ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

SPEAKER RECOGNITION OF MAGHREB DIALECTS

Leila Beltaifa - Zouari [*1], Asma Chayeh [2]
[*] Computer Department, National Engineering School of Carthage, Tunisia

DOI: 10.5281/zenodo.1066198

ABSTRACT

A few studies have focused on the west Arabic (Maghreb) dialects for which resources are rare. To handle this problem, we devoped a web-based database of speech from Tunisian, Algerian and Moroccan speakers covering the diversity of Arabic dialects spoken in north Africa. Then speaker identification and verification experiments have been conducted in order to evaluate the performance of each dialect-based system. A baseline system using Timit database have also be developed for comparison purposes. The experiments show that the performances of dialectal speaker identification systems outperform those of the baseline identification system. The dialectal speaker verification systems performances are less better than the Timit-based sytem but are close to each other.

KEYWORDS: Arabic Dialects, Gaussian Mixture Models, Maghreb database, MFCC, Speaker recognition.

I. INTRODUCTION

Automatic Speaker recognition is the process of automatically extracting the identity of a speaker on the basis of the information included in his voice [3]. This technique is developed in order to make possible the control by voice of many services such as voice dialing, telephone shopping, banking transactions over a telephone network, information and reservation services, database access services, voice mail, remote access to computers, etc.
Speaker recognition can be divided into two applications which are speaker identification and speaker verification. Speaker identification is the process of finding the identity of a speaker from a group of registered speakers. Speaker verification is the process of accepting or rejecting a person claimed identity from his voice. Similarly, speaker recognition can also be divided into text-dependent and text independent methods [5][6]. In text dependent systems, speaker recognition relies on a specific text being spoken. In text independent systems, a speaker is recognized independent of what is saying. The text-dependent methods generally achieve higher recognition performance than the text-independent method since this method can directly exploit the voice individuality associated with each phoneme or syllable. The structure of text dependent recognition systems is, therefore, rather simple since in text-independent method, the system must be able to recognize the speaker from any text.[1],[2],[4].

Although speaker recognition has reached high performance in literature [6][7][10], real systems still face to two challenging problems : noise robustness and portability to new languages [17], in particular to the Dialectal Arabic language.

In fact, the Arabic language refers to the many existing varieties of Arabic. Those varieties include one spoken and written form, Modern Standard Arabic (MSA), and many spoken only forms named regional dialects. MSA is standardized, regulated, and taught in schools, and used in written communication and formal venues. The regional dialects are used for day-to-day dealings and spoken communication [16].

Today, the dialect Arabic is becoming used in interviews, news, telephone conversations, public services, etc., and it has a strong presence today in blogs, forums, and user/reader commentaries on the internet. So, it is important to consider the dialect Arabic in the new technologies like speech recognition systems, systems Human-Machine Dialogue, etc.[12]

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

The lack of recorded and annotated resources is one of the main problems of the Dialectal Arabic. For some dialects, especially Egyptian and Levantine, there are some investigations in terms of building corpora as for example the DARPA Babylon Levantine Arabic speech corpus which gathers four Levantine dialects spoken by speakers from Jordan, Syria, Lebanon, and Palestine [12] or the CALLFRIEND Egyptian database.

In the same time very, few attempts have considered Maghreb Arabic dialects [11][13][14]; TuDiCoI [16][9] is a spontaneous dialogue speech corpus of Tunisian dialect. It contains 20 hours of recorded dialogues between staff and clients in the railway of Tunisia. The ARADIGIT Algerian database is recorded by 1800 people and contains only digits [15]. KALAM'DZ is an Arabic Spoken corpus that covers all major Algerian dialects using web resources[12].

According to our study, there is neither existing corpora containg the three main Maghreb dialects (i.e. Tunisian and Algerian and Moroccan) recorded in the same conditions nor Web-based Maghreb corpus. Thus, we present in this paper the efforts we deployed to produce annotated resources for the Maghreb dialects. Then state of the art features and classification methods are used to develop and evaluate speaker recognition systems of the Maghreb dialects. The English Timit database is considered for comparison purpuses.

In the next section, we give a brief overview of the Arabic dialects. Section 3 is dedicated to the description of the proposed database. In Section 4, we present the experiments of speaker identification and verification using the Maghreb database and compare their performance to the baseline system. Conclusions are described in Section 5.

II. THE ARABIC LANGUAGE

Arabic is the sixth most widespread languages of the world. It is spoken by more than 420 millions of people in 60 countries of the world [12]. It has two major variants:

- Modern Standard Arabic (MSA): It is the official language of all Arab countries, so it is used in administrations, schools, official radios, and press.
- Regional Dialects (also named Dialectal Arabic or Arabic Dialects) : it is the spoken language of informal daily communication [12][18].

There are many linguistic differences between MSA and the regional dialects. Regional Dialects can be classified according to geographical areas into two groups namely the Western group (the Arab Maghreb or the North African group) and the Eastern group [18]. (Figure 1)

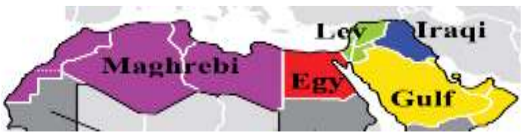


Figure 1. The arabic dialects classification [18]

a. Eastern dialects

The eastern dialects are essentially : Levantine Arabic, Gulf Arabic, Iraqi Arabic and Egyptian Arabic. Levantine is spoken by speakers from Jordan, Syria, Lebanon and Palestine. Gulf is close to MSA, perhaps because the current form of MSA evolved from an Arabic variety originating in the Gulf region. Iraqi is sometimes considered to be one of the Gulf dialects, though it has distinctive features of its own in terms of prepositions, verb conjugation, and pronunciation. Egyptian is spoken in Egypt. It is also the most widely understood dialect, due to a thriving Egyptian television and movie industry [18].

b. Maghreb dialects

The Maghreb is a large region with more variation than is seen in other regions such as the Levant and the Gulf because it is influenced by the French and Berber languages.So the Western most varieties could be unintelligible by speakers

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

from other regions in the Middle East [18]. The Maghreb Arabic dialects are mainly the Tunisian dialect, the Algerian dialect and the Moroccan dialect.

- The Tunisian Arabic, known as the Darija or Tounsi, is a set of the dialects used by urban residents, farmers and Bedouins (residents of the desert). These dialects differ considerably from each other in different levels : pronunciation, phonology, vocabulary, morphology, and syntax. We note the presence of French, Berber, Italian, Turkish, English, Spanish and also MSA words. The presence of these words is the result of many events that such as the Islamic invasions, the French colonization, migration, commercial exchanges, etc [3],[9].

- The main Algerian dialects are the Tamazight and the Arabic. The Algerian Arabic is spoken by about 80% of the population. It results from two Arabization processes due to the expansion of Islam in the 7th and 11th centuries, which lead to the appropriation of the Arabic language by the Berber population. According to both Arabization processes, Algerian Arabic dialects can be divided into two major groups: Pre-Hilali and Bedouin dialects. Both dialects are different by many linguistic features [12],

- Moroccan Darija Arabic and Berber are languages spoken in homes and on the street. Moroccan Darija is intelligible to some extent with Algerian Arabic and to a lesser extent with Tunisian Arabic. It has been heavily influenced by Berber, French, and Spanish. Darija is spoken as a first language by approximately half of Morocco's population. The other half speaks one of the three Berber languages (either Tarifit, Tachelhit or Tamazight).

III. DATABASE DESCRIPTION

Nawadays, MSA content significantly dominates dialectal content, as MSA is the variant of choice for formal and official communication and the vast majority of published Arabic is in MSA. As a result, MSA's dominance is also apparent in datasets available for linguistic research [12]. Actuelly, there is no Magreb database containing the three Arabic dialects recorded in the conditions, so we developed our TV Magreb database. In this paragraph we will desribe our Maghreb database and also Timit which is the English database used in the baseline system. The baseline system, is developed and evaluated using the protocol described in [2]. The Maghreb systems have been developed and tested using the same experimental protocol as the baseline system to be compared to the state of the art systems.

4.1. Timit description

TIMIT (Texas Instruments Massachusetts Institute of Technology) database allows automatic speaker verification to be done under almost ideal conditions. The TIMIT database is composed of 10 sentences that are spoken by 630 speakers (70 % male and 30 % female) from 8 different dialect regions in America. The dialects are New England, Northern, North Midland, south Midland, southern, New York City, Western and My Brat.

The speech is recorded using a high quality microphone in a sound proof booth at a sampling frequency of 16 kHz, with no session interval between recordings. The speech is designed to have a rich phonetic content, which consists of 2 dialect sentences (SA), 450 phonetically compact sentences (SX) and 1890 phonetically diverse sentences (SI)[8].

4.2. Maghreb database

The new Maghreb speech database (containing Tunisian, Algérian and Moroccan dialects) is collected from the Tunisian television shows "Ness Nessma News" (between 2011 and 2016) and has a duration of 108 minutes. "Ness Nessma News" is a Maghreb variaty show where the presentator invite famous people (such politicians and musicians, ..) from the the Maghreb countries.

Following the experimental protocol described in [2], we consider only 10 speakers for each database (Tunisian, Algerian and Moroccan). The speakers represent a group of native male and female speakers who speak the same dialect, ranged at the age of 25 to 65 years old. Speakers have different professions where their collected speech is totally used for public speaking mostly for those who are politicians and musicians.

After manually annoting the three bases, each speaker has 10 recordings (WAV files) in which 8 records are used to training and 2 for the test. All the speech files are sampled to 16 Khz.

Table 1: Maghreb database description

Sampling Frequency	16 kHz
Number of speakers	30
Number of dialects	3
Age of speakers	Between 25 to 65 years
Gender	6 Female and 24 Male

IV. EXPERIMENTS AND RESULTS

The state-of-the-art speaker recognition systems are based on Gaussian Mixture Models (GMMs). The speech signal taken from the person is transformed into MFCC (Mel frequency Cepstral coefficient) feature vectors. The distribution of the MFCC feature vectors is modeled with a Gaussian mixture density. Hence, the architecture of the automatic speaker recognition is as follows :

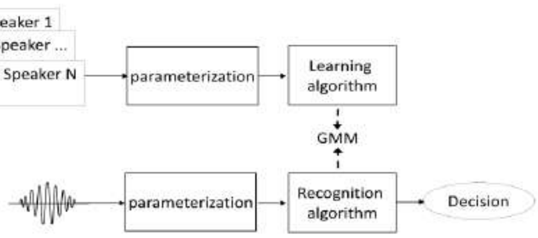


Figure2. The architecture of the automatic speaker recognition system [10]

To assess the performances of automatic speaker verification system for both Arab and English languages, we used the TIMIT English database and our corpus of Maghreb database.

For speaker identification we report the Identification Rate (IR). Results of speaker verification are reported with Detection Error Tradeoff (DET) curve and Equal Error Rate (EER). The DET graph shows the tradeoff between the False Acceptance Rate (FAR) and the False Rejection Rate (FRR).The EER is obtained at the intersection of DET curve and the affine function f(x) =x, it indicate the point at which the false acceptance equals the false rejection [6].

5.1 Baseline system

The ten utterances of each speaker were divided into 8 utterances for training (two SA, three SX and three SI sentences) and the remaining 2 utterances (two SX sentences) for the test task.

5.1.1. Speaker identification

For each dialect, we vary the Gaussian number and report the mean value of the Identification Rate (IR) in Table2.

Table2: The identification rate for 10 speakers and 8 dialects

Gaussian number	4	8	16	32	64
IR (%)	89,00	96,00	97,67	98,67	98,67

We can see that the best results are obtained with 32 Gaussians. For 32 Gaussians, the identification rate for each dialect is reported in Table3.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Table3: The identification rate for English dialects using 32 Gaussians

Dialect region	IR (%)
New England	100
Northern	100
North midland	100
South midland	100
Southern	100
New york city	90
Western	100
Army Brat	100

These results are similar to the state of the art systems using Timit database and the same experimental protocol.

5.1.2. Speaker verification

For different number of Gaussians (16, 32 and 64), we draw the DET curves and we report the Equal Error Rate (EER).

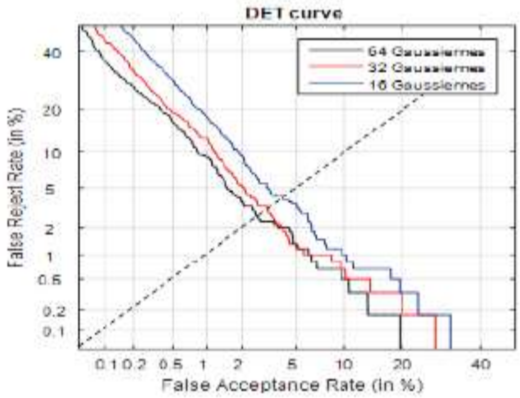


Figure3. EER for different number of mixtures for TIMIT database

As Figure3 shows, the EER are constantly improved as number of GMM components increases. The best performance corresponds to EER = 1.11%.

5.2 Maghreb systems

The goal of these experiments is to evaluate the performance of Tunisian, Moroccan and Algerian speaker recognition systems and to compare them to the baseline system.

5.2.1. Tunisian database

Following the same experimental protocol [2] as the baseline system, each speaker have 8 utterances to be used for the models training and 2 utterances for the tests.

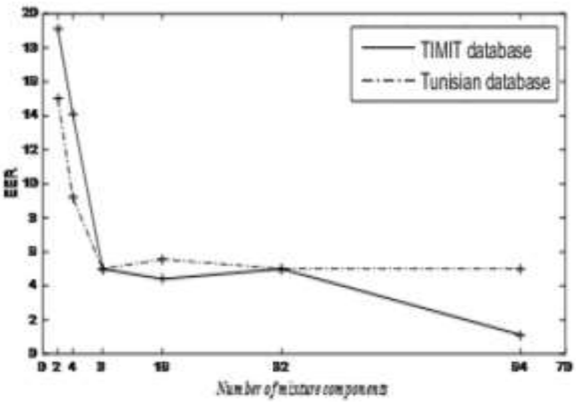
Speaker identification and verification experiments have been conducted using GMMs with 4, 8, 16, 32 and 64 Gaussians.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Table4: The identification rate for the Tunisian speakers

Gaussian number	4	8	16	32	64
IR (%)	95	100	100	100	100
EER (%)	14,8	9	5,5	5	5

As it is shown in Table4, the identification rate is good enough and is close to the Timit one. Concerning the speaker verification, the EER is decreasing to reach 5.0 % at 32 Gaussian mixtures.



In Figure 4, we report the EER for the English and the Tunisian databases.

Figure 4. EER for TIMIT and the Tunisian databases

For only 16 gaussians, the Tunisian verification system outperforms the English one. For 32 Gaussians, the two systems have the same results. With 64 components mixture, the Timit based system EER continues to decline to reach 1.11% while the Tunisian system EER is 5%.

5.2.2. Moroccan database

The same experiments of speaker identification and verification have been conducted with the Moroccan database. The identification rates are detailed in Table5.

Table5: The identification rate for the Moroccan speakers

Gaussian number	4	8	16	32	64
IR (%)	100	100	100	100	100

Table5 shows that the identification performance (IR) of the Moroccan system are better than those of the Timit one .

The EER of the Morrocan and the Timit based systems are reported in Figure5.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

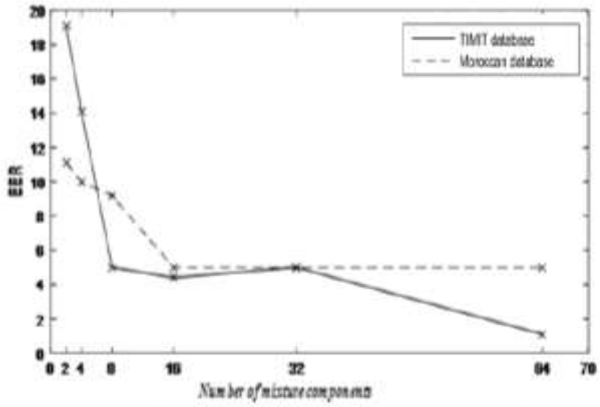


Figure 5. EER for TIMIT and the Moroccan databases

As it is mentioned in Figure5, the EER for Moroccan database decreases from 11.11 % for 2 Gaussians to be steady from 32 Gaussians at 5.0%,

5.2.3. Algerian database

Speaker identification and verification experiments are conducted with the Algerian database using the same protocol as the other Maghreb databases and Timit database.

Table6: The identification rate for the Algerian speakers

Gaussian number	4	8	16	32	64
IR (%)	100	100	100	100	100

We can see (Table6) that the Algerian system outperforms the baseline system. The speaker verification EERs are reported in Figure6.

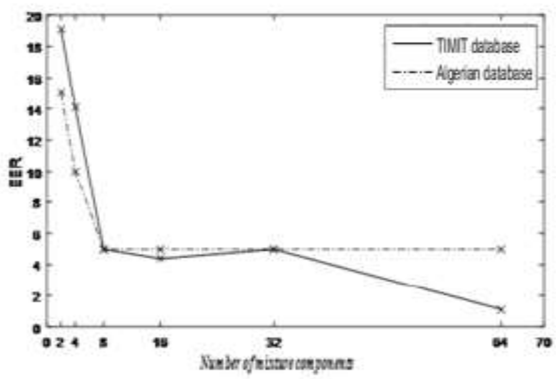


Figure 6. EER for TIMIT and the Algerian databases

For the Algerian dialect (figure.6), we notice that the EER was reduced from 15.0 % for 2 Gaussians to be stable at 32 and 64 components densities at 5.0%.

These performances losses (compared to the baseline system) can be explained by the effect of transmission degradation. In fact, TIMIT is a microphone speech but Maghreb databases contain speech broadcast on television and transmitted via IP (Internet Protocol).

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

V. CONCLUSION

Recently, dialect Arabic became a mean of communication on the Web, in chat rooms, in social media, in TV shows, etc. So the need of annotated resources for dialects to be used by human-machine interaction systems icreases.

In this paper, we have proposed a Maghreb dialectal database containing annotated speech recordings from Tunisian, Algerian and Moroccan speakers. Then, text independent speaker recognition systems for the Maghreb Arabic Dialects have been developed. In order to evaluate the performance of automatic Arabic speaker identification and verification systems, a baseline system have been developed and evaluated in the same conditions. The baseline recognition (identification and verification) system use TIMIT database.

Experiments results show that the dialectal speaker identification systems outperform the baseline system. But the dialectal speaker verification systems (Tunisian, Algerian and Moroccan) are less performant than the Timit-based system. We notice also that the ERR of the three dialect-based systems are close to each other. This can be explained by the fact that these dialects have the same origins which are mainly the MSA, the French, the Turkich and the Berber.

VI. REFERENCES

[1] G. Adda, G. Chollet, S. Essid, T. Fillon, M. Garnier-Rizet, c. Hory and L. Beltaifa-Zouari. Traitement des modalités «audio» et « parole» In M. Campedel and P. Hoogstoel. Sémantique et multi modalité en analyse de l'information. Hermès, 2011, 143-188p. ISBN : 978-2-7462-3139-9.

[2] D. A. Reynolds and R.C. Rose. Robust Text-Independent Speaker Identification Using Gaussian Mixture speaker Models. IEEE Transactions on speech and Audio Processing, january 1995. Vol.3, No.1.

[3] S. G. Bagul and R. K. Shastri. Text independent speaker recognition system using GMM. Human Computer Interactions (ICHCI), 2013.

[4] H. C. Bao and Z. C. Juan. The research of speaker recognition based on GMM and SVM. International Conference on System Science and Engineering (ICSSE), Dalian, Liaoning, 2013. pp. 373-375.

[5] J. C. Wu, A. Martin, C. Greenberg and R. Kacker. The Impact of Data Dependency on Speaker Recognition Evaluation. IEEE/ACM Transactions on Audio, Speech, and Language Processing.

[6] P. Bansal, S. A. Imam and R. Bharti. Speaker recognition using MFCC, shifted MFCC with vector quantization and fuzzy.International Conference on Soft Computing Techniques and Implementations (ICSCTI), Faridabad, 2015. pp. 41-44.

[7] Y. Yang and L. Deng. Score regulation based on GMM Token Ratio Similarity for speaker recognition. Chinese Spoken Language Processing (ISCSLP), 2014. pp. 424-424.

[8] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett and N. Dahlgren. TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM. National Institute of Standards and Technology. 1990

[9] A. Masmoudi, Y. Estève, M. Khmekhem, F. Bougares and L. Hadrich Belguith. Phonetic tool for the Tunisan arabic. SLTU-2014, St. Petersburg, Russia, 14-16 May 2014

[10] L. Beltaifa-Zouari and R. Chakroun. Bi-level Speaker Identification System: Impact on Short Utterances. International Journal of All Research Education & Scientific Methods (IJARESM). ISSN No: 2455-6211. Volume 5, Issue 8, August, 2017.

[11] N. Lachachi and A. Adla. GMM-Based Maghreb Dialect Identification System. Journal of Information Processing systems. 2015 pp22-38,

[12] S. Bougrine, A. Abdallah Lakhdari and H. Cherroun. Toward aWeb-based Speech Corpus for Algerian Arabic Dialectal Varieties. Proceedings of The Third Arabic Natural Language Processing Workshop (WANLP), pages 138–146, Valencia, Spain, April, 2017.

[13] M. Belgacem, G. Antoniadis and L.Besacier. Automatic Identification of Arabic Dialects. Proceedings LREC 2010.

[14] A. Shoufan and S. Al-Ameri. Natural Language Processing for Dialectical Arabic: A Survey.Proceedings of the Second Workshop on Arabic Natural Language Processing, pages 36–48, Beijing, China, July 26-31, 2015.

[15] K. Zargat, A. Amrouche, M. Fedila and M. Debyche. Robust Arabic speaker verification system using LSF extracted from the G.729 bitstream. IEEE international workshop on machine learning for signal processing. Southampton. UK September 2013.

[16] M, Graja, M. Jaoua, and L Hadrich-Belguith. Lexical Study of A Spoken Dialogue Corpus in Tunisian Dialect. In The International Arab Conference on Information Technology (ACIT),2010. Benghazi, Libya.

[17] M. Graciarena, S. Kajarekar, A. Stolcke and E. Shriberg. Nise Robust Speaker Identification for Spontaneous Arabic Speech. ICASSP 2007.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

[18] O. F. Zaidan and Ch. Callison-Burch. Arabic Dialect Identification. Association for Computational Linguistics 2012

CITE AN ARTICLE

Zouari , L. B., & Chayeh, A. (2017). SPEAKER RECOGNITION OF MAGHREB DIALECTS. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY*, *6*(11), 413-421.