtering. It is important to have a sufficient sampling rate to avoid aliasing. According to the Nyquist's sampling rate, the minimum sampling frequency of a signal with maximum frequency *f* should be *2f* Hz. Pre-emphasis stage increases the magnitude of higher frequency with respect to lower frequencies. FIR filter, used for this purpose and its corresponding discrete output is given in equation 1 and equation 2 respectively.[1]

1. $F(z) = 1 - kz^{-1}\ 0 < k < 1$ (1)
2. $y[n] = s[n] - k.s[n-1]\ 0 < k < 1$ (2)

Where, y[n] is the output and s[n] the signal input of the FIR filter.

The Noise-gate is applied to the pre-emphasized voice sample to remove the amplitudes (noise) below a particular threshold value.

After a noise-gate is applied the voice sample can be aligned to start from zero on the time axis. This is called zero alignment. This can reduce some of the workload for the pattern matching process later in the program since the voice samples are much closer to each other than they otherwise would be. All the above signal operations are carried out prior to MFCC extraction. They avoid interaction of noise with significant features.

Voice sample is a time variant. It must be framed with frame length within range of 20ms to 30ms. The frame length should not be too short such that we can obtain reliable spectral estimate for each frame. On the other hand, it should not be to too long such that under a particular frame voice sample is time invariant. Now each frame is multiplied with hamming window. The Hamming window function is:

3. $W[n] = 0.54 - 0.46\cos\left[\frac{2\pi n}{N-1}\right]$
4. $Y[n] = X[n] \times W[n]$

Where, N = number of samples in each frame
Y[n] = Output signal
X (n) = input signal
$W[n]$ = nth coefficient of hamming window

Fast Fourier Transform (FFT) is applied to each frame which transforms signal to frequency domain. The feature matching algorithm cannot discern the difference between two closely spaced frequencies. For this reason, we take clumps of spectral bins and sum them up to get an idea of how much energy exists in various frequency regions. This can be performed by multiplying each frame with Triangular MEL Filter banks shown in Figure 3. [1]

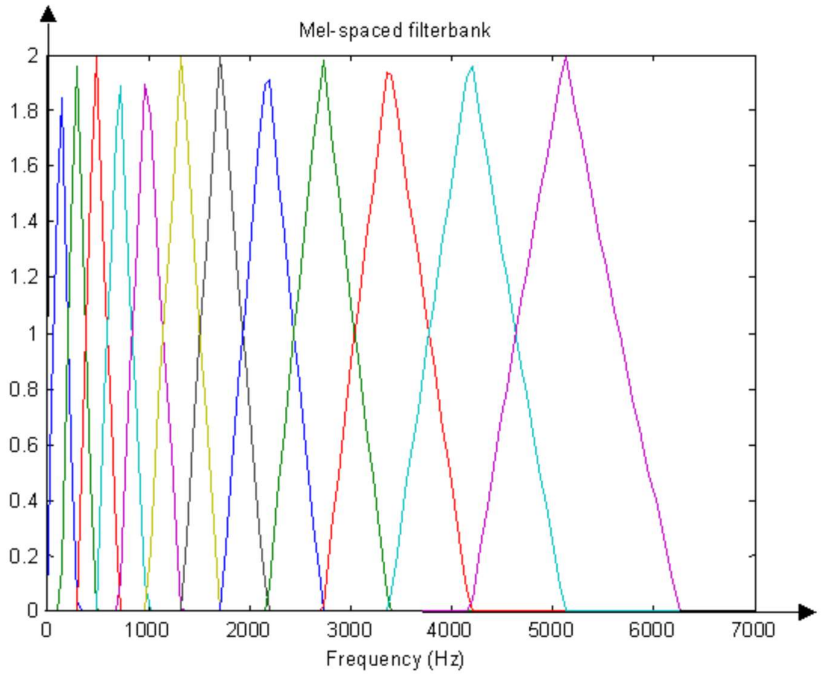

Figure 3 Mel-spaced filter banks

The first filter is very narrow and gives us indication of how much energy exists near *zero* hertz. As the frequency gets higher our filters get wider as we become less concerned about variations. The equation for calculating MEL for a given frequency is shown next:

$$5.. \quad F\,MEL = 2595 \times \log_{10}[1 + f/700]$$

We are only interested in roughly how much energy occurs at each spot. Here a set of 39 triangular filters are taken. To calculate filter bank energies, we multiply each filter bank with the energy spectrum, and then add up the coefficients. Once this is performed, we are left with 26 numbers that give us an indication of how much energy was in each filter bank. Logarithm for these 39 energy values is taken following by Discrete Cosine Transform (DCT). DCT is calculated using the equation shown [1]:

$$6.. \quad C_n = \sum_{k=1}^{k} (log\ Sk)\ [n(k - \tfrac{1}{2})\tfrac{\pi}{k}]$$

Where, n = 1, 2…K
Sk = FFT coefficients

The sample is currently in Frequency Domain (FD) after applying FFT. Converting it back to Time Domain (TD) using MEL filter and Discrete Cosine Transform (DCT) as shown in figure 4.

Figure 4 From FD to TD

Then Delta and Delta^2 coefficients are calculated for each frame. The first order derivative is called delta coefficient and the second order derivative is called delta^2 coefficient. The *nth* Delta feature and Delta-Delta feature is then defined by equation 7 and 8 as shown next:

$$7. \quad \Delta f_k[n] = f_{k+M}[n] - f_{k+M}[n]$$
$$8. \quad \Delta f_k^2[n] = \Delta f_{k+M}[n] - \Delta f_{k+M}[n]$$

Where, M typically is 2-3 frames. The differentiation is done for each feature vector separately.

C. *Distance Time Warping (DTW):*

In this stage, the features of word calculated in previous step are compared with reference audios. DTW algorithm is implemented to calculate least distance between features of words uttered from test speakers and the reference speaker. Corresponding to least value among calculated scores, the word is detected. DTW finds the optimal alignment between two times series if one time series may be "warped" non-linearly by stretching or shrinking it along its time axis. The extent of matching between two time series is measured in terms of distance factor. Dynamic time wrapping for two voice samples is illustrated in Figure 5.

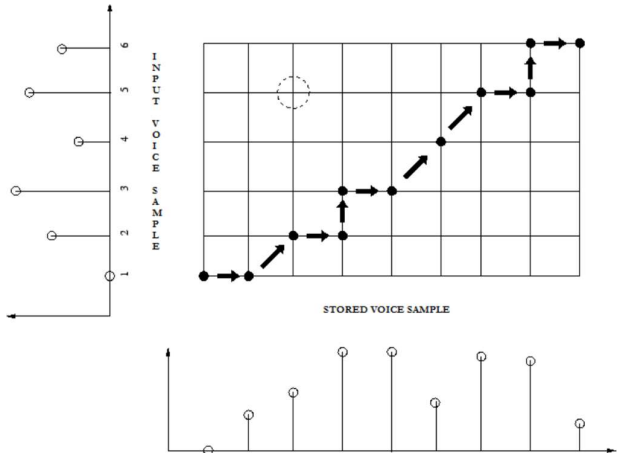

Figure 5 DTW of 2 audio samples

A matrix of order *n by m* is created whose *(i, j)* element is distance *d (ai, bj)* between points *ai* and *bj* of two times sequences. Euclidean computation is used to measure distance between features of input sample and the reference. Then, distance is measured by equation 9:

9. $D(i,j) = \min [D(i-1, j-1), D(i-1, j), D(i, j-1)] + d(i, j)$

The references distance to the same reference for the same word equals 0. The word (i) is considered pronounced correct if it has small distance to the reference's word (i) and smaller distance than word (j) which is the other word in the same pair and doesn't bypass a defined threshold.[1]

D. *Threshold (Th):*

Since DTW returns 1D vector, we have 2 words in each pair, so we have 2 1D vectors. We can calculate the users' mean DTW for given word i, and j for the same pair. The threshold is calculated by taking 20 random samples from the 2 1D vectors. And the Threshold will be a point in between. After summing the 2 means, their mean is also taken and will be returned as a Threshold.

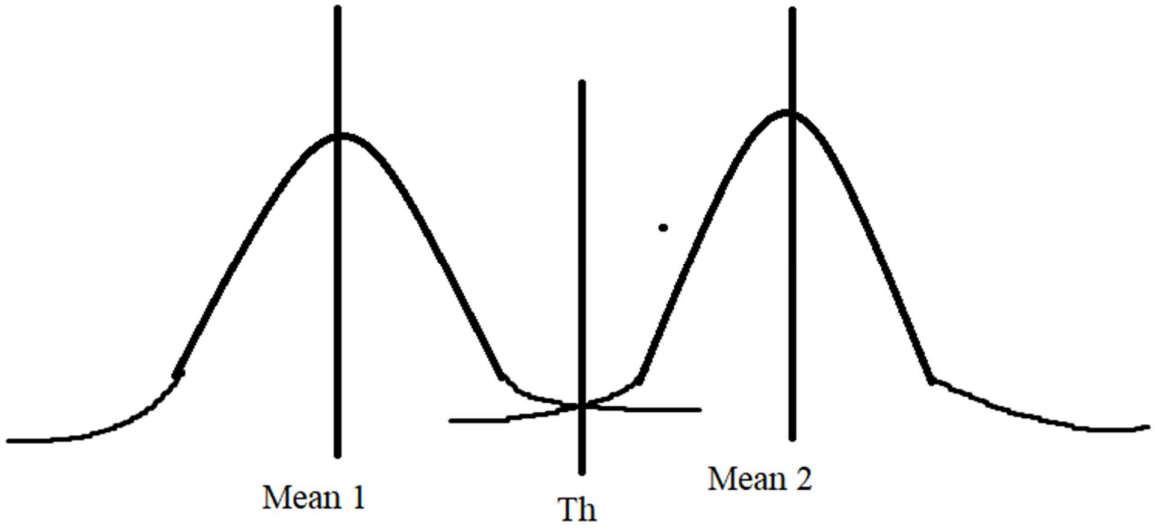

Figure 6 Threshold

III. RESULTS AND DISCUSSION:

There are 14 groups, each group for minimum should have 3 students, each student for each gender should record 123 audio samples following the set of rules. So, for males/females/children there should be at least 42 audio/word with correct formatting. Unfortunately, there are some groups which didn't record their audios at all or didn't follow the set of rules of recording. Some users as well doesn't have a 123 recording they have more or less records. There are also some students who didn't pronounce the words correctly or didn't pre-process their samples and that led to infinity values or outliers or incorrect spectrum. And finally, there are some users who took other users' samples.

The more we remove the outliers, the less the other columns values appear, the better the threshold.

The system works with word interface more than being a user interface. It take about an hour and half to run the system.

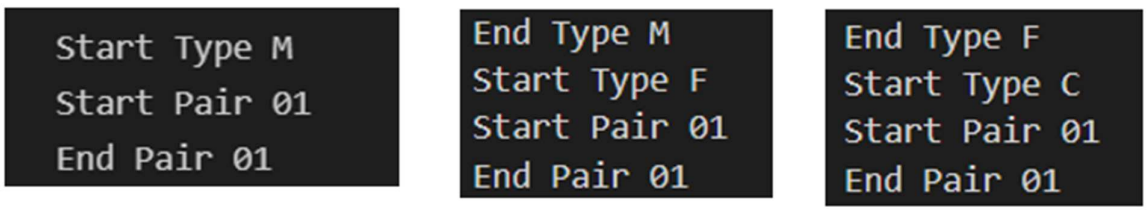

Figure 6 Shows Processing Data in the System

Table 1 Shows Male Statistics

pair	Word_Number	Words	type	total	other	other_word	correct	wrong
1	1	ذاب	M	47	5	23	19	28
1	2	ثاب	M	46	2	20	24	22
2	1	ثائر	M	45	2	17	26	19
2	2	سائر	M	48	6	16	26	22
3	1	ثمين	M	49	6	17	26	23
3	2	سمين	M	49	2	18	29	20
4	1	ثناء	M	46	8	18	20	26
4	2	سناء	M	49	3	14	32	17
5	1	نذير	M	49	7	15	27	22
5	2	نظير	M	49	1	21	27	22
6	1	نفذ	M	45	0	10	35	10
6	2	نفث	M	46	2	27	17	29
7	1	ظليل	M	47	2	20	25	22
7	2	ذليل	M	49	4	22	23	26
8	1	نذر	M	45	1	20	24	21
8	2	نظر	M	46	2	19	25	21
9	1	محظور	M	48	3	9	36	12
9	2	محذور	M	47	4	22	21	26
10	1	تين	M	47	0	24	23	24
10	2	طين	M	47	1	21	25	22
11	1	تابع	M	46	3	20	23	23
11	2	طابع	M	47	6	19	22	25
12	1	امات	M	49	5	22	22	27
12	2	اماط	M	50	7	19	24	26
13	1	طاب	M	49	3	23	23	26
13	2	تاب	M	47	7	19	21	26
14	1	امطار	M	46	2	24	20	26
14	2	امتار	M	48	1	15	32	16
15	1	دل	M	49	7	17	25	24
15	2	ضل	M	49	6	19	24	25
16	1	درب	M	39	4	8	27	12
16	2	ضرب	M	40	9	14	17	23
17	1	عد	M	48	3	17	28	20
17	2	عض	M	49	4	25	20	29
18	1	صامد	M	50	3	12	35	15
18	2	صامت	M	47	6	17	24	23
19	1	دلال	M	47	6	23	18	29
19	2	ضلال	M	48	2	16	30	18
20	1	فائده	M	46	7	7	32	14
20	2	فائضه	M	48	10	19	19	29
21	1	افاد	M	48	5	15	28	20
21	2	افاض	M	49	3	26	20	29
22	1	ضن	M	48	6	15	27	21
22	2	ظن	M	49	0	14	35	14
23	1	مضار	M	49	5	24	20	29
23	2	مدار	M	48	5	15	28	20

24	1	سوره	M	48	6	20	22	26
24	2	صوره	M	44	6	21	17	27
25	1	عسير	M	50	4	17	29	21
25	2	عصير	M	50	1	29	20	30
26	1	لمس	M	47	2	17	28	19
26	2	لمز	M	48	2	26	20	28
27	1	نسل	M	48	7	12	29	19
27	2	نزل	M	49	5	15	29	20
28	1	مسح	M	47	5	15	27	20
28	2	مسخ	M	45	4	22	19	26
29	1	صعد	M	47	5	17	25	22
29	2	سعد	M	47	1	17	29	18
30	1	صرير	M	46	1	20	25	21
30	2	سرير	M	48	8	21	19	29
31	1	مصيره	M	47	5	22	20	27
31	2	مسيره	M	46	1	19	26	20
32	1	صفر	M	48	4	28	16	32
32	2	سفر	M	48	3	15	30	18
33	1	ذكي	M	48	4	22	22	26
33	2	زكي	M	33	1	10	22	11
34	1	ذل	M	47	0	26	21	26
34	2	زل	M	50	0	25	25	25
35	1	زفر	M	46	0	15	31	15
35	2	ظفر	M	49	7	26	16	33
36	1	غمز	M	47	1	15	31	16
36	2	غمس	M	48	6	18	24	24
37	1	خير	M	47	4	8	35	12
37	2	غير	M	49	1	22	26	23
38	1	غير	M	47	6	18	23	24
38	2	خيل	M	48	4	7	37	11
39	1	خائب	M	50	1	15	34	16
39	2	غائب	M	49	5	28	16	33
40	1	قلب	M	48	3	19	26	22
40	2	كلب	M	47	0	22	25	22
41	1	تقدير	M	48	3	19	26	22
41	2	تكدير	M	46	7	17	22	24
42	1	قال	M	47	2	14	31	16
42	2	كال	M	47	11	13	23	24
43	1	خائن	M	47	2	6	39	8
43	2	كائن	M	48	4	6	38	10
44	1	خان	M	47	0	11	36	11
44	2	كان	M	47	0	11	36	11
45	1	خبير	M	50	10	2	38	12
45	2	كبير	M	46	4	18	24	22
46	1	خامل	M	49	6	7	36	13
46	2	كامل	M	47	4	8	35	12
47	1	مسك	M	46	2	11	33	13
47	2	مسخ	M	45	1	16	28	17

48	1	وعد	M	48	3	16	29	19
48	2	وأد	M	48	8	24	16	32
49	1	علم	M	49	8	13	28	21
49	2	الم	M	49	5	2	42	7
50	1	متعلم	M	46	1	13	32	14
50	2	متألم	M	49	7	26	16	33
51	1	عصابات	M	47	5	19	23	24
51	2	اصابات	M	48	1	24	23	25
52	1	حمزه	M	48	8	15	25	23
52	2	همزه	M	47	3	21	23	24
53	1	حاله	M	48	4	16	28	20
53	2	هاله	M	48	2	15	31	17
54	1	اشباح	M	48	4	15	29	19
54	2	اشباه	M	48	8	15	25	23
55	1	حرم	M	46	4	19	23	23
55	2	هرم	M	47	4	20	23	24
56	1	حامل	M	47	2	18	27	20
56	2	خامل	M	48	2	22	24	24
57	1	حامل	M	47	6	16	25	22
57	2	هامل	M	46	3	16	27	19
58	1	هزم	M	45	5	17	23	22
58	2	حزم	M	46	1	20	25	21
59	1	هان	M	48	1	18	29	19
59	2	حان	M	48	1	15	32	16
60	1	هوي	M	47	1	24	22	25
60	2	حوي	M	47	2	15	30	17
61	1	هروب	M	46	5	23	18	28
61	2	حروب	M	47	4	18	25	22
Summation		الحمد لله		5769	463	2140	3166	2603

Table 2 Female Statistics

pair	Word_Number	Words	type	total	other	other_word	correct	wrong
1	1	ذاب	F	4	1	16	23	17
1	2	ثاب	F	0	4	18	19	22
2	1	ثائر	F	41	1	17	23	18
2	2	سائر	F	41	5	19	19	24
3	1	ثمين	F	43	5	20	16	25
3	2	سمين	F	41	3	21	19	24
4	1	ثناء	F	43	4	19	19	23
4	2	سناء	F	42	4	16	22	20
5	1	نذير	F	42	0	19	23	19
5	2	نظير	F	42	0	15	26	15
6	1	نفذ	F	41	4	12	26	16
6	2	نفث	F	42	5	15	20	20
7	1	ظليل	F	40	3	19	22	22
7	2	ذليل	F	44	5	20	19	25
8	1	نذر	F	44	6	17	23	23
8	2	نظر	F	46	1	24	19	25
9	1	محظور	F	44	5	20	17	25
9	2	محذور	F	42	3	20	19	23
10	1	تين	F	42	3	14	27	17
10	2	طين	F	44	7	17	22	24
11	1	تابع	F	46	5	17	23	22
11	2	طابع	F	45	4	23	19	27
12	1	امات	F	46	6	19	20	25
12	2	اماط	F	45	0	18	25	18
13	1	طاب	F	43	5	18	21	23
13	2	تاب	F	44	4	19	21	23
14	1	امطار	F	44	3	21	21	24
14	2	امتار	F	45	3	18	23	21
15	1	دل	F	44	3	14	26	17
15	2	ضل	F	43	1	22	19	23
16	1	درب	F	42	3	21	18	24
16	2	ضرب	F	42	2	29	12	31
17	1	عد	F	43	3	20	18	23
17	2	عض	F	41	5	17	19	22
18	1	صامد	F	41	1	16	29	17
18	2	صامت	F	46	10	12	25	22
19	1	دلال	F	47	5	14	27	19
19	2	ضلال	F	46	2	23	17	25
20	1	فائده	F	42	1	23	21	24
20	2	فائضه	F	45	4	18	23	22
21	1	افاد	F	45	2	18	24	20
21	2	افاض	F	44	6	16	22	22
22	1	ضن	F	44	1	10	31	11
22	2	ظن	F	42	0	17	24	17
23	1	مضار	F	41	3	19	22	22

23	2	مدار	F	44	0	21	23	21
24	1	سوره	F	44	6	15	23	21
24	2	صوره	F	44	8	17	19	25
25	1	عسير	F	44	4	19	20	23
25	2	عصير	F	43	1	16	26	17
26	1	لمس	F	43	3	19	22	22
26	2	لمز	F	44	6	16	19	22
27	1	نسل	F	41	0	14	22	14
27	2	نزل	F	36	6	14	17	20
28	1	مسح	F	37	1	15	23	16
28	2	مسخ	F	39	6	17	18	23
29	1	صعد	F	41	4	10	29	14
29	2	سعد	F	43	5	18	21	23
30	1	صرير	F	44	7	18	21	25
30	2	سرير	F	46	5	19	21	24
31	1	مصيره	F	45	1	18	23	19
31	2	مسيره	F	42	1	17	24	18
32	1	صفر	F	42	4	23	13	27
32	2	سفر	F	40	0	15	25	15
33	1	ذكي	F	40	7	15	20	22
33	2	زكي	F	42	2	13	26	15
34	1	ذل	F	41	8	13	22	21
34	2	زل	F	43	6	20	17	26
35	1	زفر	F	43	5	19	20	24
35	2	ظفر	F	44	2	19	22	21
36	1	غمز	F	43	3	15	23	18
36	2	غمس	F	41	4	18	19	22
37	1	خير	F	41	5	16	23	21
37	2	غير	F	44	3	11	31	14
38	1	غير	F	45	0	25	18	25
38	2	خيل	F	43	4	10	31	14
39	1	خائب	F	45	4	10	33	14
39	2	غائب	F	47	7	15	23	22
40	1	قلب	F	45	5	22	17	27
40	2	كلب	F	44	5	15	25	20
41	1	تقدير	F	45	2	20	21	22
41	2	تكدير	F	43	3	22	19	25
42	1	قال	F	44	1	16	27	17
42	2	كال	F	44	6	17	21	23
43	1	خائن	F	44	3	11	31	14
43	2	كائن	F	45	2	12	31	14
44	1	خان	F	45	0	4	39	4
44	2	كان	F	43	2	14	28	16
45	1	خبير	F	44	0	16	29	16
45	2	كبير	F	45	0	8	35	8
46	1	خامل	F	43	3	9	34	12
46	2	كامل	F	46	5	25	15	30
47	1	مسك	F	45	6	13	19	19

47	2	مسخ	F	38	6	6	28	12
48	1	وعد	F	40	2	9	32	11
48	2	وأد	F	43	3	21	19	24
49	1	علم	F	43	6	4	34	10
49	2	الم	F	44	7	5	32	12
50	1	متعلم	F	44	3	11	28	14
50	2	متألم	F	42	11	16	15	27
51	1	عصابات	F	42	5	19	21	24
51	2	اصابات	F	45	4	23	19	27
52	1	حمزه	F	46	4	10	29	14
52	2	همزه	F	43	5	18	21	23
53	1	حاله	F	44	2	17	24	19
53	2	هاله	F	43	5	17	20	22
54	1	اشباح	F	42	1	15	26	16
54	2	اشباه	F	42	1	23	18	24
55	1	حرم	F	42	6	14	25	20
55	2	هرم	F	45	6	18	22	24
56	1	حامل	F	46	4	19	23	23
56	2	خامل	F	46	3	15	29	18
57	1	حامل	F	47	5	16	24	21
57	2	هامل	F	45	6	19	20	25
58	1	هزم	F	45	3	13	29	16
58	2	حزم	F	45	9	14	23	23
59	1	هان	F	46	0	17	27	17
59	2	حان	F	44	8	14	22	22
60	1	هوي	F	44	0	7	36	7
60	2	حوي	F	43	4	21	18	25
61	1	هروب	F	43	8	12	24	20
61	2	حروب	F	44	3	23	18	26
Summation		الحمد لله		44 5279	452	2020	2807	2472

Table 3 Children Statistics

pair	Word_Number	Words	type	total	other	other_word	correct	wrong
1	1	ذاب	C	34	6	9	19	15
1	2	ثاب	C	35	4	17	14	21
2	1	ثائر	C	36	5	16	15	21
2	2	سائر	C	35	4	13	18	17
3	1	ثمين	C	34	5	15	14	20
3	2	سمين	C	33	0	13	20	13
4	1	ثناء	C	40	5	15	20	20
4	2	سناء	C	39	6	17	16	23
5	1	نذير	C	42	4	20	18	24
5	2	نظير	C	41	2	18	21	20
6	1	نفذ	C	34	7	14	13	21
6	2	نفث	C	34	4	10	20	14
7	1	ظليل	C	40	1	17	22	18
7	2	ذليل	C	39	1	24	14	25
8	1	نذر	C	37	6	17	14	23
8	2	نظر	C	34	0	13	21	13
9	1	محظور	C	34	4	11	19	15
9	2	محذور	C	36	5	17	14	22
10	1	تين	C	39	4	16	19	20
10	2	طين	C	38	2	19	17	21
11	1	تابع	C	38	7	14	17	21
11	2	طابع	C	38	0	14	24	14
12	1	امات	C	29	5	10	14	15
12	2	اماط	C	29	1	12	16	13
13	1	طاب	C	33	4	10	19	14
13	2	تاب	C	33	4	14	15	18
14	1	امطار	C	39	0	21	18	21
14	2	امتار	C	39	3	18	18	21
15	1	دل	C	32	4	10	18	14
15	2	ضل	C	33	5	15	13	20
16	1	درب	C	32	0	15	17	15
16	2	ضرب	C	40	0	20	20	20
17	1	عد	C	33	4	16	13	20
17	2	عض	C	31	2	10	19	12
18	1	صامد	C	34	2	13	19	15
18	2	صامت	C	39	3	10	26	13
19	1	دلال	C	37	3	14	20	17
19	2	ضلال	C	36	2	18	16	20
20	1	فائده	C	38	3	15	20	18
20	2	فائضه	C	37	5	10	22	15
21	1	افاد	C	32	5	12	15	17
21	2	افاض	C	30	4	11	15	15
22	1	ضن	C	37	4	17	16	21
22	2	ظن	C	38	5	11	22	16

23	1	مضار	C	39	4	13	22	17
23	2	مدار	C	41	4	19	18	23
24	1	سوره	C	38	1	12	25	13
24	2	صوره	C	39	7	18	14	25
25	1	عسير	C	39	4	15	20	19
25	2	عصير	C	38	5	18	15	23
26	1	لمس	C	36	1	16	19	17
26	2	لمز	C	36	5	14	17	19
27	1	نسل	C	31	3	16	12	19
27	2	نزل	C	32	1	13	18	14
28	1	مسح	C	31	3	10	18	13
28	2	مسخ	C	31	3	14	14	17
29	1	صعد	C	39	5	12	22	17
29	2	سعد	C	39	8	15	16	23
30	1	صرير	C	37	1	13	23	14
30	2	سرير	C	37	2	19	16	21
31	1	مصيره	C	35	4	16	15	20
31	2	مسيره	C	37	3	19	15	22
32	1	صفر	C	34	3	10	21	13
32	2	سفر	C	37	4	14	19	18
33	1	ذكي	C	35	1	9	25	10
33	2	زكي	C	35	3	19	13	22
34	1	ذل	C	35	3	9	23	12
34	2	زل	C	35	5	16	14	21
35	1	زفر	C	38	4	16	18	20
35	2	ظفر	C	37	2	16	19	18
36	1	غمز	C	40	3	12	25	15
36	2	غمس	C	39	4	26	9	30
37	1	خير	C	39	1	10	28	11
37	2	غير	C	39	5	23	11	28
38	1	غير	C	34	1	12	21	13
38	2	خيل	C	38	3	17	18	20
39	1	خائب	C	39	5	14	20	19
39	2	غائب	C	39	2	19	18	21
40	1	قلب	C	35	1	19	15	20
40	2	كلب	C	34	3	11	20	14
41	1	تقدير	C	38	5	12	21	17
41	2	تكدير	C	38	4	17	17	21
42	1	قال	C	36	2	18	16	20
42	2	كال	C	31	3	8	20	11
43	1	خائن	C	32	2	11	19	13
43	2	كائن	C	33	6	6	21	12
44	1	خان	C	32	2	8	22	10
44	2	كان	C	32	1	16	15	17
45	1	خبير	C	38	0	15	23	15
45	2	كبير	C	38	2	13	23	15
46	1	خامل	C	37	0	6	31	6
46	2	كامل	C	32	4	9	19	13

47	1	مسك	C	33	0	12	21	12
47	2	مسخ	C	32	2	7	23	9
48	1	وعد	C	34	2	5	27	7
48	2	وأد	C	33	4	14	15	18
49	1	علم	C	38	0	9	29	9
49	2	الم	C	37	0	9	28	9
50	1	متعلم	C	32	2	11	19	13
50	2	متألم	C	30	2	14	14	16
51	1	عصابات	C	30	0	17	13	17
51	2	اصابات	C	36	1	18	17	19
52	1	حمزه	C	39	1	14	24	15
52	2	همزه	C	36	1	20	15	21
53	1	حاله	C	37	6	10	21	16
53	2	هاله	C	38	6	14	18	20
54	1	اشباح	C	39	1	20	18	21
54	2	اشباه	C	39	2	16	21	18
55	1	حرم	C	31	3	10	18	13
55	2	هرم	C	31	2	16	13	18
56	1	حامل	C	25	2	15	8	17
56	2	خامل	C	38	9	19	10	28
57	1	حامل	C	39	1	17	21	18
57	2	هامل	C	36	3	18	15	21
58	1	هزم	C	31	3	12	16	15
58	2	حزم	C	30	4	12	14	16
59	1	هان	C	35	5	15	15	20
59	2	حان	C	33	7	10	16	17
60	1	هوي	C	39	6	18	15	24
60	2	حوي	C	39	5	13	21	18
61	1	هروب	C	32	3	19	10	22
61	2	حروب	C	33	0	14	19	14
Summation		الحمد لله		4340	381	1742	2217	2123

Children samples are low compared to Male and Female and that proves lower accuracy in pronunciation.
A lot of data is missing due to bad records
Accuracy is overall higher than 50%
There are 3 columns dedicated to Correct percentage, error percentage and each word threshold but above tables can't handle this number of columns. Sample of the 3 columns in Figure 7
Samples of Graphical User Interface in Figure 8.

pair	Word_Nur	type	total	other	other_wor	correct	wrong	Threshold	Correct_Percenta	Wrong_Percentage
1	1	M	47	5	23	19	28	31732.82607	0.404255319	0.595744681
1	2	M	46	2	20	24	22	34956.3373	0.52173913	0.47826087
2	1	M	45	2	17	26	19	9472.018319	0.577777778	0.422222222
2	2	M	48	6	16	26	22	8572.988203	0.541666667	0.458333333
3	1	M	49	6	17	26	23	4976.725376	0.530612245	0.469387755

Figure 7

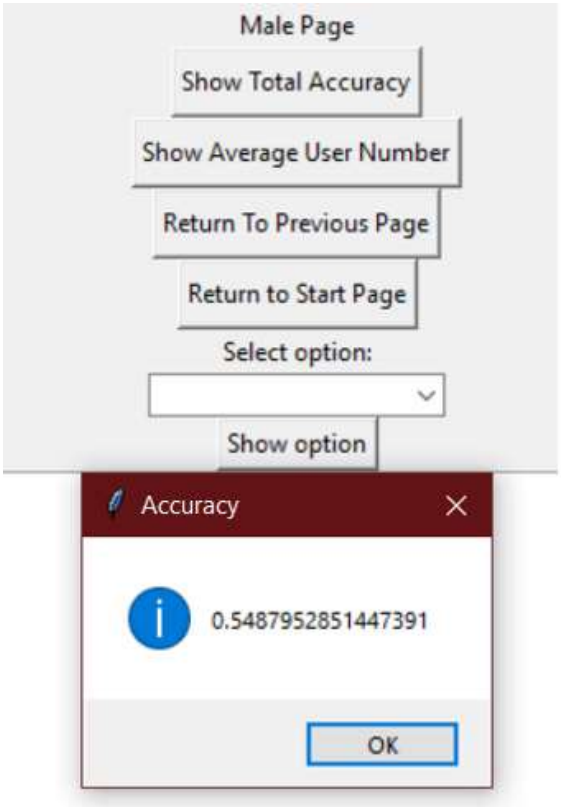

Figure 7 User Interface Sample

IV. CONCLUSIONS

As we can see our speech validation system using MFCC and DTW in Python environment. System is trained by setting references for these 123 words and comparing them to all the other speakers by DTW. This system provided higher than 50% total accuracy and for some words it reached higher than 70%. We can improve the recordings by gathering a team who will be responsible for collecting the audio samples in a noise free environment from other teams with correct pronunciation and pre-process the audio correctly so we can avoid or reduce any kind of mistakes

REFERENCES

[1] Ramesh Babu.N, *Speech recognition using MFCC and DTW*, 2014.