Automatic Arabic Dialect Classification

Esra J. Harfash
Department of Computer Science
College of Science
University of Basra, Iraq

Abdul-kareem A. Hassan
Department of Computer Science
College of Science
University of Basra, Iraq

ABSTRACT

Automatic Dialect classification (ADC) is represented important new part in automatic speech recognition (ASR). In this paper an automatic Dialect classification to independent system for Arabic languages is presented. The speakers of this system are from some Arabic countries: Egyptian, Iraq, Levantine and Kuwait, where each speaker speaks clip from the dialect of his country. The MFCC is adopted here to extract the important features from the speech signal. In the recognition task the Linear discriminant analyses (LDA) and Dynamic time warping (DTW) are used in classification stage. The LDA and DTW methods are efficient tools for the classification problems with many variations in speech signal. During the testing process, the LDA and DTW was given efficient results in identifying the classes dialect speaker, but the success rate her for DTW is somewhat better compared to LDA .

General Terms

Dialect recognition, Automatic Dialect classification, Dialect and accent recognition.

Keywords

Dialect recognition, Automatic Dialect classification, Automatic speech recognition, Dialect and accent recognition, Linear discriminant analyses (LDA).

1. INTRODUCTION

The dialect is referred to any variety of a language that is shared by a group of speakers. These variations can exist at all linguistic levels, i.e. vocabularies, idioms, grammars and pronunciation. Then the main challenge in the automatic speech recognition is understanding and modeling the individual variation in spoken language. Each speaker has their speaking styles, depending on their dialect and accent as well as their socioeconomic background; these differences give difficulties in the speaker independent systems that is designed to process any variant of any language [1]. so the dialect identification has become an important part in speech technology, also ,whether the task be carried out by a computer or a human expert ,has forensic applications [2][3].

Then dialect and accent recognition have more recently begun to receive attention from the speech science and technology communities.[4]. Where the Dialect classification is a solution which can characterize speaker traits and help in the development/selection of dynamic lexicons by selecting alternative pronunciations, generate pronunciation modeling via dialect adaptation, or train and adapt dialect dependent acoustic models. Dialect knowledge is also helpful for data mining and spoken document retrieval [5].

In Automatic dialect and accent recognizer systems are used two deferent approaches, phonotactic and spectral approaches. The phonotactic approach in recognition is based on the hypothesis that dialects or accents differed in their phone sequence distributions .The spectral modeling approach for dialect and accent recognition is based on the hypothesis that dialects or accents discriminates in terms of their spectral(acoustic) features [6].

The work of this research is about Arabic dialect identification, where in the Arabic language there are multiple variants, including Modern Standard Arabic (MSA), the formal written standard language of the media, culture and education, and the informal spoken dialects that are the preferred method of communication in daily life. While there are commercially available Automatic Speech Recognition (ASR) systems for recognizing MSA with low error rates (typically trained on Broadcast News), these recognizers fail when a native Arabic speaker speaks in his/her regional dialect. Even in news broadcasts, speakers often code switch between MSA and dialect, especially in conversational speech, such as that found in interviews and talk shows. Being able to identify dialect vs. MSA as well as to identify which dialect is spoken during the recognition process will enable ASR engines to adapt their acoustic, pronunciation, morphological, and language models appropriately and thus improve recognition accuracy[7].

This paper is introduced an independent system Arabic Dialect classification based on acoustic model .The speakers from four different countries: Egyptian , Iraq ,Levantine and Kuwait . Two methods has been their experience: : LDA and DTW,to get the best results to our database . The use of Linear Discriminant Analysis for data classification is applied to classification problem in speech recognition, where it is easily handles the case where the within-class frequencies are unequal and . This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal reparability[8] .And the Dynamic Time Warping technique made it possible to make a matching between two speech signals that are subject to alteration and variations in speed from one portion to another.

2. RELATED WORKS

There are many successful approaches in the dialect identification. Some of them adopted Acoustic model, and some of them Phonotactic Model adopted. For example, Nancy F.Chen in research[9] investigate dialect-specific and cross-dialectal phonotactic models, using both language models and support vector machines (SVMs). Techniques are evaluated both alone and in combination with a cepstral system with joint factor analysis (JFA), using a four dialect Arabic data set ., and achieve 2% average equal error rate for pairwise classification. Murat Akbacak and other [10] are used A hidden Markov model to align reference phones with dialect specific pronunciations to characterize when and how often substitutions, insertions, and deletions occur. Decision tree clustering is used to find context-dependent phonetic rules. We ran recognition tasks on 4 Arabic dialects. Ahmed Ali and other in research [11] are investigated different approaches for dialect identification in Arabic broadcast

speech, using phonetic, lexical features obtained from a speech recognition system, and acoustic features using the i-vector framework and thy are studied both generative and discriminate classifiers, combined these features using a multi-class Support Vector Machine (SVM). The results on an Arabic/English language identification task, with an accuracy of 100. Further the results to discriminate between the five most widely used dialects of Arabic: namely Egyptian, Gulf, Levantine, North African, and MSA, with an accuracy of 52%. Mohamed Belgacem and others[12] are proposed automatic recognition system of Arabic dialects . An acoustic survey of the proportion of vocalic intervals and the standard deviation of consonantal intervals in nine dialects (Tunisia, Morocco, Algeria, Egypt, Syria, Lebanon, Yemen, Golf's Countries and Iraq) is performed using the platform Alize and Gaussian Mixture Models (GMM). The results show the complexity of the automatic identification of Arabic dialects since. The better system performance of 73.33%. The researcher K. Sreenivasa Rao and others [13] are proposed Hindi dialects identification, the Prosodic and spectral features extracted from speech are used for discriminating the dialects. Spectral features are represented by Mel frequency cepstral coefficients (MFCC) and prosodic features are represented by durations of syllables, pitch and energy contours. Auto associative neural network (AANN) models are used to capture the dialect specific information from the distributions of the feature vectors. The recognition performance of the developed models is observed to be 62%, 73% and 79% using the spectral, prosodic and spectral plus prosodic features, respectively.

3. SYSTEM DESCRIPTION

In this research, the Acoustic Model is adopted in building a dialect classification system. This system is not depending on the text , and has been applied to a closed group of speakers. In this section a description of the system that have been followed in this work, including the data collection ,feature extraction and classification process. Figure 1 shows the stages of the system that has been followed in this research.
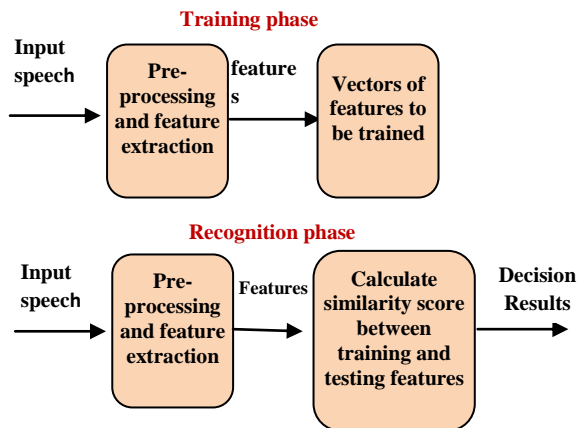

Fig. 1: The block diagram of ADC

3.1 Preprocessing

The purpose of preprocessing is to process and reduce unwanted information from the speech signal. In order to become an appropriate signal to the next processing stages . The following are the techniques that followed here:

- *Filtering :* To reduce the noise in the signal , the band pass Butterworth filters are used (which is a type of Infinite Impulse Response IIR filter) to remove the

higher or lower frequencies than the frequencies of humans.

- *Framing :* For control of the features of the signal, the framing process is applied on speech signal, the window that is used is with length 256 samples for each frame.
- *Remove silences regions :* The speech signal may contain silence regions of no importance in the process of classification, and may cause to reduce the efficiency and increase the time of features extract. In the following the steps that followed to remove silence regions:

1- Find the mean of each frame:

$$AVG_k = (\sum_{i=1}^{N} abs(X_i))/N \qquad (1)$$

$$where\ k = 1 \dots M(no.of\ frames)$$

2- Compare AVG_k with threshold value(here is equal to 0.0023 that has been found experimentally):

If $AVG_k <= 0.0023$,then this frame is deleted , Otherwise reserve the frame . Fig. 2 shows a section of the speech before and after the silence removal .

a. Before silence removal
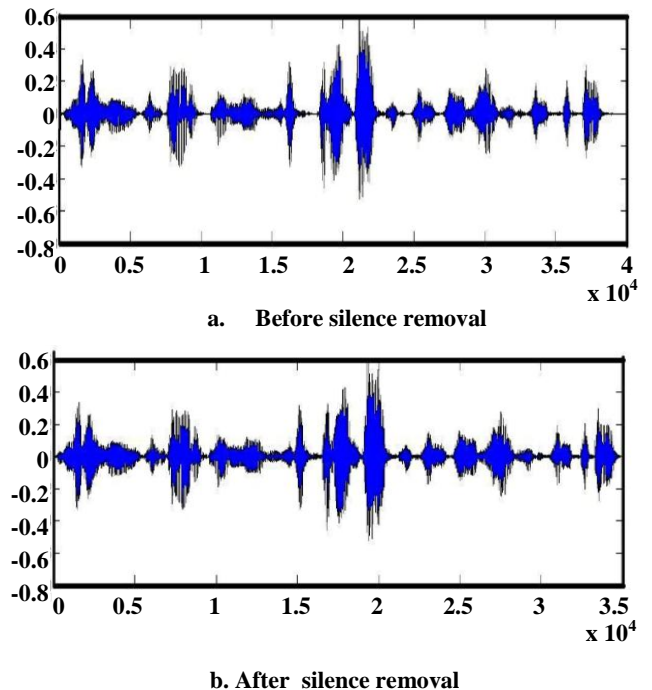

b. After silence removal

Fig. 2: The speech signal before and after silence remove

3.2 Feature selection

The Mel-Frequency Cepstral Coefficients are mostly used features in state-of-art speech recognition system based on frequency domain using the Mel scale which is based on the human ear scale. MFCCs being considered as frequency domain features are much more accurate than time domain features [14].

In this work , the MFCC are depended as feature, and in the following the calculate steps of MFCC [15][16]:

1. The first step is the pre emphases,where the signal is passed through a filter to emphasizes higher frequencies:
2. Calculate:

$$Y[n] = X[n] - 0.95 * X[n] \qquad (2)$$

3. The input speech signal is divided into frames with length 256.

4. Hamming window of length n are computed according to the formula:

$$W(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \tag{3}$$
$$Where \; 0 \le n \le N - 1$$

5. FFT is apply on each frame of signal and convert it to the frequency domain.

6. Now the Mel filter bank is calculated:

$$Melscale = [25595 \log(1+f(Hz)/700] \tag{4}$$

7. For each frame, a discrete cosine transform (DCT) of the log-magnitude of the filter outputs is computed to obtain the MFCCs

The result of each frame is MFCC with length 11 coefficients .The output is a features vectors (V_1, V_2, V_3,….., V_M) .Each V_i with length 11 coefficients, and M number of frames. Figure 3 show the general steps to MFCC Calculation[17].

Pre-emphasis

↓

Hamming Window

↓

DFT

↓

Output energy of filter on Mel-Scale
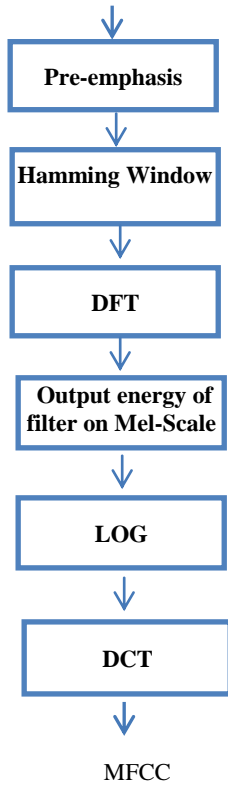
↓

LOG

↓

DCT

↓

MFCC

Fig 3 : MFCC Calculation

3.3 Dialect classification methods

In this paper, LDA and DTW techniques are used to capturing the dialect specific information from the MFCC spectral features, in the following the describe of each methods and its accuracy results.

3.3.1 LDA model

The LDA method have been used in this work to identify the class dialect of speech , where the LDA is one of the important processes in more classification problems.It is commonly used in pattern recognition , face recognition , feature extraction and data dimensionality reduction. The

basic idea of LDA is to find a linear transformation that best discriminate among classes and the classification is then performed in the transformed space based on some metric such as Euclidean distance. Mathematically a typical LDA implementation is carried out via scatter matrix analysis[18].Fisher introduced LDA for two classes and its idea was to transform the multivariate observations x to univariate observations y such that the y's derived from the two classes were separated as much as possible.

Assume there are a set of D-dimensional samples $\{ {}^{(1}, {}^{(2}, \ldots x^{(N)}\}$, N_1 of which belong to class ω_1, and N_2 to class ω_2. We seek to obtain a scalar y by projecting the samples x onto a line $y = w^T x$. To measure separation between projections of different classes[19][20]:

1. Calculate the mean vector of each class in x and y as following:

$$= \frac{1}{N_i} \sum_{x \in w_i} x \;\; and \;\; \bar{\mu} = \frac{1}{N_i} \sum_{y \in w_i} y$$

$$= \frac{1}{N_i} \sum_{x \in w_i} W^T x$$
$$= W^T \pi_i \tag{5}$$

2. Choose the distance between the projected means as our objective Function, see fig (4):

$$J(W) = |\tilde{\mu}_1 - \tilde{\mu}_2| = |W^T (\mu_i - \mu_2)| \tag{6}$$

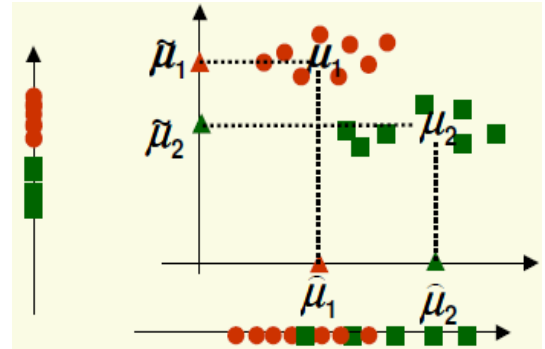**Fig. 4: the mean vector of each class in x and y**

3. The scatter of each class is:

$$\tilde{S}_i^2 = \sum_{y \in w_i} (y - \tilde{\mu}_i) \tag{7}$$

4. The Fisher linear discriminate is defined as the linear function $w^T x$ that maximizes the criterion function :

$$J(W) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{S}_1^2 + \tilde{S}_2^2} \tag{8}$$

5. The measure of the scatter define in multivariate feature space x, which are scatter matrices:

$$S_i = \sum_{x \in w_i} (x - \mu_i)(x - \mu_i)^T \tag{9}$$

$$S_1 + S_2 = S_w$$

where S_W is called the within-class scatter matrix

6. The scatter of the projection y can then be expressed as following :

$$\tilde{S}_i^2 = \sum_{y \in w_i} (y - \tilde{\mu}_i)^2 = \sum_{x \in w_i} (W^T x - W^T \mu_i) =$$

$$= \sum_{x \in w_i} W^T(x - \tilde{\mu}_i)(y - \tilde{\mu}_i)^T W = W^T S_i W$$

$$\tilde{S}_1^2 + \tilde{S}_1^2 = W^T S_W W \qquad (10)$$

7. Similarly, the difference between the projected means can be expressed in terms of the means in the original feature space:

$$(\tilde{\mu}_1 - \tilde{\mu}_2)^2 = (W^T \mu_1 - W^T \mu_2)^2$$
$$= \underbrace{(\mu_i - \mu_2)(\mu_i - \mu_2)^T}_{S_B}$$
$$= W^T S_B W \qquad (11)$$

We can finally express the Fisher criterion in terms of SW and SB as:

$$J(W) = \frac{W^T S_B W}{W^T S_W W} \qquad (12)$$

We will now seek (C-1) projections [y1,y2,...,yC-1] by means of(C-1) projection vectors w_i, which can be arranged by columns into a projection matrix W=[w1|w2|...|wC-1]:

$$y_i = W_i^T x \Rightarrow y = W^T x \qquad (13)$$

The generalization of the within-class scatter is**:**

$$S_w = \sum_{i=1}^{c} S_i \qquad (14)$$

$$where\ S_i = \sum_{x \in w_i}(x - \mu_i)(y - \mu_i)^T\ and\ \mu_i = \frac{1}{N_i}\sum_{x \in w_i} x$$

The generalization for the between-class scatter is:

$$S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T \qquad (15)$$

$$Where\ \mu = \frac{1}{N}\sum_{\forall x} x = \frac{1}{N}\sum_{x \in w_i} N_i \mu_i$$

where ST=SB+SW is called the total scatter matrix .Similarly, we define the mean vector and scatter matrices for the projected samples as:

$$\tilde{\mu}_i = \frac{1}{N_i}\sum_{y \in w_i} y \quad \tilde{S}_w = \sum_{i=1}^{c}\sum_{y \in w_i}(y - \tilde{\mu}_i)(y - \tilde{\mu}_i)^T$$

$$\tilde{\mu} = \frac{1}{N}\sum_{\forall y} y \quad \tilde{S}_B \sum_{i=1}^{c} N_i(\tilde{\mu}_i - \tilde{\mu})(\tilde{\mu}_i - \tilde{\mu})^T \qquad (16)$$

From our derivation for the two-class problem, we can write:

$$\tilde{S}_W = W^T S_W W \qquad \tilde{S}_B = W^T S_B W \qquad (17)$$

Recall that we are looking for a projection that maximizes the ratio of between-class towithin-class scatter. Since the projection is no longer a scalar (it has C-1 dimensions), wethen use the determinant of the scatter matrices to obtain a scalar objective function:

$$J(W) = \frac{|\tilde{S}_B|}{|\tilde{S}_W|} = \frac{W^T S_B W}{W^T S_W W} \qquad (18)$$

And we will seek the projection matrix W* that maximizes this ratio:

$$W^* = [W_1^*|W_2^*|...W_{c-1}^*] = argmax \frac{|W^T S_B W|}{|W^T S_W W|}$$
$$\rightarrow (S_B - \gamma_i S_W)W_i^* = 0 \qquad (19)$$

At the Classification : Once the transformation W is given, the classification is then performed in the transformed space based on some distance metric, such as Euclidean distance

$$d(x,y) = \sqrt{\sum_i (x_i - y_i)^2} \qquad (20)$$

and cosine measure

$$d(x,y) = 1 - \frac{\sum_i x_i y_i}{\sum_i x_i^2 \sum_i y_i^2} \qquad (21)$$

Then upon the arrival of the new instance B, it is classified toarg min(BW , $\tilde{\pi}_k$W),where $\tilde{\pi}_k$is the centroid of k-th class.

Results with LDA: In this part of the system, we changed the dimensions of the training data matrix, for example matrix 11 × 476 × 30 for each person changed to a matrix 11 × 14280. Because we have found through experiments that the results would be better with a change in the dimensions of the data matrix instead of taking the averages of data matrix of each person . Table (1) show the overall results of our experiments on LDA model. table (2) showing the results of each country dialect for each experiment in Table (1).

Table 1. Accuracy results

No.	experiments	accuracy
1	Iraq- Kuwait- Lebanon	62.22%
2	Iraq- Kuwait- Egypt	64.44%
3	Iraq- Lebanon- Egypt	60%
4	Kuwait- Lebanon- Egypt	53.33%
5	Iraq- Kuwait- Lebanon- Egypt	61%

Table 2. The Accuracy of LDA method of each country dialect

No.	Iraq	Kuwait	Lebanon	Egypt	Accuracy
1	73.3%	62.67%	50.67%		62.22%
2	70.3%	50%		73%	64.44%
3	73.3%		33.33%	73.3%	60%
4		63.33%	43.33%	53.3%	53.33%
5	73.3%	61.33%	40%	70%	61%

3.3.2 DTW model

The dynamic time warping (DTW) is implemented here to classify the types of dialect that are defined in this work. The DTW a method to find   the similarity between two Input sequences which may vary in speed. DTW has been applied to any type of sequences : video, audio, and graphics data. Then the DTW is a time series alignment algorithm developed originally . The goal is to   aligning two sequences of vectors by warping the time axis iteratively until an optimal match between the two sequences is found[21].

Given two sequence patterns X of length n and Y of length m, a straightforward distance measure between X and Y ,as following [22] [23]:

$$X = x_1, x_2, \ldots, x_i, \ldots, x_n$$
$$Y = y_1, y_2, \ldots, y_i, \ldots, y_m$$

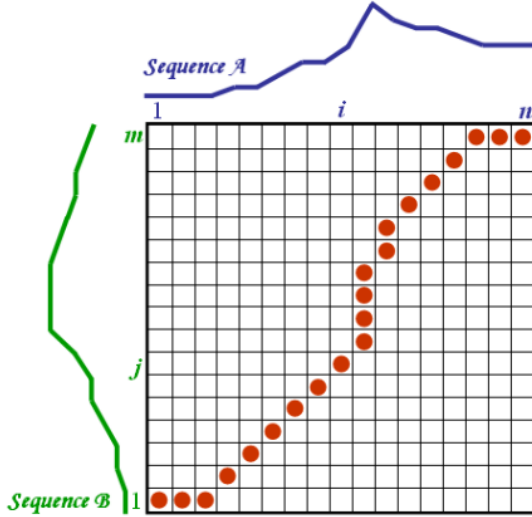Both sequences start on the bottom left of the grid, see figure (5):

Fig. 5: DTW on two sequence X and Y.

To find the *best alignment* between *X* and *Y* one needs to find the path through the grid :

$$P = p_1, \ldots, p_s, \ldots, p_k \quad, \quad p_s = (i_s, j_s)$$

which minimizes the total distance between them.**P** is called a warping function. Time-normalized distance between **X** and **Y,** :

$$D(\textbf{X},\textbf{Y}) = \left[\frac{\sum_{s=1}^{k} d(p_s) \cdot w_s}{\sum_{s=1}^{k} w_s} \right] \quad (22)$$

$d(p_s)$: distance between i_s and j_s and $w_s > 0$: weighting coefficient. The Best alignment path between **X** and **Y**:

$$\textbf{P}_0 = \arg\min (D(\textbf{X}, \textbf{Y})). \quad (23)$$

The number of possible warping paths through the grid is exponentially explosive. Seeking a weighting coefficient(Symmetric form and A Symmetric form)function which guarantees that:

$$C = \sum_{s=1}^{k} w_s$$

is independent of the warping function. Thus

$$D(\textbf{X},\textbf{Y}) = \frac{1}{C} \min_{P} \left[\sum_{s=1}^{k} d(p_s) \cdot w_s \right] \quad (24)$$

can be solved by use of dynamic programming. Now the symmetric algorithm:

Initial condition :   g(1,1) = 2d(1,1).
DP-equation:

$$g(\textbf{\textit{i}},\textbf{\textit{j}}) = \min \begin{matrix} g(\textbf{\textit{i}},\textbf{\textit{j}}-1) + d(\textbf{\textit{i}},\textbf{\textit{j}}) \\ g(\textbf{\textit{i}}-1,\textbf{\textit{j}}-1) + 2d(\textbf{\textit{i}},\textbf{\textit{j}}) \\ g(\textbf{\textit{i}}-1,\textbf{\textit{j}}) + d(\textbf{\textit{i}},\textbf{\textit{j}}) \end{matrix}.$$

Warping window : $\textbf{\textit{j}} - \textbf{\textit{r}} \leq \textbf{\textit{i}} \leq \textbf{\textit{j}} + \textbf{\textit{r}}$.
Time-normalized distance:

$$D(\textbf{X}, \textbf{Y}) = g(\textbf{\textit{n}}, \textbf{\textit{m}}) / C$$
$$C = \textbf{\textit{n}} + \textbf{\textit{m}}.$$

Asymmetric algorithm:

Initial condition: g(1,1) = d(1,1).
DP-equation:

$$g(\textbf{\textit{i}},\textbf{\textit{j}}) = \min \begin{matrix} g(\textbf{\textit{i}},\textbf{\textit{j}}-1) \\ g(\textbf{\textit{i}}-1,\textbf{\textit{j}}-1) + d(\textbf{\textit{i}},\textbf{\textit{j}}) \\ g(\textbf{\textit{i}}-1,\textbf{\textit{j}}) + d(\textbf{\textit{i}},\textbf{\textit{j}}) \end{matrix}.$$

Warping window: $\textbf{\textit{j}} - \textbf{\textit{r}} \leq \textbf{\textit{i}} \leq \textbf{\textit{j}} + \textbf{\textit{r}}$.
Time-normalized distance:

$$D(\textbf{X}, \textbf{BY}) = g(\textbf{\textit{n}}, \textbf{\textit{m}}) / C$$
$$C = \textbf{\textit{n}}.$$

 Results with DTW : In experiments with this method DTW, the data matrix for each one voice M is 476x11 .

Table (3) show the overall results of our experiments on LDA model, and Table (4) showing the results of each country dialect for each experiment in Table (3).

Table 3. Accuracy results

No.	Experiments	Accuracy
1	Iraq- Kuwait- Lebanon	71.11%
2	Iraq- Kuwait- Egypt	73.33%
3	Iraq- Lebanon- Egypt	75.56%
4	Kuwait- Lebanon- Egypt	57.78%
5	Iraq- Kuwait- Lebanon- Egypt	61.67%

Table 4. The Accuracy of each country dialect in DTW method

No.	Iraq	Kuwait	Lebanon	Egypt	Accura-cy
1	80.67%	73.33%	53.33%		71.11%
2	80%	80%		60%	73.33%
3	80%		86.67%	60%	75.56%
4		73.33%	46.67%	53.33%	57.78%
5	73.33%	73.33%	46.67%	53.33%	61.67%

Through experiments and the results obtained, there are some things we refer them here:

1. The results of the Iraqi dialect is in general better than the rest of the dialects, because the sounds were recorded by a microphone and speakers were from one region. The speakers for other countries(Kuwait- Lebanon- Egypt) were from different regions of their countries.
2.   The results of Kuwaiti dialect is good, because There are little variety in Kuwait Dialects.
3. Because there is overlap between the Egyptian and Lebanese language and there are so many differences in their dialect, we observed that the results is low sometimes .
    .

4. SUMMARY AND CONCLUTION

In this paper, Automatic dialect classification system for Arabic dialects identification is proposed. Four Arabic dialects: Egyptian, Iraq, Levantine and Kuwait are considered in this study. The MFCC features are extracted from the input speech, because it is give important information about the linguistic message. The LDA and DTW models are used in the recognition stage, where the experimental results are shown that these models are appropriate techniques in dialect classification problem which given efficient results. The results accuracy for the DTW is significantly higher than LDA where the average performance of the DTW is 67.9 and LDA is 60.2%.

Also can see , whenever the singles is longer time, the results may be better but may cause an increase in processing time, (where there are some applications need a short time signal t). The amount of training data also have an impact in the accuracy of the results, the more training data the results are better.

5. REFERENCES

[1] F. Biadsy , 'Automatic Dialect and Accent Recognition and its Application to Speech Recognition', Ph.in the Graduate School of Arts and Sciences ,COLUMBIA UNIVERSITY,[2011].

[2] E. Ferragne and F. Pellegrino, 'Automatic Dialect Identification: a Study of British English', in Speaker Classification II/2, Springer, Berlin, [2007]., pp. 243-257.

[3] S. Ellis, ' The Yorkshire Ripper enquiry: Part I. Forensic linguistics ' , [1994], pp. 197–206,.

[4] J. Vonwiller C. Blackburn and R. King ,'Automatic accent classification using artificial neural networks', Proc. EUROSPEECH,Berlin, 2:1241–1244 ,[1993].

[5] G. Liu, Y. Lei, J.. Hansen,' DIALECT IDENTIFICATION: IMPACT OF DIFFERENCES BETWEEN READ VERSUS SPONTANEOUS SPEECH', European Signal Processing Conference(EUSIP) ,[2010].

[6] H. Behravan ,'Dialect and Accent Recognition', Master's Thesis, University of Eastern Finland School of Computing December, [2012].

[7] F. Biadsy and J. Hirschberg and N. Habash ,'Spoken Arabic Dialect Identification Using Phonotactic Modeling' , Proceedings of the EACL Workshop on Computational Approaches to Semitic Languages, 31 March. , [2009] ,pp. 53–61.

[8] S. Balakrishnama, and A. Ganapathiraju, 'LINEAR DISCRIMINANT ANALYSIS FOR SIGNAL PROCESSING PROBLEMS', Institute for Signal and Information Processing ,Mississippi State University ,[1999].

[9] N. F. Chen and W. Shen, 'INFORMATIVE DIALECT RECOGNITION USING CONTEXT-DEPENDENT PRONUNCIATION MODELING', MIT Lincoln Laboratory, Lexington, MA, USA,IEEE,[2011].

[10] M. Akbacak, D. Vergyri and A. Stolcke, 'Effective Arabic Dialect Classification Using Diverse Phonotactic Models', Speech Technology and Research Laboratory, SRI International,Menlo Park, CA, USA,[2011]

[11] A. Ali1 , N. Dehak and P. Cardinal ,' Automatic Dialect Detection in Arabic Broadcast Speech', arXiv:1509.06928v2 [cs.CL], August ,[2016].

[12] M. BELGACEM, G. ANTONIADIS and L. BESACIER , 'Automatic Identification of Arabic Dialects' , University of Grenoble, France Laboratory LIDILEM & LIG : GETALP,[2005].

[13] K S. Rao, S. Nandy and S. G. Koolagudi, 'IDENTIFICATION OF HINDI DIALECTS USING SPEECH' , School of Information Technology ,Indian Institute of Technology Kharagpur, India,[2010].

[14] L. Xie, Z. Liu, 'A Comparative Study of Audio Features For Audio to Visual Cobversion in MPEG-4 Compliant Facial Animation', Proc. of ICMLC, Dalian, 13-16 Aug, [2006].

[15] L. Muda, M.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].

[16] V. Tiwari , 'MFCC and its applications in speaker recognition', International Journal on Emerging Technologies ,[2010],pp. 19-22.

[17] K. Mahkonen, 'Mel-frequency cepstral coefficients (MFCCs)', SGN-14006 Audio and Speech Processing, [2013].

[18] K. Fukunaga, 'Introduction to statistical pattern recognition',. Academic Press, [1990].

[19] C. ZHANG AND Q. RUAN,'Face Recognition Using L-Fisherfaces', JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 26, 1525-1537 ,[2010].

[20] E. ALPAYDIN , ' Introduction toMachine Learning', The MIT Press, [2004].

[21] Sakoe, H. and Chiba, S., 'Dynamic programming algorithm optimization for spoken word recognition',. IEEE Trans. on Acoust., Speech, and Signal Process., ASSP **26**, 43-49 [1978].

[22] Ralph N. ,' Dynamic Time Warping, An intuitive way of, handwriting recognition', MASTER THESIS OF Artificial Intelligence, Cognitive Science,[2004].

[23] Eiji M., 'The Dynamic Time Warping Algorithms'**,** Department of Computer Science, Tsing Hua University, Hsinchu 300 Taiwan ,June 30, [2006]