The Improvement of Real-Time ASR by using Gender recognition

Manar M. Salah El.Din[a], *Micheal N. Mikhael*[b], *Hala A. Mansour*[c]
[a] Benha University, Faculty of Engineering at Shoubra, Cairo, eng_manar_mohamed@yahoo.com
[b] Benha University, Faculty of Engineering at Shoubra, Cairo, michael.gerges@feng.bu.edu.eg
[c] Benha University, Faculty of Engineering at Shoubra, Cairo, hala.mansour@gmail.com

Abstract— Nowadays, the need of Automatic Speech Recognition (ASR) is more than before, because of the important of using computer in our life at the recent moments. So we need to improve ASR system with high quick results without decrease quality of system to keep pace with progress and evolution, where feature extraction techniques and matching or recognition techniques are the main parts of ASR system; both affect the quality and speed of the ASR system.In this paper, we examine ASR with adding gender block and using three feature extraction techniques, which are Mel-Frequency Cepstral Coefficients (MFCC), Formant Estimation Coefficients (FEC) and Linear Predictive Codes (LPC) to extract features from speech files, and then fed the output from them to one type of the matching techniques (i.e. Euclidean Distance ED and Dynamic Time Wrapping DTW).The experiment was executed with four different languages are Arabic, English, French, and Italy.

The main objective is to reach to the effect of adding gender block in ASR system on response time for the Arabic language and therefore we can use continuous speech easily with acceptable recognition time.Huge testing was done using Matlab programming with 500 speech files for Arabic, 300 for English, 50 for French and 40 for Italy each with time duration 2 sec.

It was found that MFCC with E.D or DTW are the best combinations for the Arabic language since it achieves the minimum time response over the other combinations without effecting on the quality of ASR system, also it was found the addition of gender classification in training and testing database so this decreases response time for all combination.

In addition, It was reached that adding gender will improve the results faster than not using gender with about 49.74%.

Index Terms— Gender classification, ASR, MFCC, Formant, LPC, E.D., and DTW.

1 Introduction

Speech is the most convenient way of interaction among people. Researchers have done many researches to have such easy interaction between people and machine [1]. All know that technology is progressed and growing every day like big data, artificial intelligence and machine learning etc. Beside these technologies, the field of speech recognition has also developed.
Microsoft in China is the first that introduced it by when it developed a real time application that translated English into Chinese.
Therefore, we can say that Automatic Speech Recognition (ASR) known as computer speech recognition is a process in which speech signal is converted into a sequence of words as input and the text is the output of the system by using some algorithms as shown in figure (1) [2].

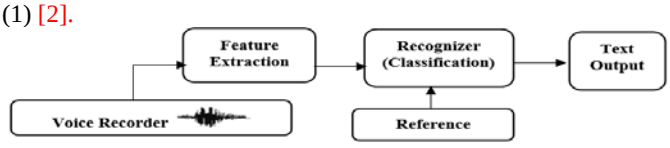


Figure 1. Automatic Speech Recognition System (ASR)

In addition, with the rapid rate of progress in technology, we need to improve ASR system with high quality and more fast time response than before. Therefore, we need to increase the speed of time response for recognition system by reach the best feature extraction technique with matching or recognition technique besides adding gender classification at the beginning to reduce time consumed of recognition process.

2 The Exiting used Techniques

2.1 Feature Extraction techniques

The feature extraction techniques are used to reduce the problem of large variability of the speech signal. This process eliminates a different source of information, that if the sound is voiced or unvoiced or silenced [Speech Feature Extraction for Gender Recognition]
The most popular feature extraction techniques in real-time speech recognition and for future also, that extract the feature of speech then use techniques to achieve the recognition of speech, therefore the ASR system will be completed.

1) Mel-Frequency Cepstral Coefficients (MFCC)
2) Formant Estimation Coefficients (FEC)
3) Linear Predictive Codes (LPC)

2.1. A) Mel-Frequency Cepstral Coefficients (MFCC)

Is the most popular and easiest feature extraction technique where extracts feature from human voice, MFCC is depend on human hearing frequencies below 1 KHz or variation of the human ear's critical bandwidth with frequency [8-10] Hz[4], so we deduced that it based on the short-term power spectrum of the human voice [3]. The main steps for computing the MFCC coefficients are shown in Figure (2). The Mel scale basic equation is given in (Eq. 1)

$$F_{Mel} = 2595 \log_{10} (1 + (F_{Hz} /700)) \quad \text{In Mel-scale} \quad (1)$$

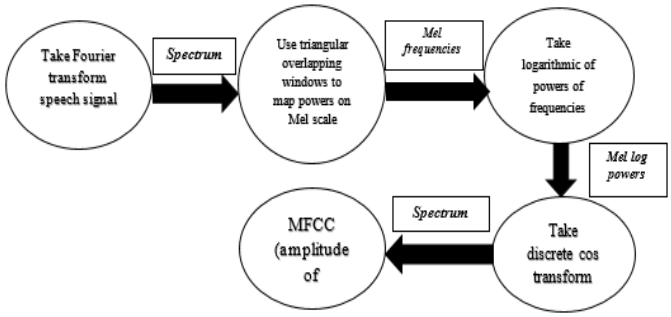


Figure 2. MFCC Feature Extraction Technique [4]

2. 1 B) Formant Estimation Coefficients (FEC).

Formants are the spectral peaks of the sound spectrum" [5].

On the other hand, the formant frequency is a relative maximum in the sound spectrum when the sound is complex. A Unit, Hertz (HZ) [6].as shown in figure (3)"

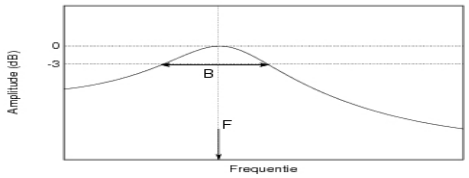


Figure 3. Formants are the peaks displayed in the spectrum

2.1 C) Linear Predictive Codes (LPC)

The Linear Predictive Coding method is a time domain approach. When a sound is pronounced, the LPC method tries to copycat the resonant structure of the human vocal tract, so presents a close and exact representation of the spectral magnitude of signals and produces coefficients related to the vocal tract pattern [7]. LPC has been a well-known technique because it exacts the approximation of the speech parameters and effective computational model of speech [8].

Table (1) Comparison between MFCC, Formant and LPC

Technique	Advantages	Disadvantages
MFCC	The Recognition accuracy is high, Low Complexity and It gives high accuracy results for pure speech after eliminating any noise or channel distortions.	Using filters might affect the performance of result of MFCC and The bandwidth of using a filter is a dependent parameter in design.
Formant	Is faster than LPC and its results are accurate.	Is slower than MFCC and more complex than MFCC
LPC	Useful for encoding speech at the low bit rate and is a reliable and accurate [5].	Is not able to distinguish the words with similar vowel sounds, slowly at a big number of words.

2. 2) THE MATCHING TECHNIQUES.

A pattern classification measures the likeness between an input speech and a reference pattern or a model obtained during training and accordingly determines a reference or a model, which best matches the input speech, as an output.

Therefore, we can say that the speaker's voice is represented by a sequence of feature vectors, which is then compared with the database i.e. the trained data set. [1]

As shown in Figure (4) and Figure (5) the steps of training and testing phases in speech recognition system.

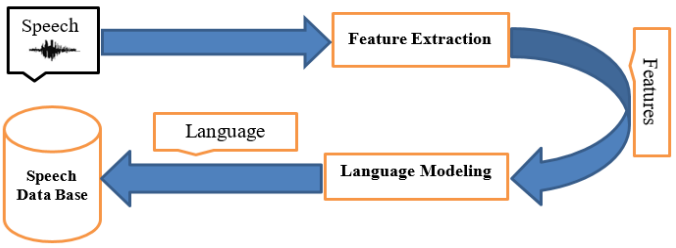


Figure (4): Training Phase

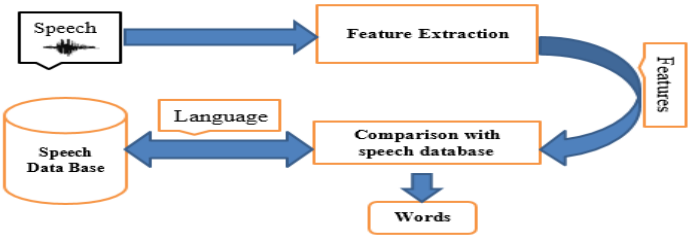


Figure (5): Testing Phase

There are many types of recognition or matching techniques, in this paper we examined the most popular matching techniques:-

- Euclidean Distance (ED)
- Dynamic Time Wrapping (DTW)

2. 2 A) Euclidean Distance (ED)

Euclidean Distance (ED) is the easiest matching recognition technique, which the concept of ED is using the feature extraction coefficients then comparing input speech signal with a model obtained through training or reference pattern, therefore determining a reference or a model, which best matches the input speech, as an output. Then the measured distance is between the test sample and the centroid for each of the clusters. The test sample belongs to the cluster with which the test sample has the minimum distance, for that, the signal is matched. Furthermore, this is known as speech recognized [10].

$$ED = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2} \quad (2)$$

2. B.2) Dynamic Time Wrapping (DTW)

DTW is known as a simple way to recognize an isolated word sample is to compare it with a number of stored word templates and decide the best match. [11]. we can say that DTW is a pattern-matching algorithm with a non-linear time normalization influence. It is based on principle of optimality of Bellman as illustrated in equation (3).

$$(i,j) = d(q_i, c_j) + \min\{(i-1, j-1), (i-1, j), \gamma(i, j-j)\} \quad (3)$$

Where $\gamma(i,j)$ is the cumulative distance of the distance d (i , j) and its minimum cumulative distance among the adjacent cells [12].

Table (2) Comparison between E.D. and DTW matching techniques

Technique	Advantages	Disadvantages
E.D.	1) Simplicity and natural, intuitive sense. 2)It does not depend on the type of language, so it can deal with any type of language	High sensitivity to noise and outliers (especially for sparse data).
DTW	1) Gives good results for the small number of templates with high speed 2) DTW is used in many applications such as cell phones, Because of the simplicity of the hardware implementation of the DTW engine.	1) For a small number of templets. 2) Narrow search space for continuous speech recognition and speaker independent, so it gives a low performance in these types of speech systems.

3. Our contribution part

3. A) Speech dataset

The speech dataset is collected from the internet speech databases. It is recorded speech signals in cleaned way (we can say noiseless). We use in this research four languages. The speech dataset parameters as shown in table 3.

Table 3. Speech files parameters

No. of speech files	Language	Ts (Sec.) Utterance "Time length of sound"	Fs (Hz) "The sampling frequency"	Ns (bits) "No. of bits per sample
500	Arabic	2	8000	16
300	English	2	8000	16
50	French	2	8000	16
40	Italy	2	8000	16

3. B) Simulation Environment

The experiments were done using Core I7/2.4 GHz processor, Windows_10 64-bit Operating System, 6.00 GB RAM. We used MATLAB R2014a as the programming language for the ASR algorithms.

4. Results Analysis.

Two main experiments must be done in order to decide the best combinations of the feature extraction and matching algorithms that will give the fastest recognition time response as the trained database growing up.

The first experiment for calculating the time consuming for the main parts of the ASR system (i.e. feature extraction and matching techniques) in normal sequence without using gender classification and the second also for the same target but with using gender classification, this gives six main probabilities tested by four languages (i.e. Arabic, English, Italian and French) the system block diagram is shown in figure (7) and the 6 probabilities are listed below.

1. MFCC features and Euclidean distances (E.D.) matching technique.
2. FORMANTS features and Euclidean distances (E.D.) matching technique.
3. LPC coefficients features and Euclidean distances (E.D.) matching technique.
4. MFCC features and Dynamic time warping (DTW) matching technique.
5. FORMANTS features and Dynamic time warping (DTW) matching technique.
6. LPC coefficients features and Dynamic time warping (DTW) matching technique.

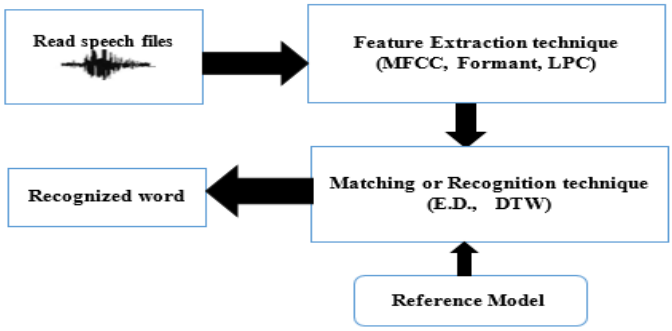


Figure 6. The Methodology of Recognition system [13]

4. 1) Relative Time Consumed by Feature Extraction and Matching Algorithms without using gender classification

The basic target of this research is to find the best combination of the feature extraction and matching techniques that will give the fastest time response for real-time Arabic ASR system without using gender recognition. Table (4) shows the consumed time for each of the three feature extraction techniques under test with respect to the different languages as the speech database grow up By using The Matlab time-consuming property.

It is essential to note that, the time duration recorded for each case depends on:

- The hardware platform (I.e. processor and memory).
- The operating system.
- The programming language (i.e. compiled of interpreted).

Since the platform used as showed above is a laptop (Core I7, 6-GB Ram) with Windows 10 and Matlab R2014a, so the consumed time will display large values. The basic problem is the use of Interpreted Matlab programming language. However, this will not be the case for the real-time Digital Signal Processing (i.e. using real DSP platforms). Also, even if the time values are larger than the real-time systems but it still has the same relative to each other. This means that if one combination is better by certain percent in this results the same will be true for real DSP systems.

Table 4. The Time consumed for each of the Feature Extraction techniques used

Language	No. of speech files	MFCC with (E.D. or DTW) (Sec.)	Formant with (E.D. or DTW) (Sec.)	LPC with (E.D. or DTW) (Sec.)
Arabic	10	0.2895	2.0705	7.373
	20	0.603	4.301	15.8845
	30	0.9735	6.7915	24.9315
	40	1.342	10.0415	34.384
	50	1.7715	11.601	43.1605
	100	3.616	31.042	89.3055
	150	5.493	41.305	138.5285
	200	7.6965	51.5705	182.9255
	250	9.062	61.918	227.8175
	300	11.281	72.623	277.38
	350	13.235	83.829	340.737
	400	17.6145	94.786	410.995
	450	22.69	105.3345	463.438
	500	24.193	117.4345	508.5985

English	10	0.3995	2.5135	9.478
	20	0.794	4.7065	18.0245
	30	1.1595	6.9225	27.238
	40	1.4125	9.0185	35.183
	50	1.81	11.409	43.7685
	100	3.735	23.187	92.1005
	150	5.6605	35.054	142.6475
	200	7.9895	46.921	191.1475
	250	9.3915	58.4195	232.049
	300	11.327	70.471	280.796
French	10	0.396	2.842	9.5295
	20	0.736	4.299	16.8935
	30	2.02	7.3025	25.441
	40	3.2445	11.362	36.6495
	50	4.175	14.5125	47.685
Italy	10	0.765	3.9815	16.1895
	20	1.265	7.787	32.2225
	30	1.8125	9.3475	37.9845
	40	2.3325	11.6	43.21

As shown from the results of the above table the fastest feature extraction technique is MFCC over the Formant or LPC algorithms respectively for all the four languages under test. The MFCC Algorithm is faster that the LPC by 22.76 times while it is faster than the Formant Technique by about 5.74 times.

Figure (6) shows the time consumed for MFCC, Formant, and LPC when working with E.D. or DTW with the four languages under test without using gender recognition. In this section a strange result was found, with change in the language, there is a difference in the time response of the same feature extraction technique although; all inputs have the equal time utterances. In the future work, this result needs more research. The results show that the fastest response was with the Arabic language then the same speed approximately for both English and French, with the slowest one with the Italy Language.

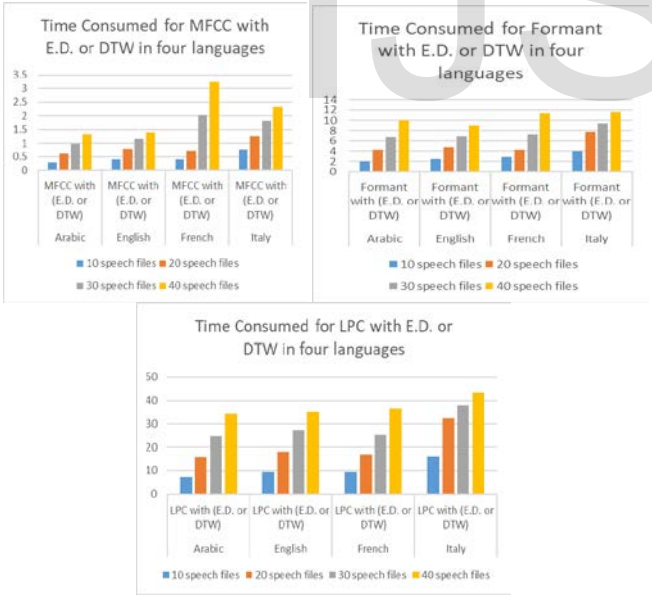


Figure 6. Time of Feature Extraction techniques in different languages

Table 5. The Time consumed for each of the Matching techniques used

Language	No. of speech files	E.D. + MFCC (Sec.)	DTW + MFCC (Sec.)	E.D. + Formant (Sec.)	DTW + Formant (Sec.)	E.D. + LPC (Sec.)	DTW + LPC (Sec.)
Arabic	10	2.898	0.466	1.140	0.194	10.869	3.610
	20	6.502	0.894	3.412	0.356	23.876	7.925
	30	10.606	1.459	6.329	0.620	37.905	12.658
	40	14.220	1.880	16.480	0.901	51.992	17.483
	50	18.254	2.248	20.218	1.029	65.014	21.508
	100	39.569	4.892	71.973	2.788	135.959	44.481
	150	59.225	7.413	193.273	3.625	215.451	68.820
	200	79.384	10.047	314.571	4.462	280.646	92.586
	250	97.338	12.563	436.512	5.361	350.336	115.502
	300	121.696	15.192	506.273	6.262	417.776	141.332
	350	142.620	17.828	685.431	7.230	513.717	181.226
	400	166.314	31.777	807.246	8.104	595.273	206.586
	450	193.175	32.390	931.142	9.581	697.245	239.123
	500	215.055	34.201	1054.021	11.443	760.097	258.892
English	10	3.999	0.607	1.157	0.219	14.103	4.781
	20	7.530	1.090	4.036	0.360	26.988	9.121
	30	11.737	1.626	6.142	0.566	41.620	13.923
	40	15.185	1.997	15.209	0.638	54.331	18.265
	50	19.351	2.779	18.713	0.842	67.902	22.559
	100	38.351	5.565	75.538	1.753	144.839	46.525
	150	59.230	8.139	199.857	2.485	232.999	70.620
	200	84.642	11.022	320.162	3.516	301.748	96.060
	250	100.340	14.029	443.168	4.126	367.396	118.578
	300	116.875	16.278	567.954	5.364	434.623	147.379
French	10	9.899	3.678	5.904	2.714	20.895	10.647
	20	17.820	5.362	9.982	3.989	45.663	20.879
	30	26.543	8.326	12.568	4.872	56.389	23.203
	40	32.603	12.987	16.995	5.886	67.980	26.949
	50	37.064	17.365	21.032	7.035	75.368	30.398
Italy	10	6.893	0.841	2.203	0.373	25.383	7.923
	20	13.838	1.580	6.771	0.767	51.720	16.199
	30	21.530	2.336	8.333	0.811	61.988	19.315
	40	28.653	3.138	10.395	0.904	72.056	22.019

Table (5) shows that the results can be reviewed by normalizing the sum of all tests with the slowest combination (i.e. E.D. + Formant) as shown in figure (10), This figure shows that the fastest matching technique was the DTW, especially with the MFCC or Formant while the slowest response was with E.D. with Formant or LPC respectively. It is important to note that the time consumed for each one of the matching techniques used may be changed with the change of the input feature extraction technique since each feature extraction technique has its own output vector size.

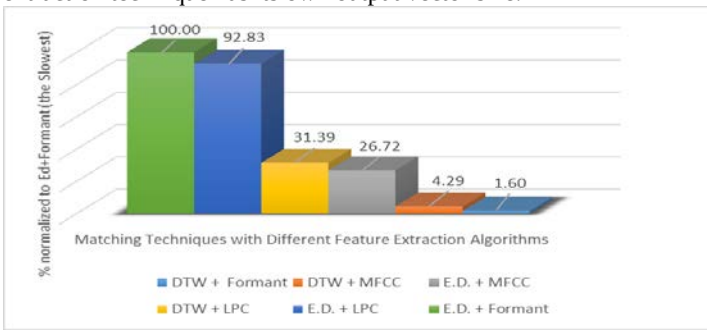


Figure 7. Relative Time Consumed by matching techniques.

Figure (7) shows the average time consumed for all combinations of the feature extraction and matching techniques with the four languages under test with respect to E.D. with Formant without gender recognition.

4. 2) Relative Time Consumed by Feature Extraction and Matching Algorithms with using gender classification

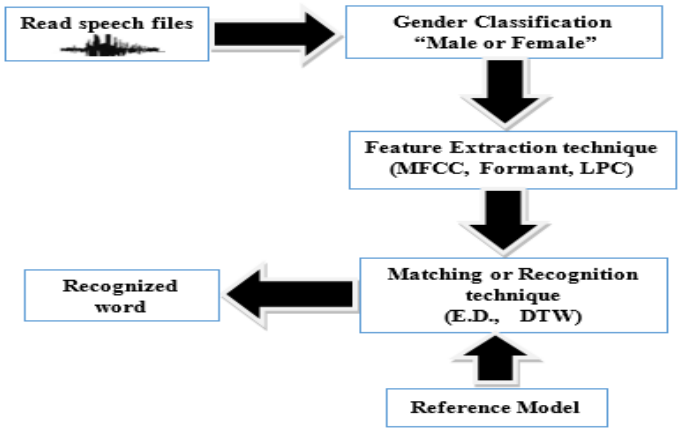


Figure 8. The Methodology of Recognition system by Gender Classification

Table 6. The Relative Time consumed for each of the Feature Extraction techniques used in gender experiment

Language	No. of speech files	MFCC with (E.D. or DTW) (Sec.)	Formant with (E.D. or DTW) (Sec.)	LPC with (E.D. or DTW) (Sec.)
Arabic	10	0.065	0.765	2.4965
	20	0.222	1.591	6.7525
	30	0.407	2.606	11.276
	40	0.5915	4.231	16.0025
	50	0.806	5.0105	20.3905
	100	1.728	14.731	43.463
	150	2.6565	19.8625	68.0745
	200	3.7585	24.9955	90.273
	250	4.441	30.169	112.719
	300	5.5505	35.5215	137.5
	350	6.5375	41.1245	169.179
	400	8.7175	46.603	204.308
	450	11.255	51.8775	230.529
	500	12.0065	57.928	253.11
English	10	0.15	1.2065	4.689
	20	0.347	2.3035	8.9625
	30	0.53	3.4115	13.5695
	40	0.6565	4.4595	17.5415
	50	0.855	5.655	21.8345
	100	1.8175	11.5355	46.001
	150	2.7805	17.4775	71.274
	200	3.945	23.411	95.524
	250	4.646	29.16	115.975
	300	5.6145	35.1865	140.349
French	10	0.3055	1.421	4.765
	20	0.4255	2.15	8.447
	30	1.0155	3.6515	12.721
	40	1.6325	5.471	18.325
	50	2.2405	7.2565	23.843
Italy	10	0.3725	1.9805	8.085
	20	0.6225	3.8835	16.1015
	30	0.8965	4.664	18.9825
	40	1.158	5.79	21.5715

As shown from the results of the above table the MFCC is relatively the fastest feature extraction technique over the Formant or LPC algorithms respectively under the effect of adding gender for all the four languages under test. The MFCC Algorithm is faster that the Formant Technique by about 5.71 times while it is faster than the LPC by 22.92 times.

Figure (9) shows the time consumed for MFCC, Formant, and LPC when working with E.D. or DTW with the four languages under test with the effect of addding gender. As shown the above results are less than the results that are found when no gender recognition is used as table (4).

Also as before, The results show that the fastest response was with the Arabic language then the same speed approximately for both English and French, with the slowest one with the Italy Language.

That we really need to search in this point.(i.e. in Future work).

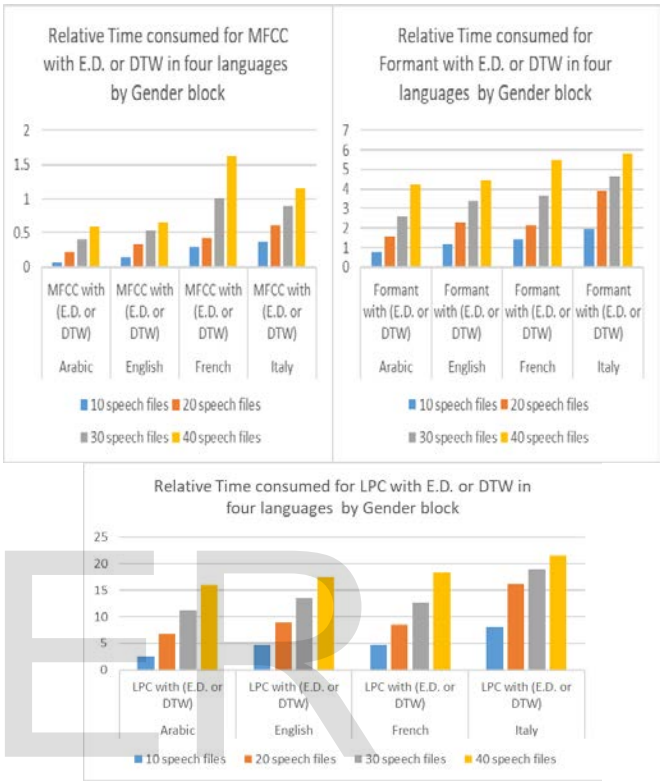


Figure (9) Relative Time Consumed for Feature Extraction Techniques by using Gender block

Table 7. The Time consumed for each of the Matching techniques used by Gender classification

Language	No. of speech files	E.D. + MFCC (Sec.)	DTW + MFCC (Sec.)	E.D. + Formant (Sec.)	DTW + Formant (Sec.)	E.D. + LPC (Sec.)	DTW + LPC (Sec.)
Arabic	10	1.399	0.183	0.520	0.047	5.385	1.755
	20	3.201	0.397	1.656	0.128	11.888	3.913
	30	5.253	0.680	3.115	0.260	18.903	6.279
	40	7.060	0.890	8.190	0.401	25.946	8.692
	50	9.077	1.074	10.059	0.465	32.457	10.704
	100	19.735	2.396	35.937	1.344	67.930	22.191
	150	29.563	3.657	96.587	1.763	107.676	34.360
	200	39.642	4.974	157.236	2.181	140.273	46.243
	250	48.619	6.232	218.206	2.631	175.118	57.701
	300	60.798	7.546	253.087	3.081	208.838	70.616
	350	71.260	8.864	342.666	3.565	256.809	90.563
	400	83.107	15.839	403.573	4.002	297.587	103.243
	450	96.538	16.145	465.521	4.742	348.573	119.512
	500	107.478	17.051	526.961	5.672	379.999	129.396

English	10	1.933	0.320	0.559	0.090	7.0315	2.371
	20	3.745	0.610	1.998	0.160	13.474	4.541
	30	5.849	0.888	3.051	0.263	20.790	6.942
	40	7.573	0.979	7.585	0.299	27.146	9.113
	50	10.156	1.870	9.337	0.401	33.931	11.275
	100	19.166	2.913	37.749	0.857	72.399	24.243
	150	29.595	4.050	99.909	1.223	120.480	35.290
	200	44.821	5.491	160.061	1.938	150.854	49.510
	250	50.150	6.995	224.564	2.043	183.678	62.769
	300	60.418	8.119	283.957	2.662	217.292	69.930
French	10	4.920	1.809	2.922	1.327	10.418	5.294
	20	8.880	2.651	4.961	1.965	22.802	10.410
	30	13.242	4.133	6.254	2.406	28.165	11.572
	40	16.272	6.464	8.468	2.913	33.960	13.445
	50	18.502	8.653	10.486	3.488	37.654	15.169
Italy	10	3.402	0.376	1.057	0.142	12.647	3.917
	20	6.874	0.745	3.341	0.339	25.815	8.055
	30	10.720	1.123	4.122	0.361	30.949	9.613
	40	14.282	1.524	5.153	0.407	35.983	10.965

As shown in table (7), the results when using gender recognition are better and faster than a results that are found in no gender recognition case as table (5).
Also in figure (10), we summarized by normalizing the sum of all tests with the slowest combination (i.e. E.D. + Formant)
The figure shows that the fastest matching technique was the DTW, especially with the Formant or MFCC while the slowest response was with ED with Formant or LPC respectively as shown in figure (7) when no gender case regarding to the fastet results in gender case from no gender case as shown.

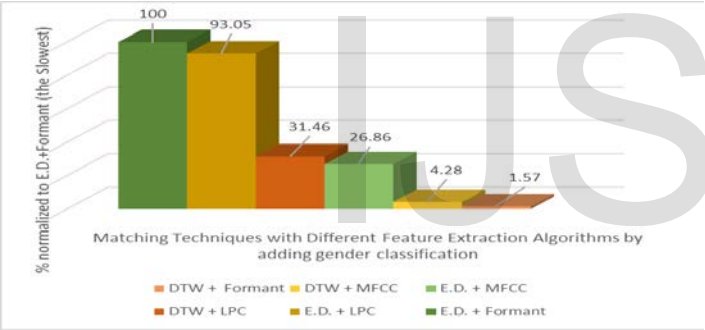


Figure (10). Relative Time Consumed by matching techniques when using Gender block

3. E) Comparison between ASR with Gender and without Gender

3. E. 1) The Relative Time consumed for each of the Feature Extraction techniques used in gender and no gender experiments

As shown in Figure (11), by using gender in ASR system that will effect on the time consumed for each feature extraction techniques
MFCC with E.D. or DTW gives results faster than Formant and LPC whatever gender or no gender case. Adding gender block reduced time consumed for each feature extraction techniques.
In gender case when MFCC works with E.D. or DTW gives results better than without gender by 2.49 times

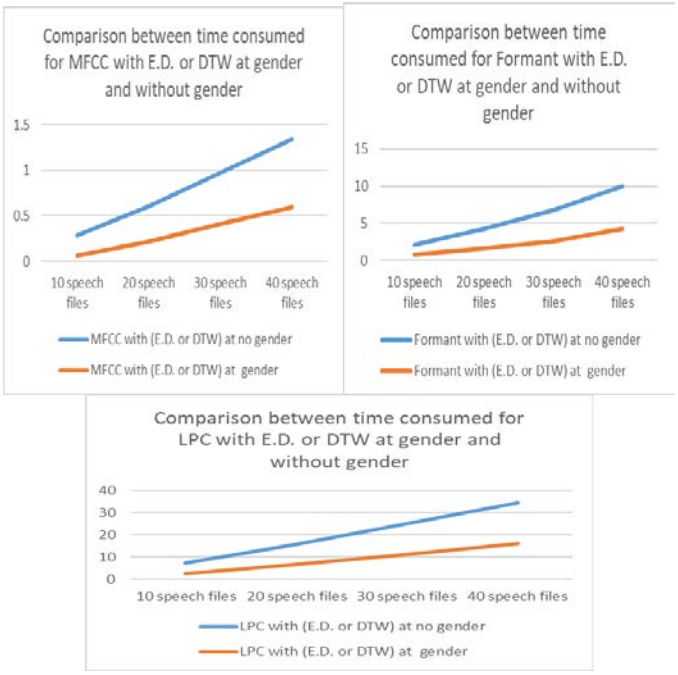
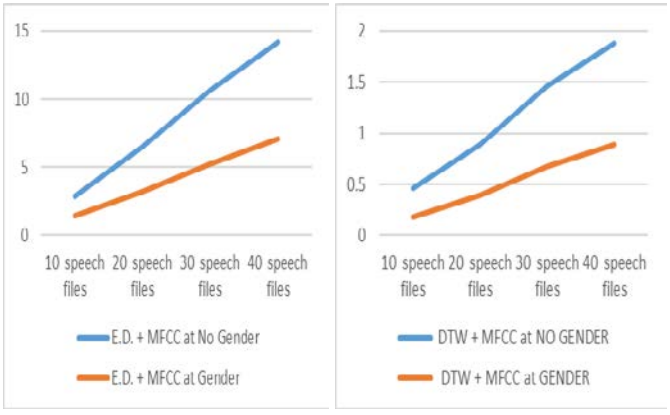


Figure (11) Comparison between Feature Extraction techniques with gender and without gender

3. E. 2) The Relative Time consumed for each of the Matching or Recognition techniques used in gender and no gender experiments

As shown in Figure (12), by using gender in ASR system that will effect on the time consumed for each matching or recognition techniques
DTW gives results faster than E.D. whatever feature extraction technique. Adding gender block reduced time consumed for each matching or recognition techniques.
In gender case, it improves results and gives quick results faster than no gender case with nearly 49.88%.



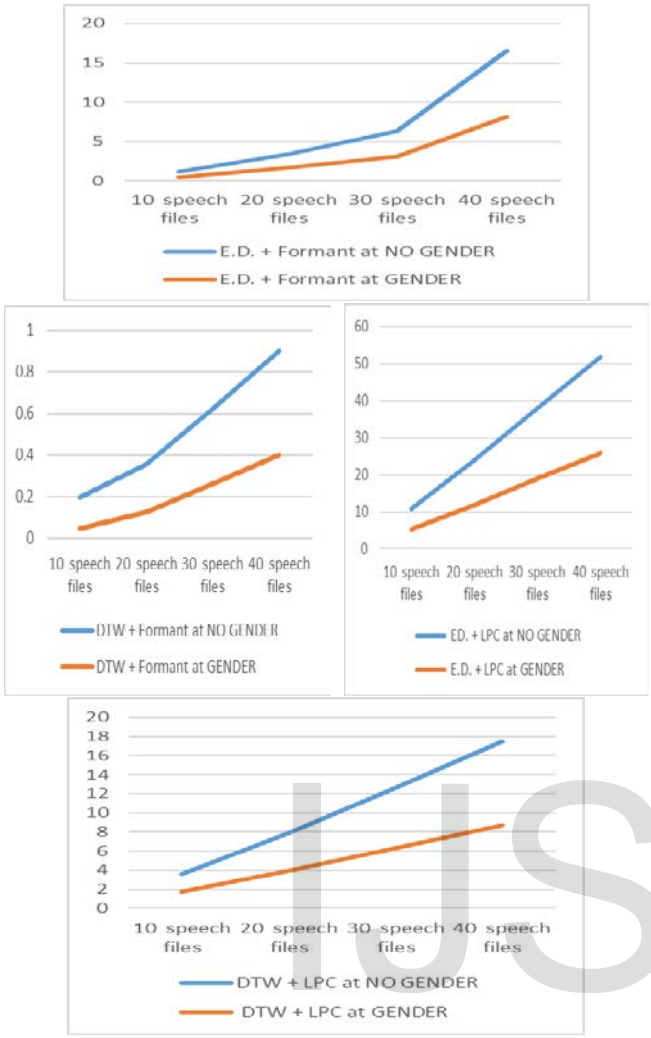


Figure (12) Comparison between matching techniques with gender and without gender

5. CONCLUSION AND FUTURE WORK

Massive testing of the ASR system was executed with different Feature Extraction Techniques (i.e. MFCC, Formant and LPC) and two common matching techniques (i.e. ED and DTW) with four languages (i.e. Arabic, English, French, and Italy). The experiments are done in two cases: First at No gender recognition and second at gender recognition. The main target was the selecting of the best combination of the Feature Extraction and Matching techniques that will give the fastest time response to be implemented with real-time Arabic ASR system. The results show that the best combination for the Arabic language is the use of MFCC with either ED or DTW, and the results are improved when using gender recognition in ASR system as shown in above.

The future work of this research will be the use of ASR in all languages specially Arabic with more matching techniques as HMM and GMM to improve efficiency of real time ASR recognition system with quick results

6. REFERENCES

1. Sanjivani S. Bhabad Gajanan K. Kharate, "An Overview of Technical Progress in Speech Recognition", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3, March 2013
2. Suma Shankaranand1, Manasa S2, Mani Sharma3, Nithya A.S4, Roopa K.S.5, K.V. Ramakrishnan6, "An Enhanced Speech Recognition System", International Journal of Recent Development in Engineering and Technology Volume 2, Issue 3, March 2014
3. Gurpreet Kaur, Ravinder Khanna, and Amod Kumar, "Automatic Speech and Speaker Recognition Using MFCC: Review", International Journal of Advances in Science and Technology (IJAST) Vol 2 Issue 3(September 2014)
4. Shreya Narang, Ms. Divya Gupta, "Speech Feature Extraction Techniques: A Review", International Journal of Computer Science and Mobile Computing, Vol.4 Issue.3, March- 2015
5. DORRA GARGOURI, Med ALI KAMMOUN, and AHMED BEN HAMIDA, "A Comparative Study of Formant Frequencies Estimation Techniques", Proceedings of the 5th WSEAS International Conference on Signal Processing, Istanbul, Turkey, May 27-29, 2006 (pp15-19)
6. Surabhi Vaishnav1, Saurabh Mitra2, " Speech Emotion Recognition: A Review", International Research Journal of Engineering and Technology (IRJET), Volume: 03 Issue: 04 | Apr-2016
7. Michelle Cutajar, Edward Gatt, Ivan Grech, Owen Casha, Joseph Micallef, "Comparative study of automatic speech recognition techniques", The Institution of Engineering and Technology, Vol. 7, Iss. 1, January 2013
8. Urmila Shrawankar and Dr. Vilas Thakare, ResearchGate "arXiv", "Techniques for Feature Extraction In Speech Recognition System: A Comparative Study" May 2013
9. UmaraniJ.Suryawanshi and Prof. Dr. S. R. Ganorkar, "Hardware Implementation of Speech Recognition Using MFCC and Euclidean Distance", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 8, August 2014
10. Akanksha Singh Thakur and Namrata Sahayam,"Speech Recognition Using Euclidean Distance", International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 3, March 2013
11. Mayur Babaji Shinde and Dr. S. T. Gandhe, "Speech processing for isolated Marathi word recognition using MFCC and DTW features", International Journal of Innovations in Engineering and Technology (IJIET), Vol. 3 Issue 1 October 2013
12. Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", JOURNAL OF COMPUTING, VOLUME 2, ISSUE 3, MARCH 2010, ISSN 2151-9617.
13. Manar M. Salah El.Din, Micheal N. Mikhael and Hala A. Mansour, "Evaluation for the Best Combination of Feature Extraction and Matching Techniques used for Real-Time Arabic Speech Recognition ", ESRJ The Engineering and Scientific Research Journal of the Faculty of Engineering, Shoubra, VOLUME No. 01, issue No. 01, Octobor 2008