	Academic Year:	2022/2023	Term:	Fall 2022	
	Course Code:	ELC 4011	Course Title:	Introduction to Digital Signal Processing	

Cairo University
Faculty of Engineering
Electronics and Communications Engineering Department – 4th Year

Speech Recognition Project

Presented in partial fulfillment of the requirements for ELC - 4011

Under the supervision of

Dr. Mohsen Rashwan

Presented by:

Std. No:	B.N.	Sec.	NAME
1	46	2	عبدالرحمن محمد نورالدين
2	5	3	على ماهر عبدالسلام
3	7	3	عمار اشرف عمار
4	29	3	محمد احمد العدوى
5	35	4	مينا حنا فايز وهبة

DEC. 2022

Table of Contents

Table of Figures

Abstract

ASR (Automated speech recognition) is a technology that allows users to enter data into information systems by speaking rather than punching numbers into a keyboard. ASR is primarily used for providing information and forwarding phone calls. In recent years, ASR has grown in popularity among large corporation customer service departments. It is also used by some government agencies and other organizations. Basic ASR systems recognize single-word entries such as yes-or-no responses and spoken numerals.

Introduction

The field of automatic speech recognition has witnessed a number of significant advances in the past 5 – 10 years, spurred on by advances in signal processing, algorithms, computational architectures, and hardware. These advances include the widespread adoption of a statistical pattern recognition paradigm, a data-driven approach which makes use of a rich set of speech utterances from a large population of speakers, the use of stochastic acoustic and language modeling, and the use of dynamic programming-based search methods. [1]

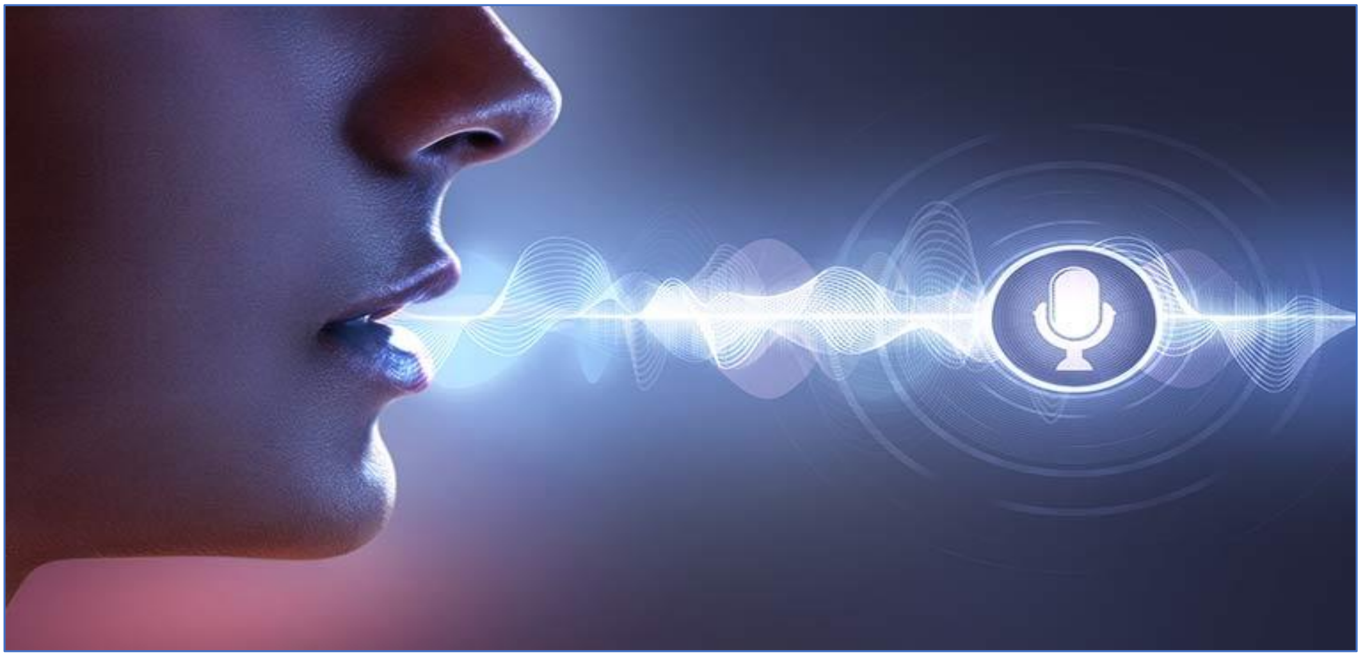


Figure 1 - Speech Signals

Requirements

We need to recognize words spoken by 3 genders “Male – Female – Child” through extracting features from the speech signal and collect the data set and store them to be as data base information Then select one from the males and females and child to be the Golden Data which will be a reference of to compare all of the records with it, speech can be identified. The process of speech recognition initially starts with digitalization and quantization of human sound to allow digital signal processing on the speech. After digitalization the feature extraction comes, which can be done by using MFCC (Mel-frequency cepstral coefficients) used for feature extraction from the speech signal and using dynamic time warping (DTW) to generate the distance and comparing the distance with the threshold of the word to decide if it’s correct or wrong.

Project Road Map

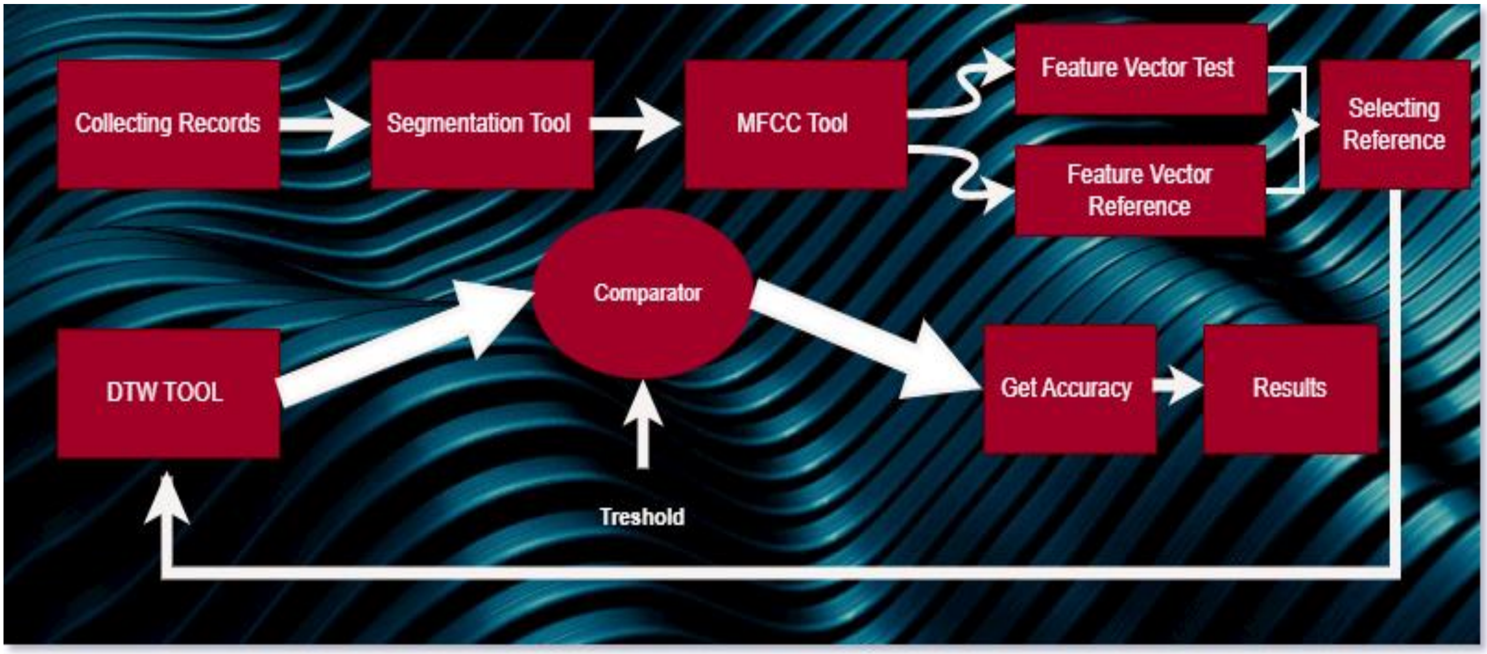


Figure 2 - Road Map Block Diagram

Collecting Records

Each member in the team is required to collect 6 records of some specific pairs of words, first 2 records for a male and 2 for a female and 2 for 2 children.

Segmentation Tool

In this phase, it's required to separate all the words from each record into a single file then rename each word record with a specific name. We used the MATLAB built-in function *detectSpeech* which detects boundaries of speech in an audio signal and returns the indexes of each word. [2]

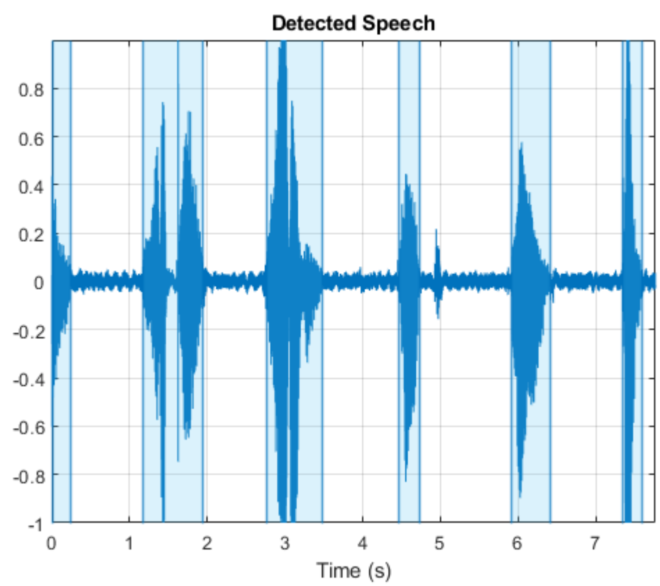


Figure 3 – DetectSpeech function

Outputs are a nicely arranged set of records.

Figure 4 - Segmentation output

Totally 47 records each one represents 1 word. For easier readability all naming followed a specific standard stating student number, gender of the recorder, his age, source of recording, specifying which word in which pair and finally if the recorder is chosen as test or reference.

MFCC Tool

It's stands for Mel-Frequency Spectral Coefficients. [3]

The first step is to apply a pre-emphasis filter on the signal to amplify the high frequencies. A pre-emphasis filter is useful in several ways:

1) Balance the frequency spectrum since high frequencies usually have smaller magnitudes compared to lower frequencies,
2) Avoid numerical problems during the Fourier transform operation and
3) May also improve the Signal-to-Noise Ratio (SNR).

After pre-emphasis, the signal should be split into short time frames. By performing the Fourier transform on this short time frame, we can get a good approximation of the signal's frequency contour by concatenating adjacent frames. Each set of Frames of a single word is analyzed into set of windows with standard of being 25ms and those windows are overlapped that helps characterize this word. Then we transform it into frequency domain by doing fft then we get the log scale of this frequency domain and return back to what's called the cepstrum to simple the analysis.

The final step in computing the filter bank is to apply a triangular filter (usually 40 filters) to the power spectrum to extract the frequency bands. The Mel scale is intended to mimic the non-linear human ear's perception of sound by being more discriminating at low frequencies and less discriminating at high frequencies.

After that we get DCT (discrete cosine transform) and that is because of the compression capabilities of the DCT as it concentrates most of its values on lower range thus allowing us to get only a part of that data thus helped us in compressing the coefficients. Finally, we subtract the average from the coefficients as it helps remove any dc value between two speakers as any speakers are in different conditions, different voice strength for example, another reason that is It turns out that filter bank coefficients computed in the previous step are highly correlated, which could be problematic in some machine learning algorithms. Therefore, we can apply Discrete Cosine Transform (DCT) to decorrelate the filter bank coefficients and yield a compressed representation of the filter banks.

To implement MFCC we used the MATLAB built-in function *mfcc* which extract the Mel-Frequency spectral coefficients, log energy, delta and delta-delta of an input audio signal. [4]

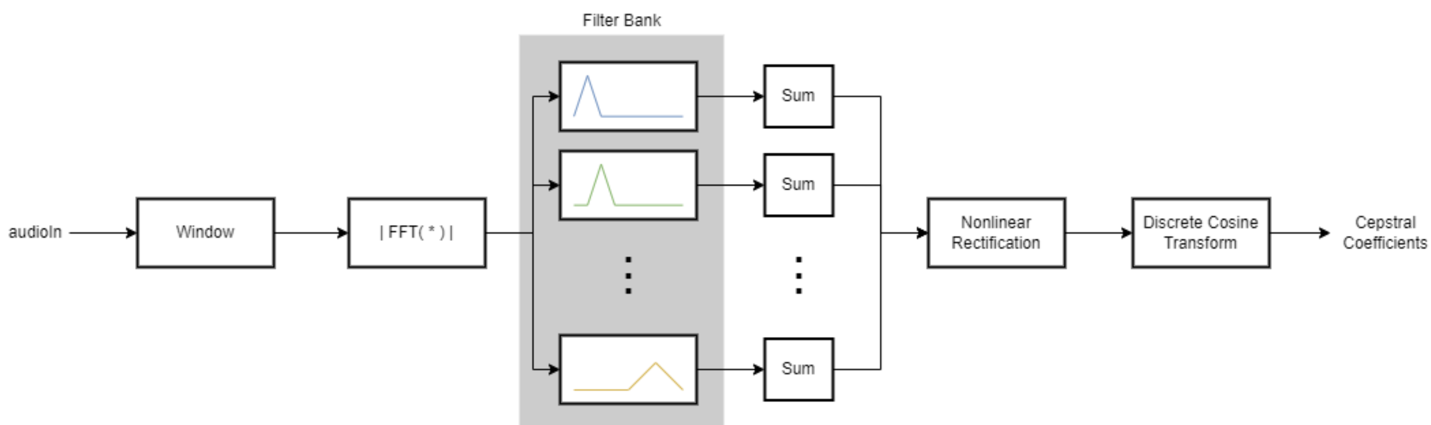


Figure 5 - MFCC Block Diagram

After calculating all mfcc coefficients from all tests and references, Saved all the mfcc outputs into matrix files and considered it as database to use it in later calculations.

Figure 6 - mfcc output

Dynamic Time Warping (DTW)

In time series analysis, dynamic time warping (DTW) is an algorithm for measuring similarity between two temporal sequences, which may vary in speed. For instance, similarities in walking could be detected using DTW, even if one person was walking faster than the other, or if there were accelerations and decelerations during the course of an observation. DTW has been applied to temporal sequences of video, audio, and graphics data. Indeed, any data that can be turned into a linear sequence can be analyzed with DTW.

The idea of comparing sequences of different lengths is to create one-to-many and many-to-one matches so that the overall distance between the two can be minimized.

The main application for us is speech recognition, to cope with different speaking speeds. Other applications include speaker recognition and online signature recognition. It can also be used in partial shape matching applications.

In general, DTW is a method that calculates an optimal match between two given sequences (e.g. time series) with certain restriction and rules which are:

1) Each index from sequence x must be matched with one or more indices from sequence y and vice versa.
2) The first index from sequence x must be matched with at least the first index of sequence y.
3) The last index from sequence x must be matched with at least the last index of sequence y.
4) The mappings of indices in sequence x to indices in sequence y must always be increasing.

We used the MATLAB built-in function *dtw* to implement this block. [5]

Saved the dtw scores of all tests in 3 folders for male, female and child. And for easier readability saved the outputs into excel sheets as following. [6]

the content of these sheets are the distance scores of each word inside each pair of test mfcc with the same corresponding word of the test. In addition, we calculated the distance between this word and the opposite word inside the same pair as we will need it in our approach of threshold calcualtion.

	A	B	C
1	Word	Same	Opposite
2	P1W1	238.053672	228.843511
3	P1W2	149.949221	187.994929
4	P2W1	105.131719	130.166223
5	P2W2	78.141243	92.7398819
6	P3W1	300.288752	309.876293
7	P3W2	244.702007	232.847317
8	P4W1	183.115974	244.776747
9	P4W2	158.253324	118.495592
10	P5W1	285.559079	318.403701
11	P5W2	191.074614	206.646363
12	P6W1	238.155001	322.816299
13	P6W2	242.291208	224.332784
14	P7W1	358.965327	348.436928
15	P7W2	213.21933	284.785635
16	P8W1	228.109191	241.030668
17	P8W2	287.587771	249.791057
18	P9W1	235.805627	308.588847
19	P9W2	207.990571	188.33483
20	P10W1	312.044143	304.811238
21	P10W2	271.333319	254.888909
22	P11W1	310.571717	312.124537
23	P11W2	327.727202	350.531639

	A	B	C
1	Word	Same	Opposite
2	P1W1	178.857454	166.495341
3	P1W2	205.085374	235.294869
4	P2W1	187.579642	266.481055
5	P2W2	412.045113	320.353403
6	P3W1	227.544388	246.200362
7	P3W2	230.925722	215.470922
8	P4W1	177.194194	200.170967
9	P4W2	193.315522	203.490101
10	P5W1	191.988707	243.863681
11	P5W2	265.560633	220.834232
12	P6W1	403.745849	388.416246
13	P6W2	419.055589	441.832269
14	P7W1	384.629057	393.103612
15	P7W2	352.806051	373.37702
16	P8W1	228.891601	324.561579
17	P8W2	288.879511	255.622351
18	P9W1	279.836184	240.334065
19	P9W2	209.2406	307.975456
20	P10W1	202.809751	231.875302
21	P10W2	250.102414	261.061729
22	P11W1	238.693588	280.423328
23	P11W2	326.359026	291.966625

	A	B	C
1	Word	Same	Opposite
2	P1W1	173.243506	196.367566
3	P1W2	202.073556	151.476796
4	P2W1	224.903468	318.308513
5	P2W2	267.002957	164.496474
6	P3W1	266.832543	261.120172
7	P3W2	294.933325	258.32768
8	P4W1	206.840498	229.726395
9	P4W2	264.194737	281.362444
10	P5W1	220.373208	274.030866
11	P5W2	274.119567	241.858106
12	P6W1	282.239016	297.37472
13	P6W2	262.92324	244.614799
14	P7W1	250.874778	295.547599
15	P7W2	276.588225	284.976907
16	P8W1	376.528614	474.799685
17	P8W2	386.743473	251.658657
18	P9W1	288.602338	225.91394
19	P9W2	300.067985	331.992888
20	P10W1	234.591765	256.394444
21	P10W2	279.894407	279.979416
22	P11W1	346.226949	341.644536
23	P11W2	407.427607	375.563087

	A	B	C
1	Word	Same	Opposite
2	P1W1	242.167076	232.483202
3	P1W2	250.499219	212.304914
4	P2W1	273.727624	300.530862
5	P2W2	286.648356	244.785448
6	P3W1	312.936124	342.123742
7	P3W2	339.707159	311.120218
8	P4W1	358.644633	399.473468
9	P4W2	441.321758	392.458214
10	P5W1	265.735607	293.244243
11	P5W2	290.701799	272.67585
12	P6W1	358.330204	357.347963
13	P6W2	365.982846	375.856616
14	P7W1	417.175422	356.285528
15	P7W2	477.439865	531.220255
16	P8W1	447.92685	473.097383
17	P8W2	400.634385	399.539625
18	P9W1	277.428627	280.433894
19	P9W2	305.392685	301.004823
20	P10W1	310.191131	362.639121
21	P10W2	385.879148	367.614905
22	P11W1	423.214785	475.001087
23	P11W2	400.644062	387.344821

Figure 7 - Examples of dtw output

These are some examples of our dtw results. The numbers ranges from 80~300 which is probably has to do with our approach of calculating the mfcc.

Threshold Calculation

An upper limit of accepted distance is needed to make the decision whether the speaker pronounced the word correctly or not. The approach we took to experimentally calculate this threshold is by running *dtw* between each word from test speaker with the corresponding same word of the reference. In addition, running cross *dtw* between each word and the opposite word inside the same pair.

Then, calculate the median between these 2 distance values. Finally, calculate the mean average through the results of the median of all tests for the same word. So, this approach could be more precise and accurate by adding more tests subjects.

	A	B	C
1	Pair	Word1	Word2
2	Pair 1	230.049834	221.258167
3	Pair 2	172.760346	162.283786
4	Pair 3	305.818663	287.039061
5	Pair 4	181.884647	216.42389
6	Pair 5	229.313604	226.157787
7	Pair 6	279.978566	288.082609
8	Pair 7	289.994188	266.151593
9	Pair 8	241.372003	210.803083
10	Pair 9	215.056761	237.754523
11	Pair 10	309.360623	276.128922
12	Pair 11	303.868561	339.664686
13	Pair 12	289.077318	282.993398
14	Pair 13	294.007626	262.87854
15	Pair 14	195.303352	214.503972
16	Pair 15	200.433031	221.468949
17	Pair 16	297.892353	265.564312
18	Pair 17	276.055358	276.794301
19	Pair 18	231.619199	210.034179
20	Pair 19	194.149325	189.064795
21	Pair 20	294.388984	273.105508
22	Pair 21	206.076751	162.135089
23	Pair 22	255.894896	243.796563
24	Pair 23	245.657712	261.237199

	A	B	C
1	Pair	Word1	Word2
2	Pair 1	215.182358	172.754303
3	Pair 2	227.361465	231.111755
4	Pair 3	249.361548	265.63058
5	Pair 4	230.407343	237.690055
6	Pair 5	233.186567	239.154635
7	Pair 6	363.355635	363.198617
8	Pair 7	288.205367	285.41191
9	Pair 8	291.998157	289.361076
10	Pair 9	269.232806	287.549908
11	Pair 10	253.051019	268.989379
12	Pair 11	306.929138	311.405177
13	Pair 12	246.988778	265.478344
14	Pair 13	254.986437	262.458072
15	Pair 14	251.64528	254.257971
16	Pair 15	251.832375	244.893169
17	Pair 16	313.11836	297.288164
18	Pair 17	301.147681	314.500398
19	Pair 18	275.569886	259.903786
20	Pair 19	219.201564	199.008505
21	Pair 20	200.675673	194.574386
22	Pair 21	232.254311	214.599446
23	Pair 22	263.384079	284.467252
24	Pair 23	271.213945	281.426042

	A	B	C
1	Pair	Word1	Word2
2	Pair 1	216.565067	222.452442
3	Pair 2	236.733803	218.589827
4	Pair 3	295.92426	297.691215
5	Pair 4	239.378945	270.251269
6	Pair 5	267.860557	241.292347
7	Pair 6	347.85833	340.063692
8	Pair 7	318.497464	352.96797
9	Pair 8	333.847984	305.872902
10	Pair 9	251.357426	236.516201
11	Pair 10	316.765294	285.023484
12	Pair 11	303.843788	292.223285
13	Pair 12	240.122291	253.815421
14	Pair 13	290.363972	271.100195
15	Pair 14	272.974058	266.357963
16	Pair 15	273.06128	228.449416
17	Pair 16	340.825194	307.573595
18	Pair 17	303.562055	253.280953
19	Pair 18	288.375814	257.788365
20	Pair 19	266.709063	269.25508
21	Pair 20	224.432773	215.84763
22	Pair 21	353.659441	329.19218
23	Pair 22	260.751673	265.128931
24	Pair 23	232.127801	250.751601

Figure 8 - Thresholds

These are the threshold values for all words for child, female and male tests in order.

Accuracy

After collecting the distance scores from *dtw*, compare each word score of the test with its distance threshold, if it is lower then the word is pronounced correctly. If the score is higher than the threshold then it is either pronounced the opposite word of the same pair or totally spelled wrong but in either case it is decided as wrong.

Save the scores for each test separately as a table stating which word score correct and wrong.

G03S1C09W_Results	G03S1C11W_Results
G03S1F22W_Results	G03S1F27W_Results
G03S1M22W_Results	G03S1M24W_Results
G03S2C07W_Results	G03S2C08W_Results
G03S2F39W_Results	G03S2F50M_Results
G03S2M22M_Results	G03S2M52W_Results
G03S3C06M_Results	G03S3C10M_Results
G03S3F23M_Results	G03S3M20M_Results
G03S4C09W_Results	G03S4F24W_Results
G03S4F50W_Results	G03S4M21W_Results
G03S4M22W_Results	G03S5C07W_Results

Figure 9 - test results

Here are some of the test scores

	A	B	C
1	Pair	Word 1	Word 2
2	Pair 1	Correct	Correct
3	Pair 2	Correct	Wrong
4	Pair 3	Correct	Wrong
5	Pair 4	Correct	Correct
6	Pair 5	Correct	Correct
7	Pair 6	Correct	Correct
8	Pair 7	Correct	Correct
9	Pair 8	Correct	Correct
10	Pair 9	Correct	Correct
11	Pair 10	Correct	Wrong
12	Pair 11	Correct	Correct
13	Pair 12	Correct	Correct
14	Pair 13	Correct	Correct
15	Pair 14	Correct	Correct
16	Pair 15	Correct	Correct
17	Pair 16	Correct	Correct
18	Pair 17	Correct	Correct
19	Pair 18	Wrong	Correct
20	Pair 19	Wrong	Correct
21	Pair 20	Correct	Correct
22	Pair 21	Correct	Correct
23	Pair 22	Correct	Wrong
24	Pair 23	Correct	Wrong

	A	B	C
1	Pair	Word 1	Word 2
2	Pair 1	Wrong	Correct
3	Pair 2	Wrong	Correct
4	Pair 3	Wrong	Correct
5	Pair 4	Wrong	Correct
6	Pair 5	Wrong	Wrong
7	Pair 6	Wrong	Wrong
8	Pair 7	Wrong	Correct
9	Pair 8	Wrong	Wrong
10	Pair 9	Correct	Correct
11	Pair 10	Correct	Wrong
12	Pair 11	Correct	Wrong
13	Pair 12	Wrong	Wrong
14	Pair 13	Wrong	Wrong
15	Pair 14	Correct	Correct
16	Pair 15	Wrong	Wrong
17	Pair 16	Wrong	Correct
18	Pair 17	Correct	Wrong
19	Pair 18	Wrong	Correct
20	Pair 19	Wrong	Wrong
21	Pair 20	Wrong	Wrong
22	Pair 21	Wrong	Correct
23	Pair 22	Wrong	Wrong
24	Pair 23	Wrong	Wrong

	A	B	C
1	Pair	Word 1	Word 2
2	Pair 1	Correct	Wrong
3	Pair 2	Correct	Wrong
4	Pair 3	Correct	Correct
5	Pair 4	Correct	Correct
6	Pair 5	Correct	Wrong
7	Pair 6	Wrong	Wrong
8	Pair 7	Wrong	Wrong
9	Pair 8	Correct	Correct
10	Pair 9	Wrong	Correct
11	Pair 10	Correct	Correct
12	Pair 11	Correct	Wrong
13	Pair 12	Wrong	Correct
14	Pair 13	Correct	Wrong
15	Pair 14	Correct	Correct
16	Pair 15	Correct	Correct
17	Pair 16	Correct	Correct
18	Pair 17	Wrong	Correct
19	Pair 18	Correct	Correct
20	Pair 19	Wrong	Correct
21	Pair 20	Correct	Correct
22	Pair 21	Correct	Correct
23	Pair 22	Correct	Correct
24	Pair 23	Correct	Wrong

Figure 10 - Score examples

Tabulate Results

Save all the results and scores of the test inside excel sheets tabulating needed details and information to make it easy to read. Stating details about each word alone which are helpful to find which words are the most difficult to pronouns or get which pair of words gets mixed up.

Child Accuracy Table

	A	B	C	D	E	F	G	H	I
1				Tot.words	judgment				
2	Pair	Word2	Word1	childs	other_	Word2_1	Word1_1	Correct	Wrong
3	1		ذا	9	3	3	3	3	6
4		ثا		9	2	7	0	7	2
5	2		ثا	9	4	0	5	5	4
6		سا		9	1	8	0	8	1
7	3		نذير	9	2	0	7	7	2
8		نظير		9	2	7	0	7	2
9	4		تين	9	3	1	5	5	4
10		طين		9	1	8	0	8	1
11	5		دل	9	3	1	5	5	4
12		ضل		9	2	7	0	7	2
13	6		عد	9	4	1	4	4	5
14		عض		9	3	5	1	5	4
15	7		صامد	9	3	2	4	4	5
16		صامت		9	3	6	0	6	3
17	8		ضن	9	2	0	7	7	2
18		ظن		9	3	5	1	5	4
19	9		لمس	9	2	2	5	5	4
20		لمز		9	2	7	0	7	2
21	10		مسح	9	4	0	5	5	4
22		مسخ		9	2	5	2	5	4
23	11		مصير	9	1	4	4	4	5
24		مسير		9	4	5	0	5	4
25			ذل	9	3	0	6	6	3
26	12	زل		9	2	4	3	4	5
27	13		زفر	9	4	0	5	5	4
28		ظفر		9	3	6	0	6	3
29	14		خير	9	3	0	6	6	3
30		غير		9	2	7	0	7	2
31	15		قلب	9	4	0	5	5	4
32		كلب		9	4	5	0	5	4
33	16		خائن	9	4	0	5	5	4
34		كائن		9	2	7	0	7	2
35	17		مسك	9	3	0	6	6	3
36		مسخ		9	2	7	0	7	2
37	18		علم	9	5	1	3	3	6
38		ألم		9	2	7	0	7	2
39	19		حمزة	9	2	1	6	6	3
40		همزة		9	3	6	0	6	3
41	20		حالة	9	4	0	5	5	4
42		هالة		9	2	6	1	6	3
43	21		حرم	9	3	0	6	6	3
44		هرم		9	1	8	0	8	1
45	22		حامل	9	3	0	6	6	3
46		خامل		9	3	5	1	5	4
47	23		هزم	9	3	0	6	6	3
48		حزم		9	3	5	1	5	4
49	Total			414				262	152
50	%							63.285	36.715

Figure 11 - Child Accuracy Table

After calculating all results for out 9 test children. In total the accuracy scored 63% correct words and 36% wrong. Most word spelled wrong is “ذا” which is hard to differentiate between it and the opposite word of the same pair.

Female Accuracy Table

	A	B	C	D	E	F	G	H	I
1				Tot.words			judgment		
2	Pair	Word2	Word1	Females	other_	Word2_1	Word1_1	Correct	Wrong
3	1		ذا	9	1	0	8	8	1
4		ثا		9	2	6	1	6	3
5	2		ثا	9	1	0	8	8	1
6		سا		9	3	2	4	2	7
7	3		نذير	9	1	1	7	7	2
8		نظير		9	4	5	0	5	4
9	4		تين	9	2	0	7	7	2
10		طين		9	3	6	0	6	3
11	5		دل	9	1	0	8	8	1
12		ضل		9	4	2	3	2	7
13	6		عد	9	5	0	4	4	5
14		عض		9	5	4	0	4	5
15	7		صامد	9	3	0	6	6	3
16		صامت		9	2	6	1	6	3
17	8		ضن	9	2	0	7	7	2
18		ظن		9	3	3	3	3	6
19	9		لمس	9	2	5	2	2	7
20		لمز		9	3	6	0	6	3
21	10		مسح	9	3	0	6	6	3
22		مسخ		9	4	4	1	4	5
23	11		مصير	9	3	0	6	6	3
24		مسير		9	2	4	3	4	5
25	12		ذل	9	3	0	6	6	3
26		زل		9	3	6	0	6	3

	A	B	C	D	E	F	G	H	I
26	12	زل		9	3	6	0	6	3
27	13		زفر	9	2	1	6	6	3
28		ظفر		9	4	4	1	4	5
29	14		خير	9	2	1	6	6	3
30		غير		9	2	7	0	7	2
31	15		قلب	9	2	1	6	6	3
32		كلب		9	1	8	0	8	1
33	16		خائن	9	1	1	7	7	2
34		كائن		9	2	7	0	7	2
35	17		مسك	9	3	3	3	3	6
36		مسخ		9	2	7	0	7	2
37	18		علم	9	1	6	2	2	7
38		ألم		9	1	8	0	8	1
39	19		حمزة	9	3	1	5	5	4
40		همزة		9	1	8	0	8	1
41	20		حالة	9	3	0	6	6	3
42		هالة		9	2	7	0	7	2
43	21		حرم	9	6	0	3	3	6
44		هرم		9	3	6	0	6	3
45	22		حامل	9	2	0	7	7	2
46		خامل		9	4	5	0	5	4
47	23		هزم	9	1	0	8	8	1
48		حزم		9	2	3	4	3	6
49	Total			414				258	156
50	%							62.3188	37.6812

Figure 12 - Female Accuracy Table

The accuracy score 62% correct and 37% wrong. Most words spelled wrong are “سا” which was mostly mistaken with the opposite word in the same pair “ثا”. In addition, the word “لمس” was mostly mistaken with its opposite word.

Male Accuracy Table

	A	B	C	D	E	F	G	H	I
1				Tot.words		judgment			
2	Pair	Word2	Word1	Males	other_	Word2_1	Word1_1	Correct	Wrong
3	1		ذا	9	3	0	6	6	3
4		ثا		9	4	4	1	4	5
5	2		ثا	9	4	0	5	5	4
6		سا		9	2	3	4	3	6
7	3		نذير	9	5	0	4	4	5
8		نظير		9	3	6	0	6	3
9	4		تين	9	2	0	7	7	2
10		طين		9	3	6	0	6	3
11	5		دل	9	3	0	6	6	3
12		ضل		9	2	4	3	4	5
13	6		عد	9	4	0	5	5	4
14		عض		9	4	5	0	5	4
15	7		صامد	9	2	1	6	6	3
16		صامت		9	2	7	0	7	2
17	8		ضن	9	3	0	6	6	3
18		ظن		9	3	5	1	5	4
19	9		لمس	9	4	1	4	4	5
20		لمز		9	4	5	0	5	4
21	10		مسح	9	3	0	6	6	3
22		مسخ		9	3	5	1	5	4
23	11		مصير	9	4	0	5	5	4
24		مسير		9	4	5	0	5	4
25	12		ذل	9	3	0	6	6	3
26		زل		9	3	5	1	5	4
27	13		زفر	9	3	0	6	6	3
28		ظفر		9	2	5	2	5	4
29	14		خير	9	3	0	6	6	3
30		غير		9	3	6	0	6	3
31	15		قلب	9	3	0	6	6	3
32		كلب		9	2	4	3	4	5
33	16		خائن	9	4	0	5	5	4
34		كائن		9	4	5	0	5	4
35	17		مسك	9	3	0	6	6	3
36		مسخ		9	2	5	2	5	4
37	18		علم	9	2	0	7	7	2
38		ألم		9	2	4	3	4	5
39	19		حمزة	9	4	0	5	5	4
40		همزة		9	2	5	2	5	4
41	20		حالة	9	4	0	5	5	4
42		هالة		9	4	4	1	4	5
43	21		حرم	9	3	2	4	4	5
44		هرم		9	4	5	0	5	4
45	22		حامل	9	3	0	6	6	3
46		خامل		9	3	6	0	6	3
47	23		هزم	9	5	0	4	4	5
48		حزم		9	3	6	0	6	3
49	Total			414				241	173
50	%							58.2126	41.7874

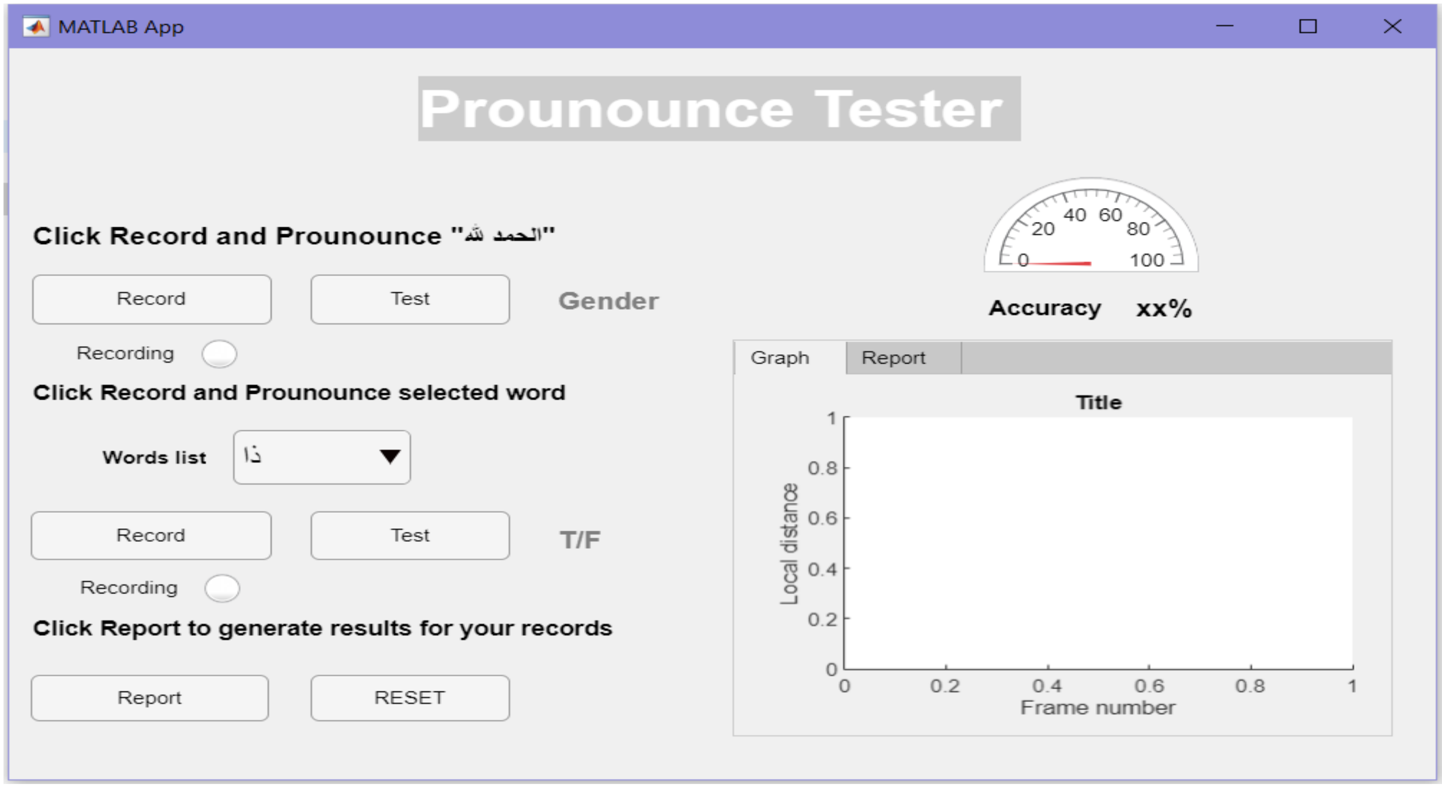
Figure 13 - Male Accuracy Table

The accuracy score 58% correct and 41% wrong. Words that are spelled wrong aren’t consistent on one specific word and it is mostly depends on the speaker so we got a wide variety of different results.

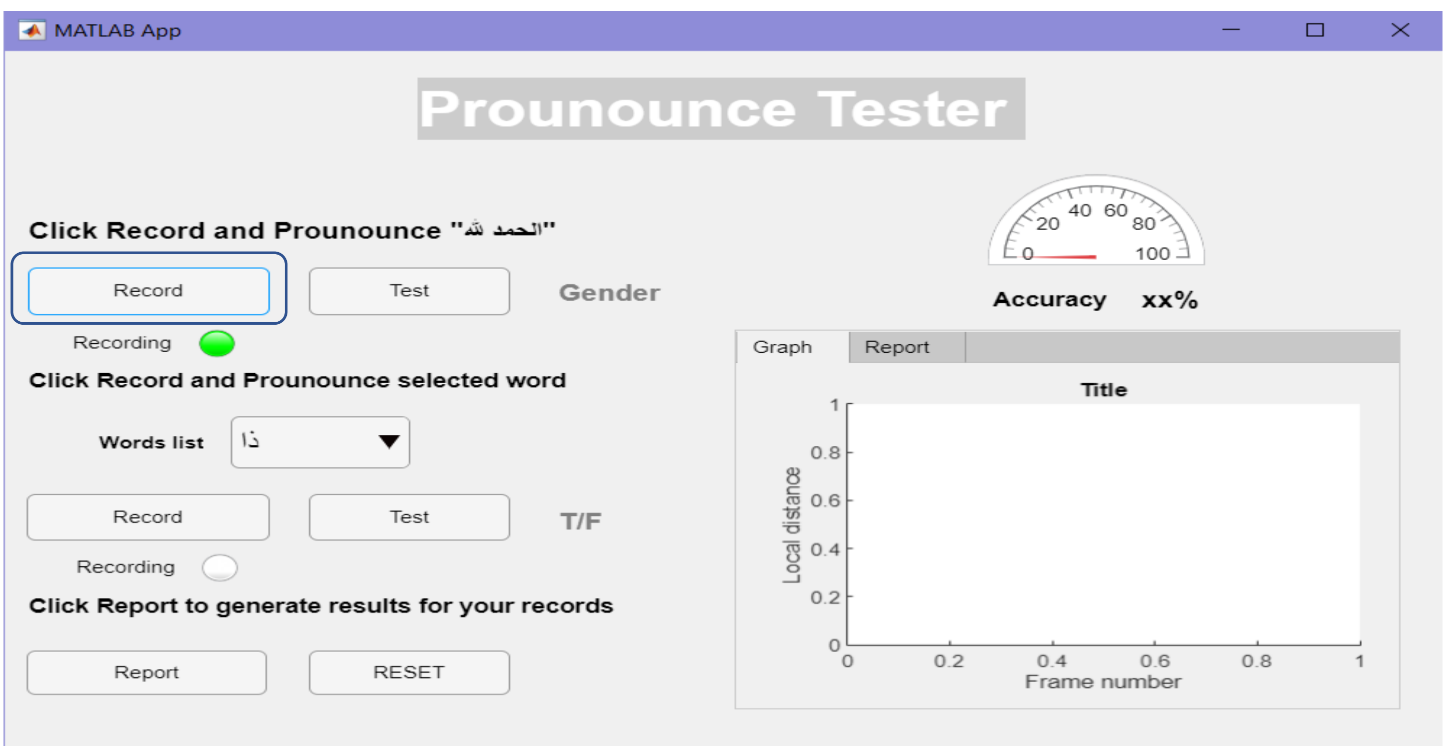
In general, due to children being very close range of age and similar voice it got a fine result comparing with the reference child speaker. For females we had a wider range of age which causes to fluctuate the numbers of accuracy. Same applies for male testers.

Graphical User Interface & Application

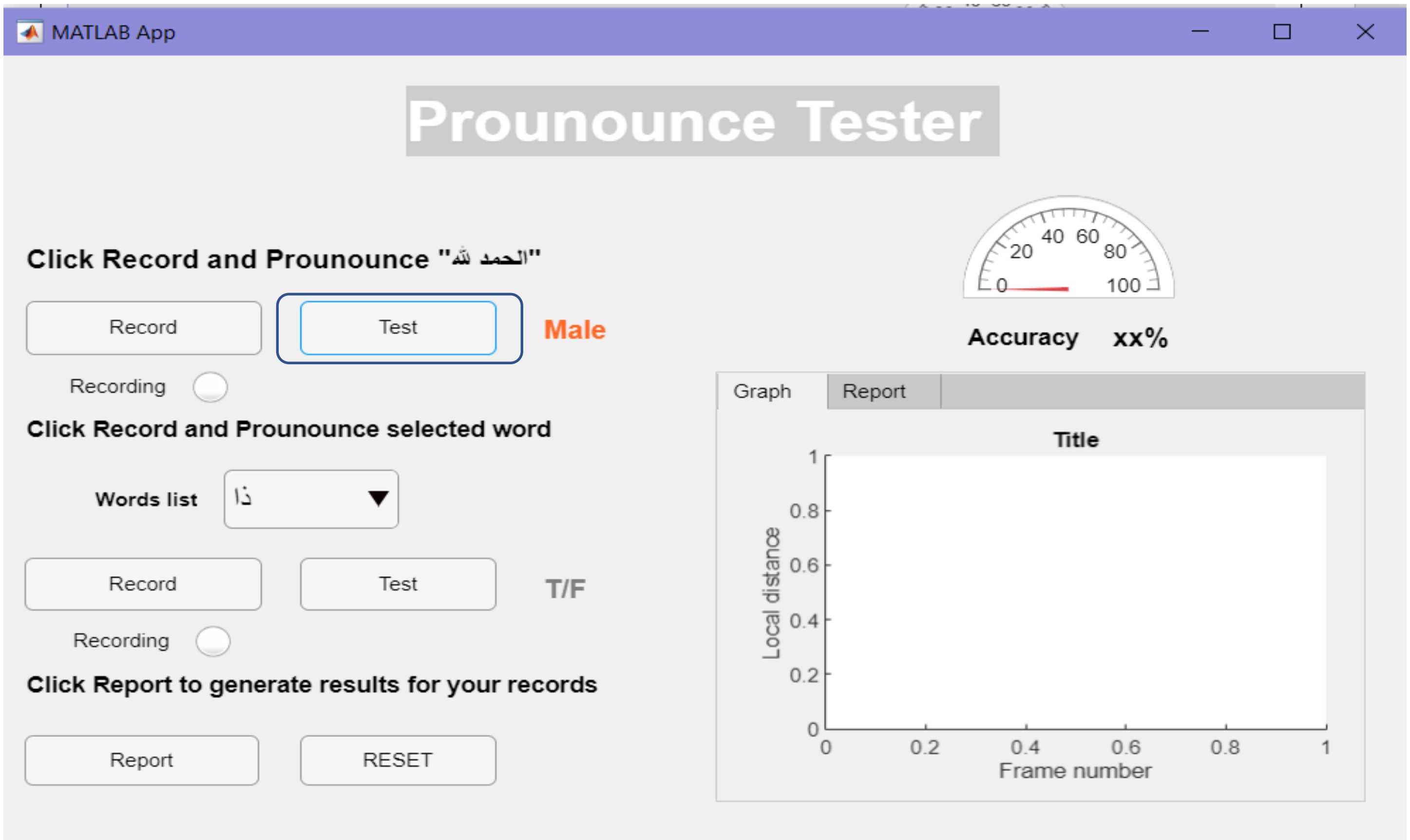
This is our App interface



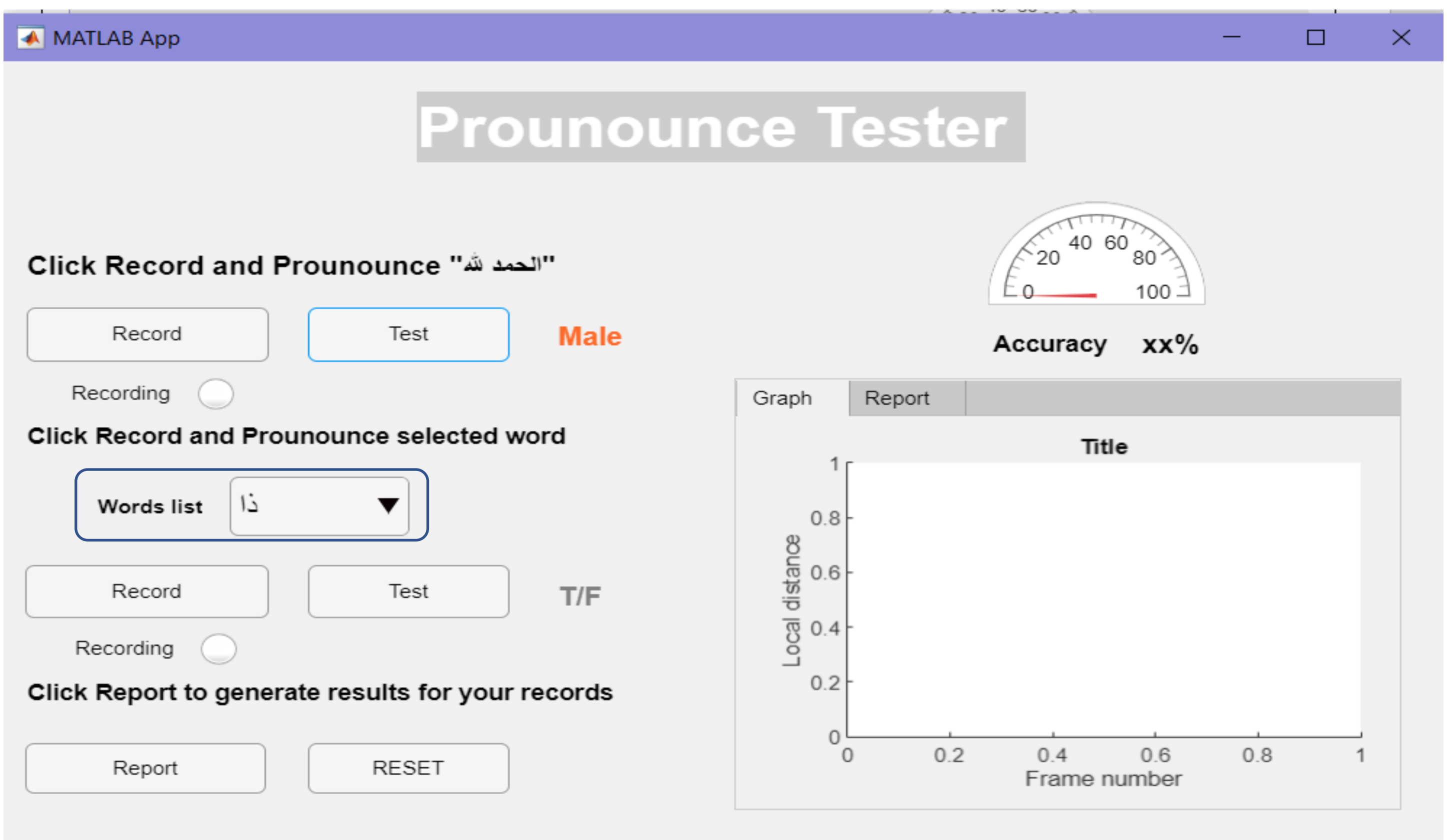
If U want to Test Gender U Will Press the Record Button and Start Recording The word " الحمد لله" Then the program will light the lamp and test the record and tell u the gender



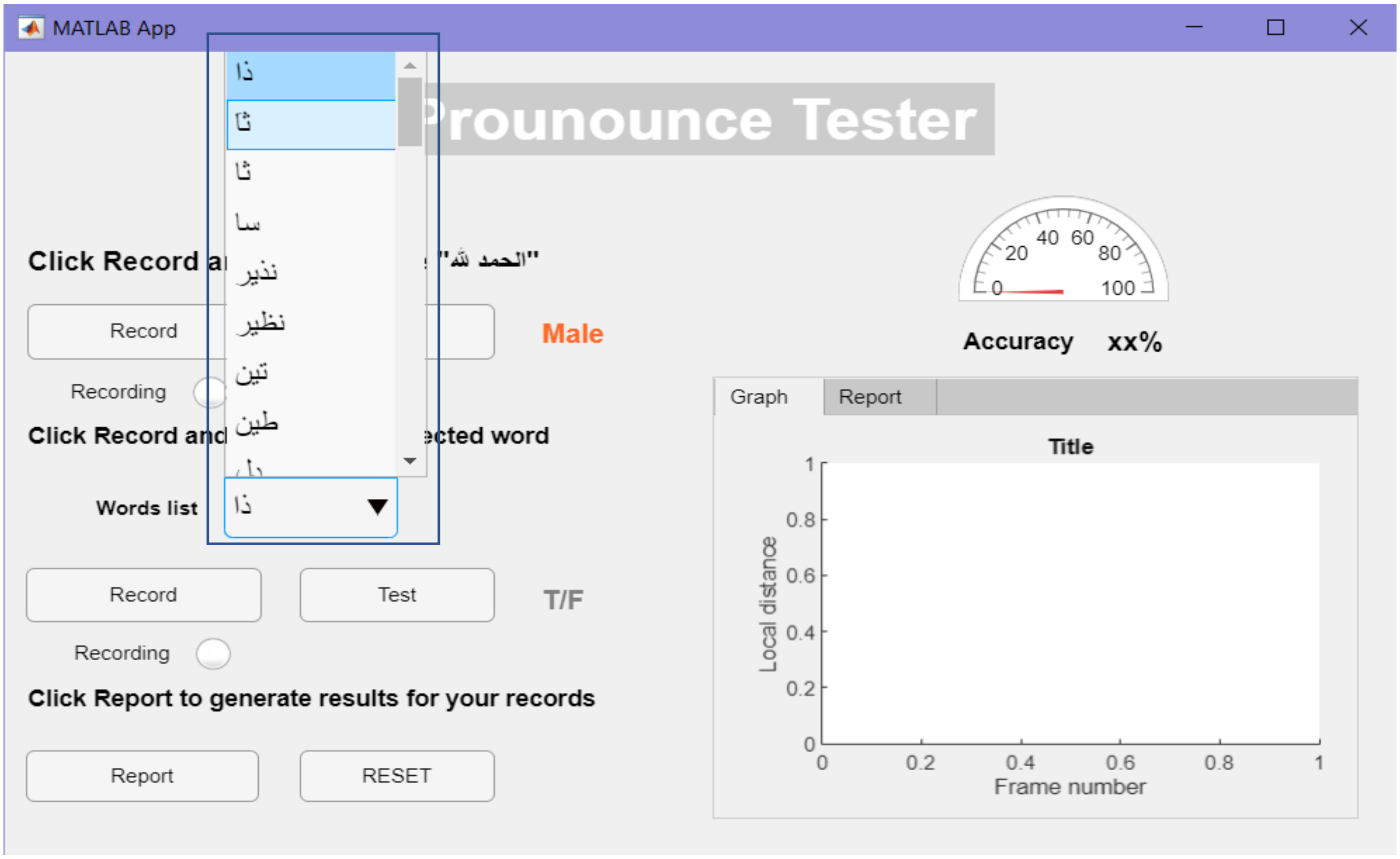
After that u must press the button Test to see the decision of the gender and the output



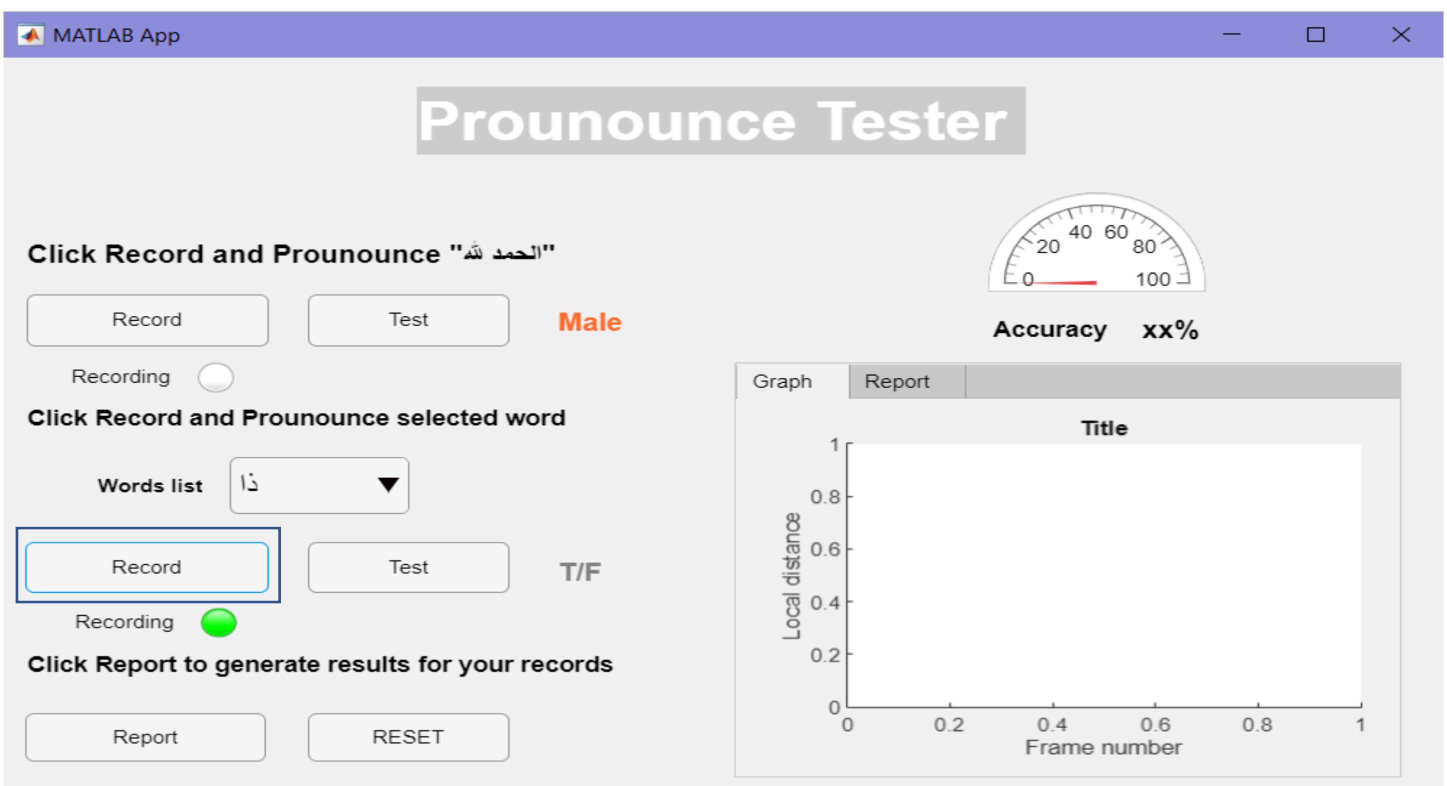
Also u can choose any word from the list Words list



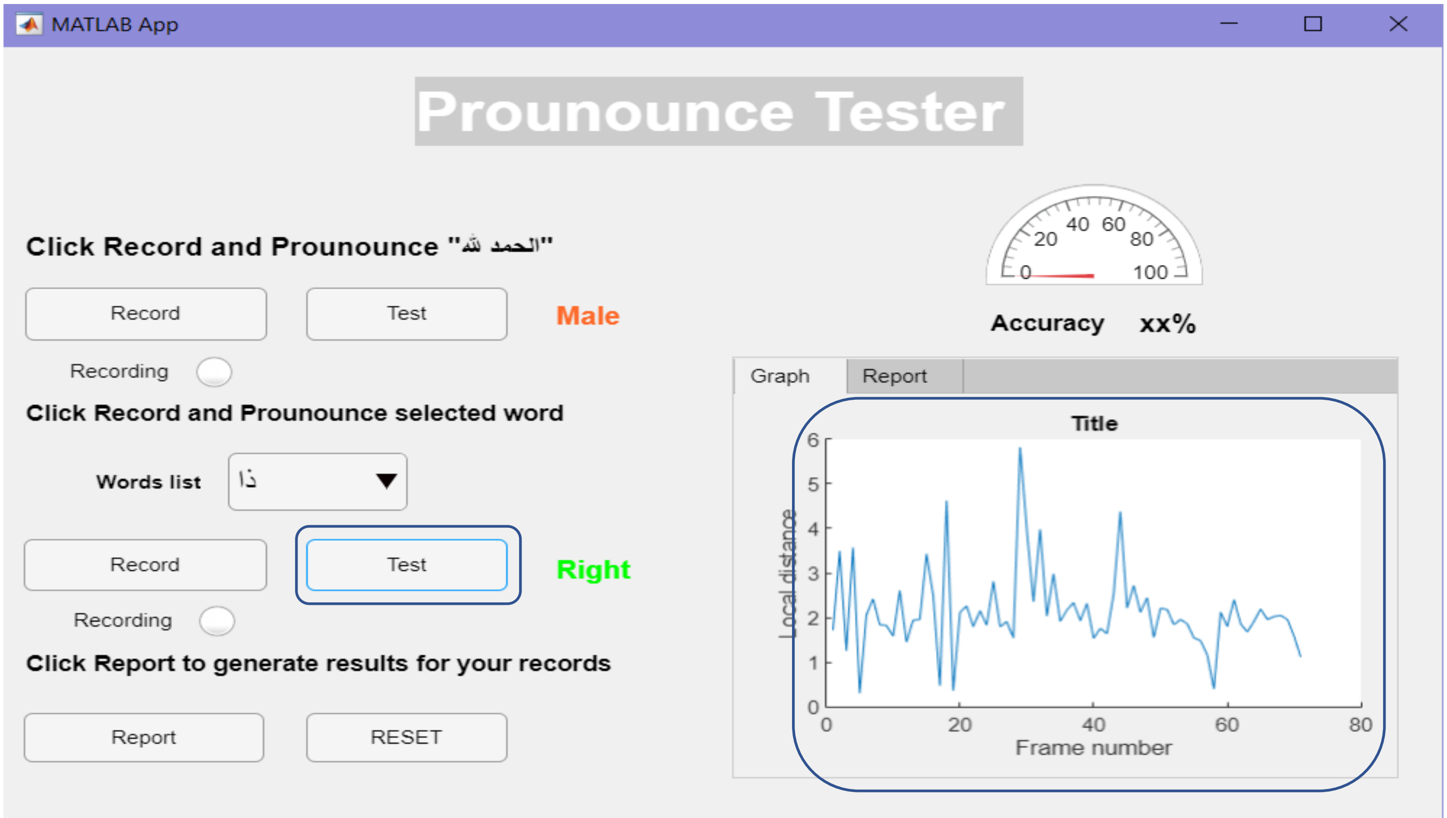
U can choose the required word from ("ذا" , "ثا" , ………….. etc)



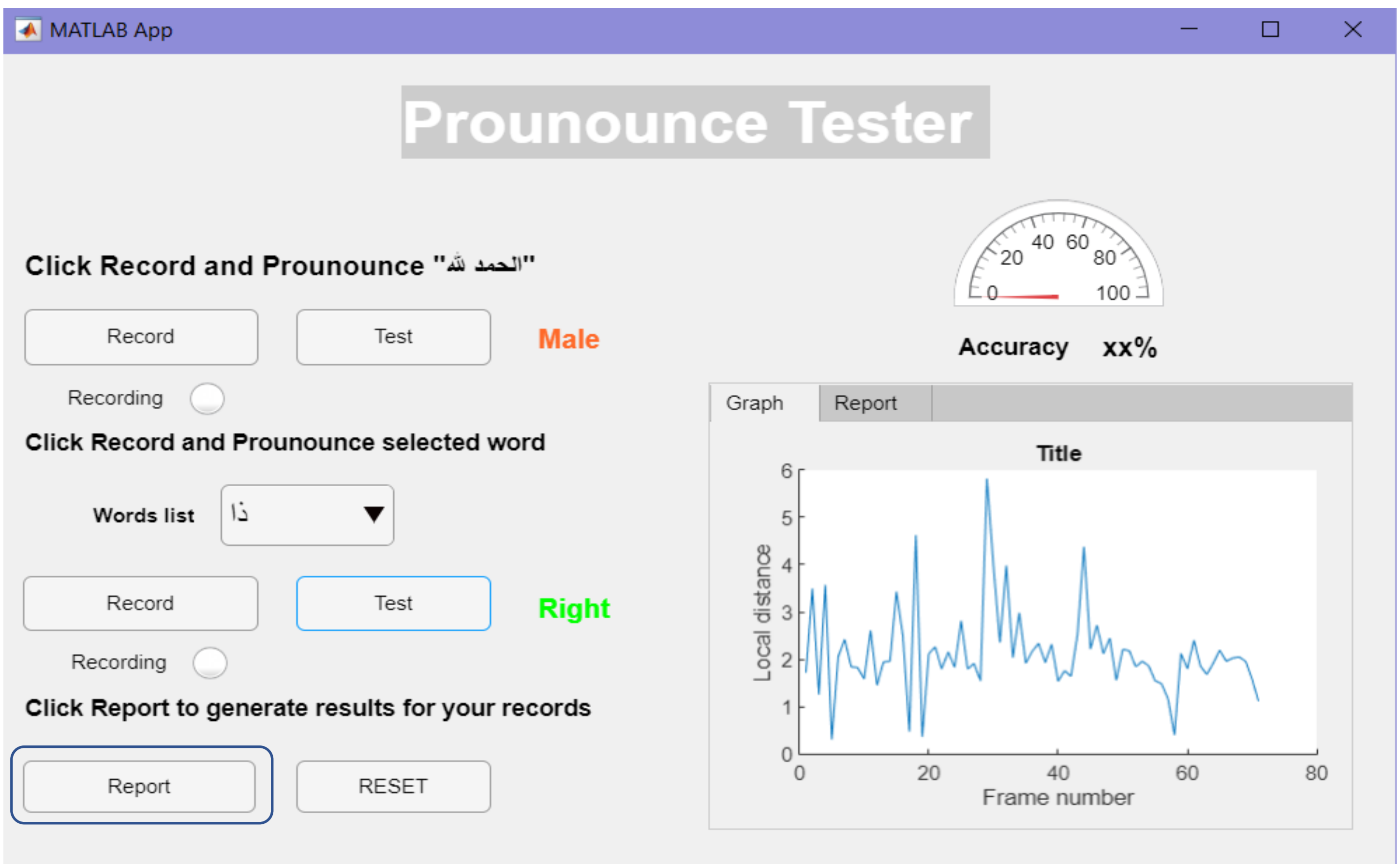
After That U Must Press the Record Button and The App Will light The lamb and Record the word



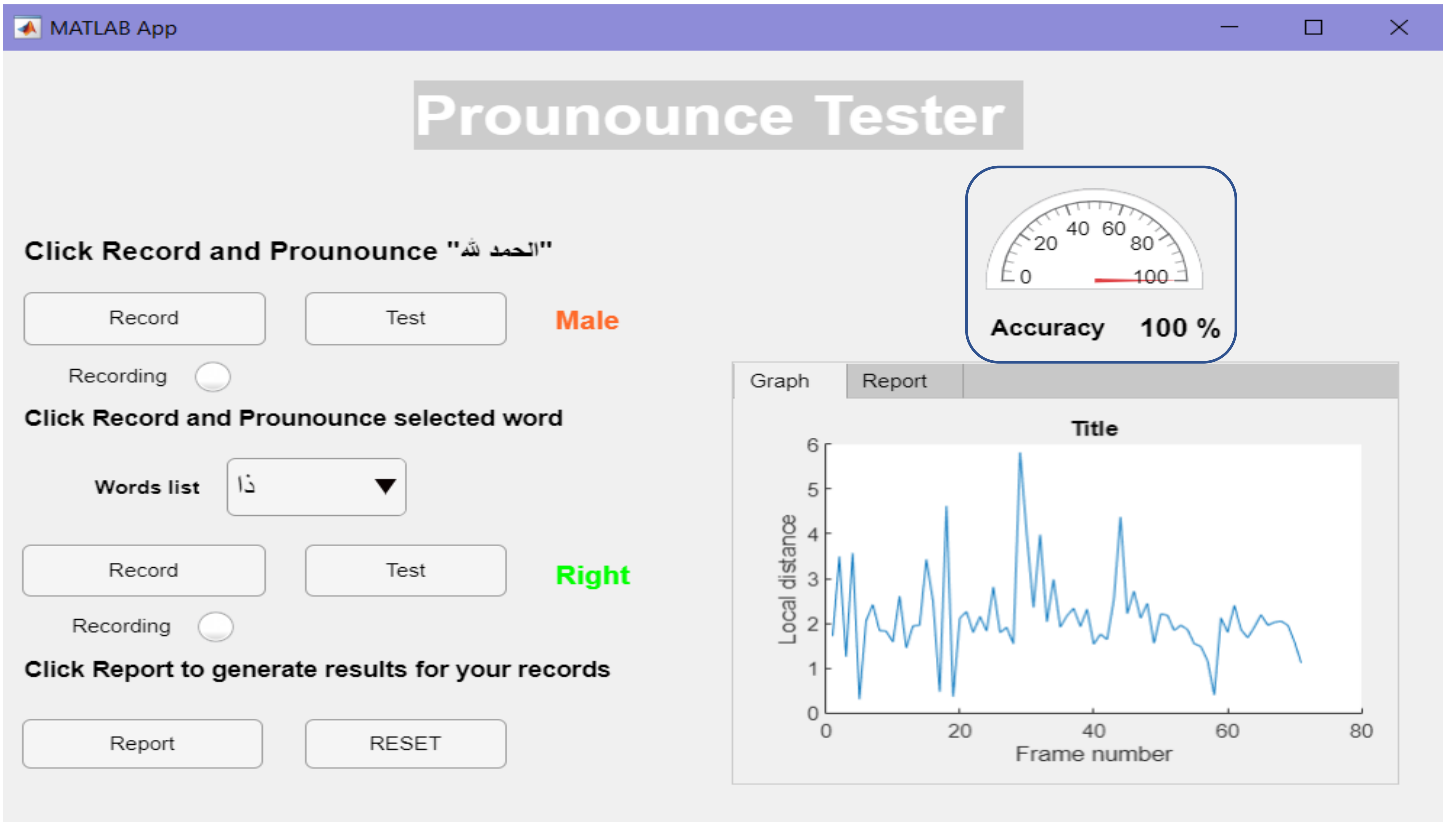
Then If u want to see the result of the program u must press the Test Button



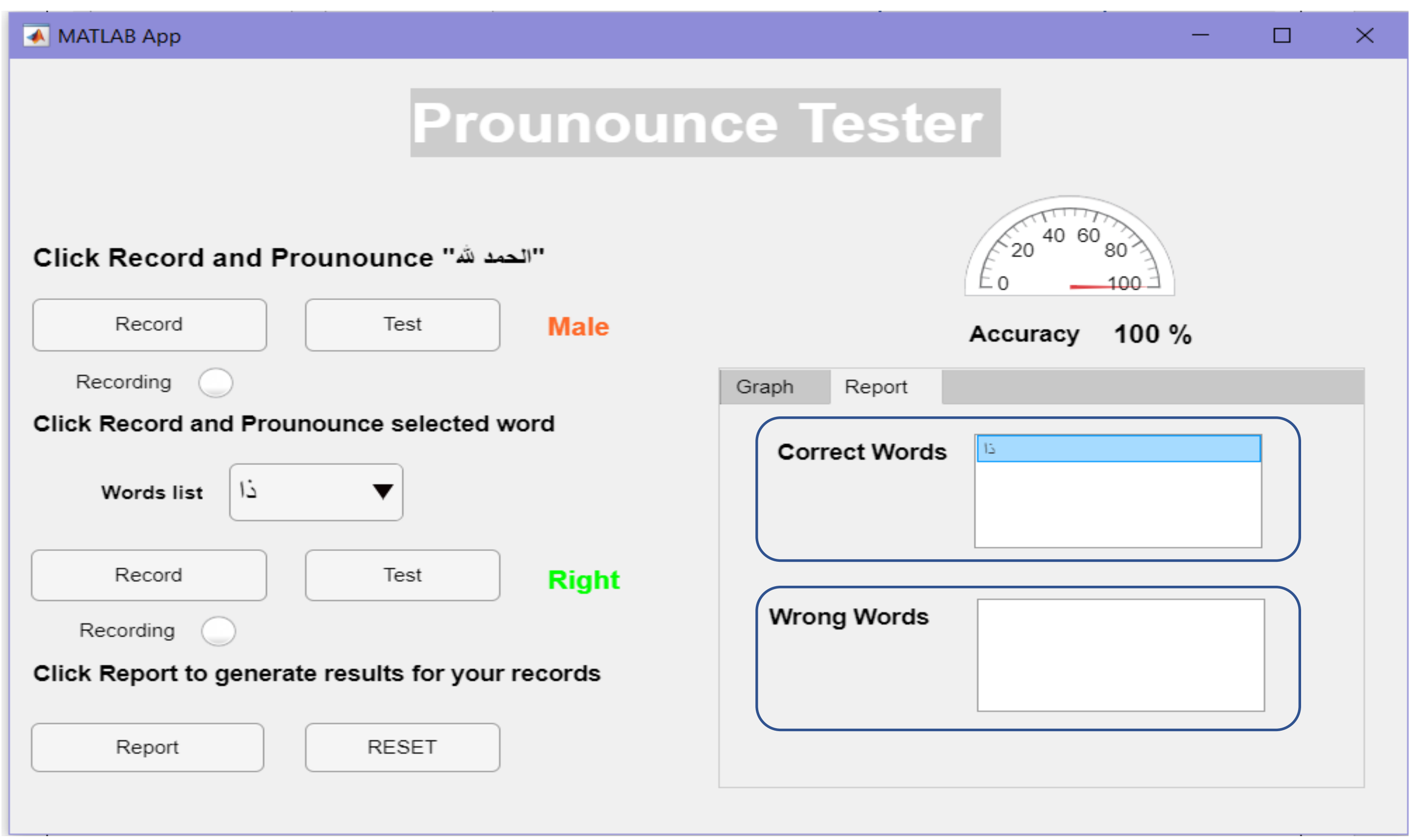
Then if u want to see the report program and the Accuracy u must press the Report Button



Now we will see the Accuracy output and the Accuracy percentage



Then we press the Report Button on the Graph and see the total Correct words and the total Wrong Words



Conclusion

Automatic speech recognition will continue to find applications, such as meeting/conference summarization, automatic closed captioning, and interpreting telephony. It is expected that speech recognizer will become the main input device of the "wearable" computers that are now actively investigated. In order to materialize these applications, we have to solve many problems. The most important issue is how to make the speech recognition systems robust against acoustic and linguistic variation in speech. In this context, a paradigm shifts from speech recognition to understanding where underlying messages of the speaker, that is, meaning/context that the speaker intended to convey are extracted, instead of transcribing all the spoken words, will be indispensable and if we have more time in the future, we make more features to our application to obtain complex capabilities to users to meet and use the powerful science like that speech recognition.

References

[1] "Speech Recognition," [Online]. Available: https://paperswithcode.com/task/speech-recognition. [Accessed 2022].

[2] MathWorks, "Detect boundaries of speech in audio signal - MATLAB detectSpeech," [Online]. Available: https://uk.mathworks.com/help/audio/ref/detectspeech.html. [Accessed 2022].

[3] J. Hui, "Speech Recognition — Feature Extraction MFCC & PLP," 28 August 2019. [Online]. Available: https://jonathan-hui.medium.com/speech-recognition-feature-extraction-mfcc-plp-5455f5a69dd9. [Accessed 2022].

[4] MathWorks, "Extract MFCC, log energy, delta, and delta-delta of audio signal - MATLAB mfcc," [Online]. Available: https://uk.mathworks.com/help/audio/ref/mfcc.html. [Accessed 2022].

[5] MathWorks, "Distance between signals using dynamic time warping - MATLAB dtw," [Online]. Available: https://uk.mathworks.com/help/signal/ref/dtw.html. [Accessed 2022].

[6] MathWorks, "Write table to file - MATLAB writetable," [Online]. Available: https://uk.mathworks.com/help/matlab/ref/writetable.html. [Accessed 2022].