Palestinian Arabic regional accent recognition

Conference Paper · October 2015

DOI: 10.1109/SPED.2015.7343088

4 authors, including:

Stephen Taylor
University of West Bohemia

Abualsoud Hanani
Birzeit University

Palestinian Arabic Regional Accent Recognition

Abualsoud Hanani, Hanna Basha, Yasmeen Sharaf
Electrical and Computer Engineering
Birzeit University
Ramallah, Palestine
ahanani@birzeit.edu

Stephen Taylor
Computer Science
Fitchburg State University
Fitchburg, Massachusetts USA
staylor@fitchburgstate.edu

***Abstract*—We attempt to automatically recognize the speaker's accent among regional Arabic Palestinian accents from four different regions of Palestine, i.e. Jerusalem (JE), Hebron (HE), Nablus (NA) and Ramallah (RA). To achieve this goal, we applied the state of the art techniques used in speaker and language identification, namely, Gaussian Mixture Model – Universal Background Model (GMM-UBM), Gaussian Mixture Model – Support Vector Machines (GMM-SVM) and I-vector framework. All of these systems were trained and tested on speech of 200 speakers. GMM-SVM and I-vector systems outperformed the baseline GMM-UBM system. The best result (accuracy of 81.5%) was obtained by an I-vector system with 64 Gaussian components, compared to an accuracy of 73.4% achieved by human listeners on the same testing utterances.**

Keywords— Accent recognition, Palestinian Arabic accents, I-vector, Gaussian mixture model, support vector machines

I. INTRODUCTION

The speech signal contains paralinguistic information in addition to its linguistic content, such as gender, accent, language, emotional state and age of speaker. Accent variation is considered one of the major sources of degradation of the performance of Automatic Speech Recognition (ASR) systems [1]. Dealing with accent variation within a language is still attracting the interest of many researchers. Recognizing accent of speaker prior to speech recognition can be used for adapting ASR model parameters for that accent, and hence improve the performance of speech recognition [2].

The most successful adaptation techniques used for adapting ASR model parameters to specific speaker and/or specific accent are the Maximum A Posterior (MAP) adaptation and Maximum Likelihood Logistic Regression (MLLR) [3]. Using these techniques depends on automatically recognizing speaker accent prior to speech recognition and then selecting appropriate model parameters.

In addition, recognizing speaker accent automatically can help in personalizing synthetic speech of a text-to-speech (TTS) system. Consequently, it can also be beneficial for personalizing a speech-to-speech translation (S2ST) system for synthesizing the recognized and translated speech from one language to a specific regional accent in another language.

Several languages of great practical importance, for example, Arabic, English, Spanish and Chinese, comprise a variety of regional dialects that can differ significantly from each other, and can even be mutually unintelligible. The Arabic language can be viewed as a family of related languages, with limited vocabulary overlap between dialects, but with a high overlap among the phoneme inventories. One thing that makes the problem dealing with Arabic dialects more difficult than dialects in the other languages, is that not only does each Arabic country have its own dialect to the extent that it may be considered as a separate language but sometimes there may also exist different dialects within the same country.

Traditionally, Arabic dialects are divided into two major groups, eastern dialects and western dialects. Eastern dialects include Levantine, Gulf, Iraqi and Egyptian dialects. Western dialects (also known as Maghrebi) include Morroccan, Algerian, Tunisian and Libyan dialects. Levantine dialects include Syrian, Lebanese, Palestinian and Jordanian sub-dialects. However, there is a clear distinction of accents of people from Levantine countries, and moreover each region within these four countries has its own accent. This means that an Arabic ASR trained on, or adapted to, Levantine dialectal Arabic speech will not work perfectly on Palestinian dialectal speech. Furthermore, adapting Arabic ASR using Palestinian dialectal speech will not solve the problem of accent variation completely as there is a clear distinction between accents of speakers from different regions inside Palestine. For example, the accent of speakers from the city of Hebron is clearly differentiable from the accent of speakers from Ramallah. In addition to Modern Standard Arabic (MSA), most of the previous work (e.g. [4]) on Arabic dialects divided Arabic dialects into five major regional dialects; Levantine, Gulf, Iraqi, Egyptian and Maghrebi.

Traditionally, it is common to distinguish three sub-dialects within each Arab country: city dwellers, peasants/farmers and Bedouins. The three degrees are often associated with a social class hierarchy from rich, settled city-dwellers down to Bedouins. Different social associations exist as is common in many other languages around the world.

However, in Palestine, the Bedouin accent is very rare nowadays. There is clear differentiation of city dwellers' accents from different cities. Similarly, the accent of people living in the rural area of northern Palestine is distinguished from the accent of people in the south. This motivates us to divide Palestinian accents regionally. To the best of our knowledge, no previous work has been done on the regional accent recognition task of Palestinian Arabic accents.

In this paper we consider the problem of identifying individual Arabic accents from among four regional accents, namely Jerusalem (JE), Hebron (HE), Nablus (NA), and rural area of Ramallah (RA). An accuracy of 81.2% is obtained using a I-vector based system. This performance is compared to an accuracy of 73.4% achieved by human listeners.

The remainder of this paper is organized as follows: in the following section the materials and methods are presented. First the data set used for analysis is described and then the methods of GMMs, SVMs and I-vectors are briefly presented. Next, experiments and results are presented and discussed. Finally, conclusions and future directions are included.

II. Prior Work

Over the last few years, much research has been conducted in the field of accent/dialect recognition. The most successful approaches applied for accent recognition can be divided into two major classes: phonotactic based approaches and acoustic based approaches [4-5]. Phonotactic approaches, such as Phone Recognition followed by Language model (PRLM) [6], use the difference in sequence of sounds for each particular accent for modeling accents, where acoustic approaches use the difference in realization (or pronunciation) of these sounds for building accent dependent models. The most common and successful acoustic approaches applied for accent recognition are those which use Gaussian Mixture Models (GMM) for building an accent independent model (called Universal Background Model, or UBM) and then use MAP adaptation for adapting UBM parameters for each target accent [7]. This system is known as GMM-UBM in the literature.

The GMM-UBM approach is extended by introducing GMM supervectors, high dimensional vectors representing speakers' pronunciation of all of the phones in the training data set. Supervectors are formed by concatenating MAP adapted UBM means from each training or test sample, resulting in vectors of tens of thousands of entries for each speaker. This can be thought of as projecting acoustic features (e.g. MFCC) into a high dimensional space where the accents become easily separable. Support Vector Machines (SVM) are then used for accent classification in the supervector domain [8]. This system is referred to as GMM-SVM in the literature.

GMM-SVM has been improved by incorporating Joint Factor Analysis (JFA) concepts for minimizing the inter-session variation within a single accent [9]. Inter-session variation includes different speakers, different channels, different background noise, different microphones, etc. and is usually referred to as Total Variability (TV). The concept of estimating TV leads to state of the art technique which is referred to as I-vectors [10]. I-vectors are low-dimensional vectors (e.g. 100-400 dimensions, which is low-dimensional compared to the supervectors of the last paragraph) representing speakers. Average of I-vectors in the enrollment data of each accent is often used for representing accents. Similarity between predefined accent-dependent I-vectors and testing I-vector is used for recognizing accent of testing speaker. Dot product and cosine between I-vectors are often used as a measure for accent recognition [11].

Recently, acoustic modeling using Deep Neural Networks (DNN) has been proven to outperform state of the art I-vector approach in many fields such as speech recognition, speaker verification, and language recognition [12]. However, few attempts have been conducted in the field of accent recognition [2]. The conclusions of [21] suggest that many more hours of training data than we have might be required for a DNN accent recognition system.

III. Palestinian Arabic Accents

The Palestinian accent is considered one of the broad class of Levantine Arabic dialects, which includes Arabic dialects of people in Syria, Lebanon, Jordan and Palestine. However, the Palestinian accent contains huge variations, and it can be divided into subsets of accents. We decided to divide Palestinian accents regionally. Each region in Palestine can be assumed to have its own sub accent. Since covering all Arab regions in Palestine is difficult, we decided to consider only four major regions in this study: Jerusalem, Hebron, Nablus, and rural areas around Ramallah. Accents in these four regions are distinctively distinguishable and well known in Palestine.

The noticeable differences between Palestinian Arabic and the Northern forms of Levantine Arabic, such as Western Syrian Arabic and Lebanese Arabic, are stronger in non-urban dialects. The main differences between Palestinian and northern Levantine Arabic can be summarized in the following points [14-16]:

- Phonetically: Palestinian dialect differs from Lebanese regarding the classical diphthongs /ay/ and /aw/, which have simplified to [eː] and [oː] in Palestinian as in Western Syrian. In Lebanon, people have retained a diphthongal pronunciation: [eɪ] and [oʊ].
- Palestinian dialect differs from Western Syrian as far as short stressed /i/ and /u/ are concerned. In Palestine, people keep a more or less open [ɪ] and [ʊ] pronunciation, and are not neutralized to [ə] as in Syrian.
- The Lebanese and Syrian dialects are more prone to ‘imala’ of /a:/ than Palestinian is. For instance, شتا 'winter' is ['ʃɪta] in Palestinian, but ['ʃəte] in Lebanese and Western Syrian.
- In morphology, the plural personal pronouns are إحنا ['ɪħna] 'we', همه ['hʊmme] 'they', كم -[-kʊm] 'you', هم - [-hʊm] 'them' in Palestinian, while they are in Syria/Lebanon نحنا ['nɪħna] 'we', هنه ['hʊnne] 'they', كن -[-kʊn] 'you', هن - [-hʊn] 'them'.
- The conjugation of the imperfect 1st and 3rd person masculine has different prefix vowels. Palestinians say باكتب ['baktʊb] 'I write' باشوف [baʃuːf] 'I see' where Lebanese and Syrians say بكتب ['bəktʊb] and بشوف [bʃuːf]. In the 3rd person masculine, Palestinians say بكتب ['bɪktʊb] 'He writes' where Lebanese and Western Syrians say بيكتب ['byəktʊb].

- Hamza-initial verbs commonly have an [o:] prefix sound in the imperfect in Palestinian. For example, Classical Arabic has اكل /akala/ 'to eat' in the perfect tense, and آكل /ākulu/ with [a:] sound in the first person singular imperfect. The common equivalent in Palestinian Arabic is اكل /akal/ in the perfect, with imperfect 1st person singular بوكل /bōkel/ (with the indicative b- prefix.) Thus, in the Galilee and Northern West Bank, the colloquial for the verbal expression, "I am eating" or "I eat" is commonly ['bo:kel] / ['bo:tʃel], rather than ['ba:kʊl] used in the Western Syrian dialect. Note, however, that ['ba:kel] or even ['ba:kʊl] are used in the South of Palestine.

- The conjugation of the imperative is different too. 'Write!' is اكتب ['ʊktʊb] in Palestinian, but كتوب [kto:b], with different stress and vowel and length, in Lebanese and Western Syrian.

- For the negation of verbs and prepositional pseudo-verbs, Palestinian dialect is similar, to some extent, Egyptian Arabic. Typically, suffixes ش [ʃ] on top of using the preverb negation /ma/, e.g. 'I don't write' is مابكتبش [ma bak'tʊbʃ] in Palestinian, but مابكتب [ma 'bəktʊb] in Northern Levantine.

- In vocabulary, Palestinian is closer to Lebanese than to Western Syrian, e.g. 'is not' is مش [məʃ] in both Lebanese and Palestinian while it is مو [mu] in Syrian; 'How?' is كيف [ki:f] in Lebanese and Palestinian while it is شلون [ʃlo:n] in Syrian as in Iraqi. However, Palestinian also shares items with Egyptian, e.g. 'like' is زي [zeii] in Palestinian instead of مثل [mɪtl], as found in Syrian and Lebanese Arabic.

IV. SPEECH DATA

Based on an extensive review of corpora used for accent identification and the methods used for data collection, a script of prompts (5 sentences, 5 vocabulary lists, one 150-word paragraph in a form of cultural story) was prepared to be read by randomly selected subjects from the four target regions. Prompts contain carefully selected words which were prepared with the advice of a linguistic consultant expert in Palestinian accents. These items are selected to motivate people to speak in their native accents. The words contain most of the consonants, vowels and clusters of the Levantine Arabic dialect.

In total, 300 speakers, born and raised in the four selected regions, were chosen randomly and asked to read prompts in their own accent. All of recordings were collected by making interviews with volunteers with ages varying from 18 to 70. Recordings length varies from 2.5 to 15 minutes. Most of interviews were recorded in a quiet setting using a head mounted microphone and sampled at 44.1 kHz.

The speech recordings are then divided into two sets; recordings of 200 speakers (50 from each regional accent) are used as training and the remaining 100 speakers (25 from each accent) as testing. The gender and age distribution of speakers in train and test sets are worked to be balanced as possible. There is no overlap between training and test speakers. The minimum duration of testing files is around 2.5 minutes, which can be divided into up to five 30s segments. Therefore, five short cuts, with an average length of 30s, were extracted from each of 100 testing speakers (recordings). This results in 500 testing segments used for evaluation. More information about the speech database is presented in Table I.

TABLE I. SPEAKER DISTRIBUTION IN COLLECTED DATABASE

REGIONAL ACCENT	NO. OF TRAINING SPEAKERS	NO. TESTING SPEAKERS	TOTAL DURATION OF TRAIN SET [MIN]	TOTAL DURATION OF TEST SET [MIN]
JE	50 (24f,26m)	25 (12f,13m)	67	16.02
HE	50 (18f,29m)	25 (12f,13m)	58	16.24
NA	50 (23f,27m)	25 (12f,13m)	69	15.4
RA	50 (28f,22m)	25 (13f,12m)	64	18.03
Total	200	100	258	63.6

V. SYSTEM DESCRIPTION

A. Front-end processing

In front-end processing, acoustic features are extracted for both training files and testing files. The speech is segmented into frames by a 20-ms window progressing at a 10-ms frame rate then 19 Mel-scale Cepstral Coefficients (MFCC) are extracted from the speech frames. Next, Shifted-Delta Cepstra [17], with 7-3-1-7 configuration, are computed and appended to the MFCC feature vectors resulting in feature vectors with dimension equal to 68. RASTA filtration is applied to the power spectra. A simple energy-based voice activity detection (VAD) is performed to discard the non-speech frames.

Finally, Cepstral mean and variance normalization (CMVN) is applied on the resulting 68-dimensional feature vectors.

B. GMM-UBM System

The Gaussian Mixture Model is widely and successfully used in various speech processing applications such as speaker, language and accent identification [4, 5, and 18].

A UBM is a GMM trained on acoustic features (68 feature vectors) extracted from all training dataset of all accents. The K-means clustering algorithm is used for finding initial parameters of UBM GMM (means, diagonal covariance matrices and weights).

An accent-dependent GMM is obtained by MAP adaptation (means only) of the UBM using the accent specific enrollment features. The result is one UBM model and one accent-dependent model for each of our four target accents.

We have tried different number of Gaussians for GMMs 16, 32, 64 and 128. No gains have been found with UBM of larger number of components.

C. GMM-SVM System

In our GMM-SVM system, each single utterance is used to estimate the parameters of a GMM by MAP adaptation of the UBM. The adapted GMM mean vectors are then stacked together to form a ‘supervector’. Hence each speech utterance is mapped from the Cepstral feature vector domain to the high-dimensional supervector domain. The aim of this process is to “construct” a good separating hyper-plane in a high dimensional feature space. This process also normalizes the length of the utterances. The supervectors are used to build one SVM model for each accent, by taking one accent as a ‘target’ class and the others as a ‘background’ class. More details about GMM-UBM and GMM-SVM systems can be found in our previous study in [5].

D. I-vector based system

Our third accent recognition system is based on I-vectors, a technique introduced in [10] for speaker identification. This technique has also been proven to work well in language and accent identification [19, 2]. An I-vector classifier is based on a configuration determined by the size of the UBM, the number of factor dimensions for the total variability subspace, as well as the various compensation methods to attenuate within-accent speaker variability.

Feature vectors of each utterance in the training and testing data are used for adapting means of UBM (which is trained on all available training data) in order to estimate an utterance dependent GMM using eigenvoice adaptation technique.

The eigenvoice adaptation technique assumes that all the pertinent variability is captured by a low rank rectangular, total variability matrix T. Then the GMM supervector (vector created by concatenating all mean vectors from the word dependent GMM) for a given word utterance can be modeled as follows:

$$M = m + Tx + \varepsilon \qquad (1)$$

Where m is the UBM supervector, the I-vector x is a random vector having a normal distribution N (0, I), and the residual noise term $\varepsilon \sim N(0, \Sigma)$ models the variability not captured by the matrix T. In training total variability matrix for accent recognition, we assume that every utterance for a given accent group is considered a different class. Additional details on the I-vector extraction procedure are described in [10].

Linear Discriminant Analysis (LDA) is used for reducing I-vectors dimension. The LDA procedure consists of finding the basis that maximizes the between classes variability while minimizing the intra-accent variability.

Recently, Gaussian-PLDA has been used to make the I-vector distribution more normal, which improves performance of I-vector system based on standard LDA [20]. A Gaussian-PLDA model has been trained on dimensionally-reduced I-vectors of training data, and then used for scoring in our I-vector system.

E. Visualization

Our I-vector system maps a word utterance into a 100 dimensional vector space for classification. To obtain insight into how I-vector works, this space can be visualized by

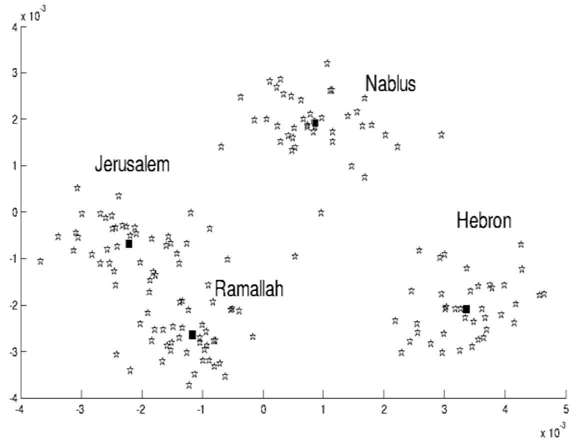

Fig. 1: Visualization of Palestinian accents I-vector space.

projecting it onto a suitable 2-dimensional subspace using LDA. Fig. 1 shows 2-dimensional I-vectors representing four accents. These I-vectors are obtained from training data and using UBM of 64 components. It is clear from the figure that four accents in the I-vector space are to some extent separable. However, there is a clear overlap between accents of people in Jerusalem and people in Ramallah. This may be due to the fact that they are geographically close to each other. Accents of people in Nablus and Hebron cities, which are located in the north and south of West Bank respectively, are clearly distinguishable from each other and from Jerusalem and Ramallah accents.

VI. EXPERIMENTS AND RESULTS

A. Experimental setup

One gender-independent UBM was represented as a diagonal covariance GMM. It was trained on the whole training set of the four accents. The variance flooring was used in each iteration of EM algorithm during the UBM training. The same UBM was used for GMM-UBM, GMM-SVM and I-vector accent recognition systems.

Four accent dependent GMMs were MAP-adapted from the UBM using accent specific data. The UBM means were also MAP adapted using data from each speaker of each accent, generating the GMM supervectors which were used to train the GMM-SVM system.

In order to investigate the effect of number of Gaussian components on our three systems, UBMs with different number of components (16, 32, 64 and 128) were trained and used in the three systems.

B. Results and discussion

In order to study the effect of TV dimension on the performance of our I-vector based accent recognizer, four different numbers of dimensions were tried (50, 100, 150 and 200) using UBM of 64 components and LDA dimension of 3 (one less than the number of classes.) Results of these experiments are presented in figure 2 below. These results suggest that a system with 100 dimensional I-vectors gives the best performance. Based on this, dimension of I-vectors in the subsequent experiments is fixed to 100.

The experimental results of our three systems evaluated on the 500 testing short utterances (with an average length of 30s, five segments from each testing speakers) for different number of Gaussians are as shown in table II.

It is clear from results presented in table II, that by increasing number of Gaussian components of UBM, the performance of three systems increases up to 64 components. At 128 components, performance of all systems degrades slightly. This degradation may be due to over-fitting phenomena because the amount of available training data is relatively small. Another important note: we can observe from the results, that there is no significant difference between the performances of GMM-SVM and I-vector systems. This is not surprising, as both systems are based on the GMM supervectors. However, GMM-SVM system classify accents in the supervector high-dimensional domain using SVM, whereas, I-vector system projects high-dimensional supervectors to a low dimensional space, where TV factors can be estimated easily. The best performance was achieved by I-vector system with 64 components.

In order to investigate the results further, a confusion matrix of our I-vector system with 64 components and total variability dimension of 100, is shown in table III below. Validating 2-dimensional I-vectors representation in figure 1 above, the most confusion is between Jerusalem and Ramallah accents.

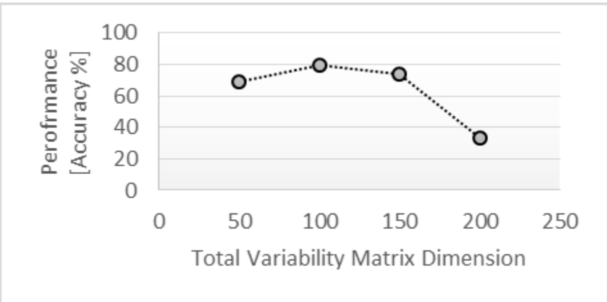

Fig. 2: Effect of TV dimension on I-vector system performance

TABLE II. ACCURACY [%] OF THREE PROPOSED SYSTEMS ON 500 TESTING SEGMENTS

System/UBM order	GMM-UBM	GMM-SVM	I-vector
16	45	75	76.25
32	47.5	77.5	78
64	50	80	81.5
128	47.5	78.75	79.75

TABLE III. CONFUSION MATRIX FOR I-VECTOR REGIONAL ACCENT RECOGNITION EXPERIMENTS

Recognized Accents					
		JE	HE	NA	RA
True Accents	JE	37	3	6	4
	HE	8	42	0	0
	NA	3	2	45	0
	RA	9	2	0	39

C. Human Performance

To provide a baseline against which the automatic accent recognition systems performance could be compared, a web-based human perceptual experiment was conducted using exactly the same 500 test utterances that were used in evaluating the automatic accent recognition systems described above. Twenty subjects took a part in this experiment. Each subject listened to a set of randomly chosen twenty-five test utterances from the total test utterances.

For each utterance, subjects were asked to identify the accent of the speaker, to indicate their confidence in their decision, to estimate the gender and age of the speaker, and to indicate the factors (acoustic quality, use of particular words or phrases, intonation, grammar, or other factors) that influenced their decision. The human listeners scored an average accuracy of 73.4% for the accent identification task.

VII. CONCLUSION

The objective of this paper was to automatically recognize the speaker’s accent among four regional Arabic Palestinian accents by using state of the art techniques. To achieve this task, three different modeling techniques were investigated, the GMM baseline technique, the GMM-SVM and I-vectors.

The results have validated our first hypothesis of using speaker and language identification techniques in the task of regional accent recognition. It was shown that the GMM-SVM and I-vector systems outperform the GMM baseline confirming our second hypothesis that supervectors could create a more discriminative feature space and achieve a higher performance.

As noted in the introduction, this is a preliminary work on Arabic Palestinian regional accent recognition, that the authors are interested in developing further. In the future we intend to deal with the problem of the lack of data by incorporating additional speakers to the database. We also would like to investigate the impact of Palestinian accents specifically, and Arabic dialects generally, on the performance of Arabic ASR and also to extend modeling techniques to include DNN.

VIII. REFERENCES

[1] M. Benzeghiba, R. D. Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi, and C. Wellekens, “Automatic speech recognition and speech variability: A review,” Speech Communication, vol. 49, no. 1011, pp. 763 – 786, 2007.

[2] M. Najafian, A. DeMarco, S. Cox, and M. Russell. "Unsupervised model selection for recognition of regional accented speech." In Fifteenth Annual Conference of the International Speech Communication Association. 2014.

[3] C. J. Leggetter,, and P. C. Woodland, “Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models”, Computer Speech and Language, Vol. 9, Issue 2, 171-185, 1995.

[4] F. Biadsy, 2011, Automatic dialect and accent recognition and its application to speech recognition, Columbia University Academic Commons, http://hdl.handle.net/10022/AC:P:10243.

[5] A. Hanani, M. J. Russell, and M. J. Carey. "Human and computer recognition of regional accents and ethnic groups from British English speech." Computer Speech & Language 27.1,2013: 59-74.

[6] M. A. Zissman, T. P. Gleason, D. M. Rekart and B. L. Losiewicz, “Automatic dialect identification of extemporaneous conversational, Latin American Spanish speech”, Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP-96.

[7] D. A. Reynolds, Thomas F. Quatieri, Robert B. Dunn, Speaker verification using adapted gaussian mixture models, Digital Signal Processing, Volume 10, Issues 1–3, January 2000, Pages 1941, ISSN 10512004, http://dx.doi.org/10.1006/dspr.1999.0361.

[8] Campbell, W.M.; Sturim, D.E.; Reynolds, D.A., "Support vector machines using GMM supervectors for speaker verification," Signal Processing Letters, IEEE , vol.13, no.5, pp.308,311, May 2006 doi: 10.1109/LSP.2006.870086

[9] C. Vair, C. Daniele, C. Fabio, D. Emanuele, and P. Laface. "Channel factors compensation in model and feature domain for speaker recognition." In Speaker and Language Recognition Workshop. IEEE Odyssey 2006, 2006, pp. 1-6.

[10] N. Dehak, P. Kenny, R. Dehak, P. Ouellet, and P. Dumouchel, “Front end factor analysis for speaker verification,” in IEEE Transactions on Audio, Speech and Language Processing, 2011.

[11] D., Najim, R Dehak, P Kenny, N Brümmer, P Ouellet. "Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification." Interspeech. Vol. 9. 2009.

[12] M. Grégoire. "Deep learning for spoken language identification." NIPS Workshop on deep learning for speech recognition 2009.

[13] M Halloun, ‘A Practical Dictionary of the Standard Dialect Spoken in Palestine’. Bethlehem University, 2000.

[14] Arabic Language. Microsoft Encarta Online Encyclopedia. Retrieved on 10 Jan 2015.

[15] O. Boyd Jenkins, Population Analysis of the Arabic Languages, March 2000.

[16] J. Watson, The Phonology and Morphology of Arabic, Introduction, pg. xix.Oxford: Oxford University Press, 2007.

[17] P. A. Torres-Carrasquillo, E.S., M.A. Kohler, R.J. Greene, D.A Reynolds, and J.R. Deller,, "Approaches to language identification using Gaussian mixture models and shifted delta cepstral features". Proc. ICSLP 02, 2002: p. 89-92.

[18] A. Hanani, M. J. Carey, M. J. Russell. Improved language recognition using mixture components statistics. In INTERSPEECH 2010, 2010, pp. 741-744.

[19] D. Martinez, O. Plchot, L. Burget, O. Glembek, and P. Matejka,“Language recognition in I-vectors space.” in Interspeech. ISCA, 2011, pp. 861–864.

[20] B. Pierre-Michel, A. Larcher, D. Matrouf, J. Bonastre, and O. Plchot. "Variance-spectra based normalization for I-vector standard and probabilistic linear discriminant analysis." In Odyssey: The Speaker and Language Recognition Workshop, 2012, pp. 157-164.

[21] I. Lopez-Moreno, Javier Gonzalez-Dominguez, D. Martınez, O. Plchot, J. Gonzalez-Rodriguez, and P. Moreno. Automatic language identification using deep neural networks. In Proceedings of ICASSP 2014, pages 5374–5378.