Mixed ASR System for Amazigh and Arabic Under-Resourced Dialects in Maghreb Region

Mohamed Hamidi[1], Hassan Satori[1], Ouissam Zealouk[1], Khaled Lounnas[2, 3], Mourad Abbas[2], Mohamed Lichouri[2] and Hocine Teffahi[3]

[1]LISAC -FSDM-USMBA University, Fes, Morocco
[2] Computational linguistics Dept-CRSTDLA, Algiers, ALGERIA
[3] LCPTS-USTHB University, Algiers, ALGERIA
{mohamed.hamidi.5@gmail.com, hsatori@yahoo.com, ouissam.zealouk@gmail.com, klounnas@usthb.dz, m.abbas@crstdla.dz, m.lichouri@crstdla.dz, hteffahi@gmail.com}

Abstract. Automatic Speech Recognition (ASR) technology plays an essential role in human-machine interaction. In this paper, we describe our conducted speech experiments that are realized to develop and adapt a mixed automatic speech recognition system based on spoken digits for Amazigh and Arabic dialects that are considered as under-resourced dialects in the Maghreb region. Our used database includes speech samples collected from 24 Moroccans and Algerian speakers including both males and females. The designed system is implemented based on the combination of hidden Markov models and Gaussian mixture models, as well as the Mel frequency spectral coefficients (MFCCs) feature extraction method.

Keywords: Speech recognition; Mixed dialects; HMMs; MFCCs; GMMs.

1 Introduction

Automatic Speech Recognition (ASR) is an interdisciplinary subfield of computational linguistics that creates systems that enable the recognition and translation of spoken words into text by machines. ASR is a multidisciplinary technique; it includes knowledge and research in computer science, linguistics and engineering fields.

The last few years have seen tremendous advances in language and dialect recognition that are starting to gain the attention of speech technology research, especially speech recognition for regional dialect. A wide range of work on recognition of spoken digits and alphabets is being performed in the field of ASR and many algorithms have been developed using a variety of technologies. In general, ASR researchers have targeted alphabets and spoken numbers for different languages. Most of the research has been done in English, Japanese and Mandarin [1], but very little research can be found in Amazigh and Arabic dialects that are Under-Resourced dialects in Maghreb region.

In this paper, we aim to create a mixed ASR system that allows recognizing the ten fist spoken digits for Amazigh and Arabic dialects used in Morocco and Algeria that

are considered as under-resourced dialects in the Maghreb region. In our realization, we use the static approach with hidden Markov Models combined to Gaussian mixture models.

This paper is organized as follows, an introduction in Section 1. Section 2 presents the related works. Section 3 presents the used materials and methods. Section 4 shows the system preparation. Experimental Results are given in Section 5. Finally, the conclusion in Section 6.

2 Related works

Table 1. Some research on Amazigh speech recognition systems

Authors	Description	Method	Results
Zealouk et al., [2]	Evaluation of Amazigh digits (0-9) under car and grinder noisy environments	HMM-GMM	72.92% with 5 dB and 0.17% with 35 dB
Hamidi et al, [3]	Study of the performance of the Amazigh interactive digital voice recognition system in a noisy train environment	HMM-GMM	72,43% with 3 dB and 0% with 39 dB
Barkani et al., [4]	Development of an ASR Amazigh control system with Raspberry Pi board	HMM-GMM	90.43%
Zealouk et al., [5]	Detecting the differences between the voice of normal and pathological speakers based on the ASR system.	HMM-GMM	84.00% for normal speakers 27.50% for pathological speakers
Hamidi et al, [6]	Amazigh speech recognition system via interactive voice response (IVR)	HMM-GMM	89.64%
Addarrazi et al, [7]	Recognizing the lip movement for lip-reading system	DCT-HMM	84.99%
Telmem et al., [8]	Build an Amazigh speech recognition system	HMMs and CNN	92%
Addarrazi et al, [9]	Audio visual speech recognition system	Viola-Jones approach	99% for face detection. 96.6% for mouth detection.
Satori et al., [10]	Create an ASR system that allows detecting the smoker speakers.	HMM-GMM	The recognition rate for smokers is 50% lower than for non-smokers.
Satori et al., [11]	Create an Amazigh speech recognition system based on digits and alphabets	HMM-GMM	92.89 %

The majority of ASR's previous work in both Amazigh and Arabic has focused on the official languages known as Amazigh Language (AL) and Modern Standard Arabic (MSA). However, we find that these languages are not the language of Daily

communication in some countries, such as Morocco and Algeria, where other types called dialects are used, which are a less researched area. Tables 1 summarizes some ASR previous work for Amazigh language. Tables 2 presents some ASR previous work for Arabic language.

Table 2. Some research on Arabic speech recognition systems

Authors	Description	Method	Results
Meftah et al, [12]	Emotional Speech Recognition by using Arabic language	MLP-SVM	54.07% and 84.14%
Eljawad et al, [13]	Speech Recognition of Arabic language based on Fuzzy Logic and Neural Network	Fuzzy logic and neural network	77.1% and 94.5%
Elharati et al, [14]	MFCC and HMMs based Arabic ASR system	MFCC-HMMs	92.92%
Alshayeji et al, [15]	Studying the effect of diacritics on Arabic ASR systems.	DNN	from 4.68% to 42%
Alsharhan et al, [16]	Enhancement of Arabic ASR system based on the automatic generation of accurate audio text.	HMM	71%
Frihia et al, [17]	Building a large vocabulary continuous speech recognition system	SVM-HMM	0.05%
Khelifa, et al, [18]	Building a teaching and learning system based on Arabic ASR technology	HMM-GMM	97%
Wahyuni, [19]	Recognize spoken Arabic letters	ANN	92.42%
Satori et al, [20]	Build an Arabic Automated Speech Recognition System	HMM	96,67%

In [21] authors have presented a novel method that combines the automatic speech recognition and language identification systems. Their work aims to develop a speech system that allows identifying the used dialect and recognizing the spoken digits based on modern standard Arabic and Amazigh Moroccan language. Their system was based on the SVM and HMM methods and their findings present that the proposed system performs 33% better than an ordinary speech recognition system. Researchers in [22] have described an automatic speech recognition system based on Darija Moroccan Dialect. The implemented system is able to recognize the ten first Darija digits collected from 20 speakers (males and females). In their work, the HMM-GMM combination system was used with MFCC feature extraction method and their best-obtained rate is 96.27 %.

3 Materials and methods

In this section, we describe our used Automatic speech recognition system, extraction technique and modeling algorithms.

3.1 Automatic Speech Recognition

Speech recognition [23] is the process of decoding speech signals picked up by a microphone and converting them into words. These words can be exploited as commands, data input, or application control. Recently, this technology-based applications are often found in many fields like military, commercial, industrial, healthcare, telephony, etc.Fig. 1 presents the structure of ASR system. Recently, for Moroccan Amazigh ASR systems were targeted by our lab researchers [24-28] ;

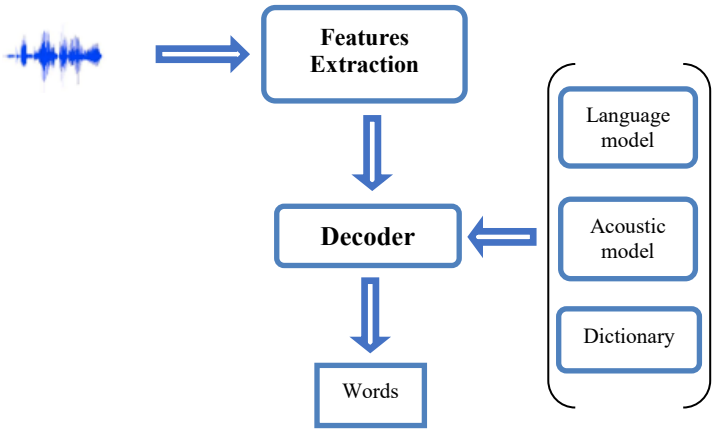

Fig. 1. ASR system Architecture

3.2 Hidden Markov model

The hidden Markov model (HMM) [29] is a statistical modeling method, his structure includes a finite ensemble of states and each one is associated with a probability distribution where the transition probabilities govern the transitions among the states. Fig. 2 presents a case of 3 states of the Hidden Markov Model.

HMMs have efficient learning algorithms that allow consistent handling of insertion and deletion penalties. They can also handle variable-length entries. However, HMMs

have a large number of unstructured parameters and produce a number of sub-optimal modeling assumptions that limit their effectiveness

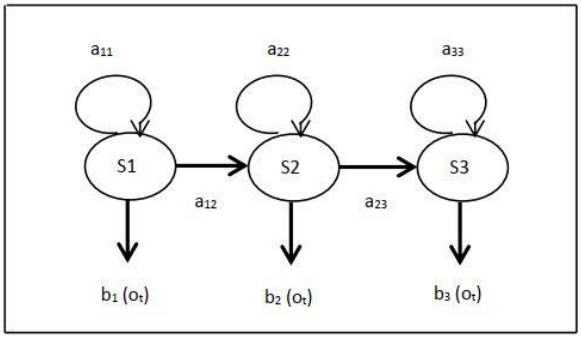

Fig. 2. Three states HMM architecture

3.3 Mel Frequency Cepstral Coefficients (MFCC)

The extraction of Mixed Frequency Cepstrum Coefficients (MFCC) [30] comprises a frame-by-frame analysis of an input speech where the speech signal is segmented into a sequence of frames. Each frame allows a Fast Fourier Transform (FFT) to generate certain parameters, which are then subjected to a Mel perception scale and decorrelation. The result is a sequence of feature vectors describing a useful logarithmically compressed amplitude and simplified frequency information. Fig. 3 shows the process of the Mel frequency ceptral coefficient technique.

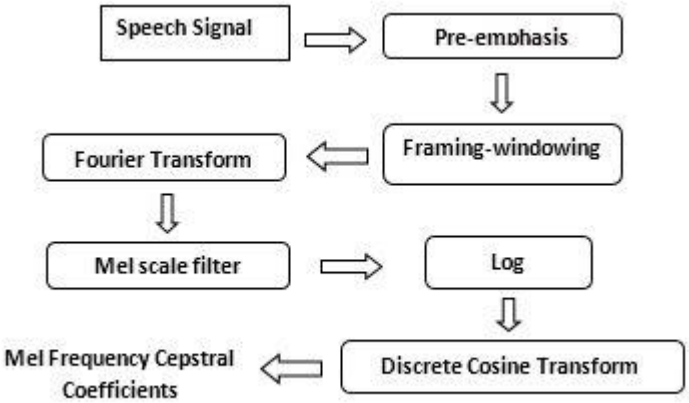

Fig. 3. The MFCC architecture

4 Mixed System Preparation

Our proposed mixed speech system is implemented on Ubuntu 14.04 LTS operating system and the used hardware is laptop with Intel Core i5 CPU of 2.4 GHz speed and RAM 4G.

4.1 Corpus Preparation

In the field of automatic Amazigh speech recognition, voice databases are rare. All the works published in this area have been tested on databases that were created in our lab as part of this work. Our vocal corpora are recorded using a microphone using the WaveSurfer recording tool in .wav format. Our database includes four corpuses consist of ten separate spoken digits collected from 24 Moroccan and Algerian speakers (male and female) aged 18 to 50 years. The recording sample rate is 16 kHz, with a resolution of 16 bits. During the recording sessions, speakers were asked to pronounce the digits sequentially. Each speaker's recordings have been saved in a ".wav" file. During the recording session, each file was replayed to ensure that all words were included in the recorded signal. Wrongly spoken recordings were ignored and only correct recordings are kept. Table 3 presents the audio database information. Table 4 presents the used digits with four dialects.

Table 3. Database description

Corpus	Parameter	Value
Am_digits	Description	Moroccan Amazigh dialect
	Speakers	6 (four for training and two for testing)
Dar_digits	Description	Moroccan Darija dialect
	Speakers	6 (four for training and two for testing)
Al_digits	Description	Algerian Darija dialect
	Speakers	6 (four for training and two for testing)
Kb_digits	Description	Algerian Kabyle dialect
	Speakers	6 (four for training and two for testing)

Table 4. The used Ten first digits

Number	Amazigh Dialect	Darija Dialect	Kabyle Dialect	Algerian Dialect
0	ILEM	SIFR	OULECH	SIFER
1	YEN	WAHD	YEWAN	WAHED
2	SIN	JOJ	SIN	ZOUJ
3	KRAD	TLATA	THYATHA	TLATHA
4	KUZ	RABAA	REBAA	RBAA
5	SEMUS	KHAMSA	KHEMSA	KHMSA
6	SEDISS	STTA	SETSA	STTA
7	SA	SBAA	SEBAA	SEBAA
8	TAM	THMANYA	THMANIA	THMANYA
9	TZA	TSAAOD	TESSAA	TESAA

4.2 Acoustic Models Preparations

Our acoustic models include the representations of ten first spoken digits created in the training phase by using audio data. To design our acoustic model, we gathered a set of input data and processed it using the SphinxTrain tool as shown in Figure 4. The following list shows the input data and the files used.

- A set of audio data (specific data for training).
- Configuration files that exclude the relationship between training and testing of text and audio files
- Language model gives a representation of the probability of occurrence for each used digit.
- The dictionary determines the pronunciation of the used digits

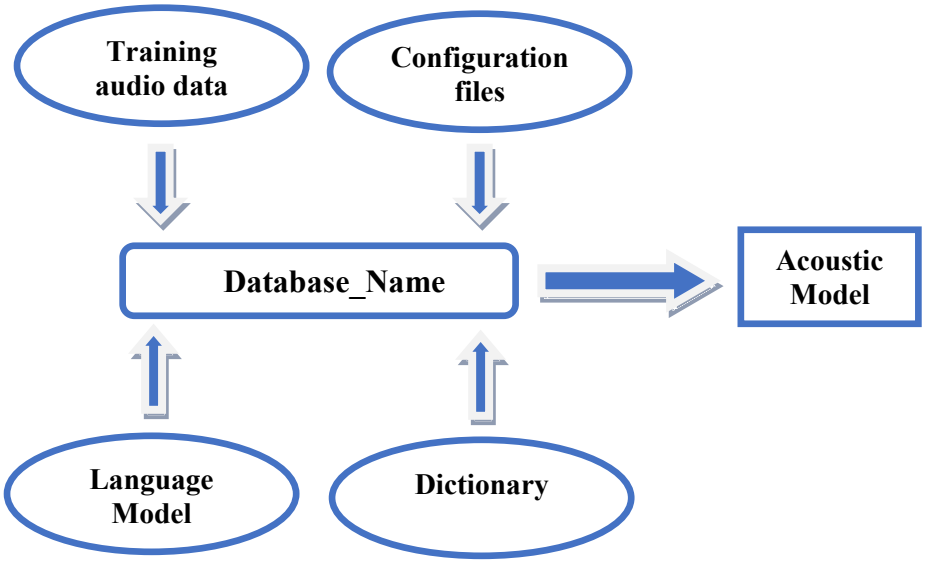

Fig. 4. Acoustic model preparation

4.3 Pronunciation Dictionary

The dictionary performs the role of the intermediary between the Acoustic Model and Language Model. In our work, we have designed a mixed dictionary that includes the different used dialects. In this work, we have created two pronunciation dictionaries; the first one is com_dictionary (See Table 4) which includes 40 digits and their pronunciations where we have combined the dictionaries of four dialects that are Amazigh dialect, Darija dialect, Kabyle dialect and Algerian dialect. The second proposed dictionary is mix_dictionary (See Table 5) contains 32 words where we have designed a mixed and inclusive dictionary that includes digits with two or more variants where the alternate transcriptions marked with parenthesis like (1) stand for second pronunciation. For example, the word YEN appears with two variants YEN and YEN (1) whose phonetic descriptions are respectively YEN and Y A N. So, the same method we will apply with the digits which have the same or close pronunciation such as THEMANYA digit in Algerian dialect and TMANIA digit in Kabyle dialect.

Table 4. The combination of the separated used dictionaries

ILEM	I L E M	RABAA	R A B AA	SABAA	S A B AA
YEN	Y E N	KHAMSA	KH A M S A	THMANIA	TH M A I A
SIN	S I N	STTA	S TT A	TESSAA	T E SS AA
KRAD	K R A D	SBAA	S B AA	SIFER	S I F E R
KUZ	K U Z	THMANYA	TH M A N Y A	WAHED	W A H E D
SEMUS	S E M U S	TSAAOD	T S AA O D	ZOUJ	Z O U J
SEDISS	S E D I SS	OULECH	OU L E CH	TLATHA	T L A TH A
SA	S A	YEWAN	Y E W A N	RBAA	R B AA
TAM	T A M	SIIN	S I I N	KHMSA	KH M S A
TZA	T Z A	THYATHA	TH Y A TH A	STTA	S TT A
SIFR	S I F R	REBAA	R E B AA	SEBAA	S E B AA
WAHD	W A H D	KHEMSA	KH E M S A	THMANYA	TH M A N Y A
JOJ	J O J	SETSA	S E T S A	TESAA	T E S AA
TLATA	T L A T A				

Table 5. The proposed mixed dictionary

ILEM	I L E M	ZOUJ	Z O U J
YEN	Y E N	TLATA	T L A T A
YEN(1)	Y A N	TLATA(1)	T L A TH A
SIN	S I N	REBAA	R A B AA A
KRAD	K R A D	REBAA(1)	R E B AA
KUZ	K U Z	KHAMSA	KH A M S A
SEMUS	S E M U S	KHAMSA(1)	KH E M S A
SEDISS	S E D I SS	SETTA	S TT A
SA	S A	SETTA (1)	S E TT A
TAM	T A M	SEBAA	S A B AA
TZA	T Z A	SEBAA(1)	S E B AA
OULECH	OU L E CH	THEMANYA	TH M A N Y A
YEWAN	Y E W A N	THEMANYA(1)	T M A N I A
WAHED	W A H E D	TSAAOD	T S AA OU D
JOJ	J OU J	TESAA	T E S AA

5 Experimental Results

In order to create a mixed speech recognition system, that includes the ten first digits of Moroccan Amazigh, Algerian Kabyle, Moroccan Darija and Algerian Darija dialects, we have conducted several experiments with 3 and 5 HMM states and different Gaussian Mixture Models (4, 8, 16, 32 and 64). The recognition rates were observed and recorded for each experiment. Figure 5 presents the recognition rates of combined-system, which is based on com_dictionary, and mixed-system that is based on the mix_dictionary.

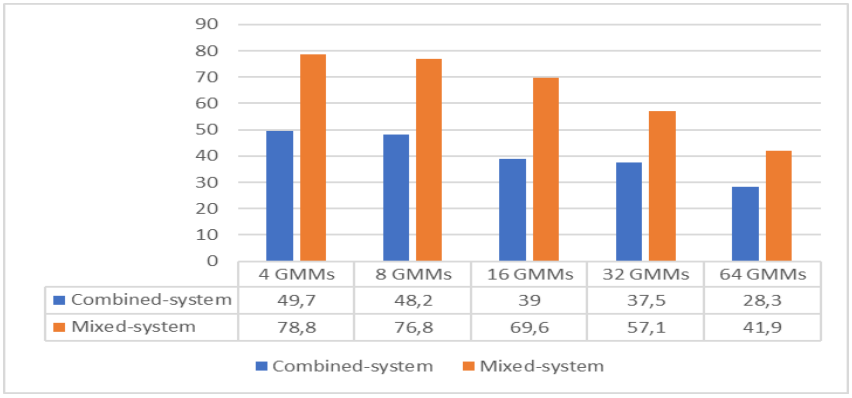

Fig. 5. The recognition rates of combined-system and mixed-system systems based on 3 HMMs and several GMMs values

The differences between two systems are 29.1, 28.6, 30.6, 19.6 and 13.6% are found for using 4, 8,16, 32 and 64 GMMs, respectively. It is found that 4 Gaussian mixture distributions obtained the best recognition rate of 78,8 %.

For mixed system, the best recognition rate is 78,8 % was found with 4 GMMs and the lower rate is 41,9% found with 64 GMMs. For combined system, the heigher rate is 49,7 % was observed with 4 GMMs and the lower rate is 28,3% found with 64 GMMs.

Figure 6 illustrates the recognition rates of mixed system with 5 HMMs, the obtained results are 75.8, 70.8, 67.8, 39.2 and 20.6 % found with using 4, 8,16, 32 and 64 GMMs, respectively. The best obtained recognition rate is 75.8 % found with 4 GMMs.

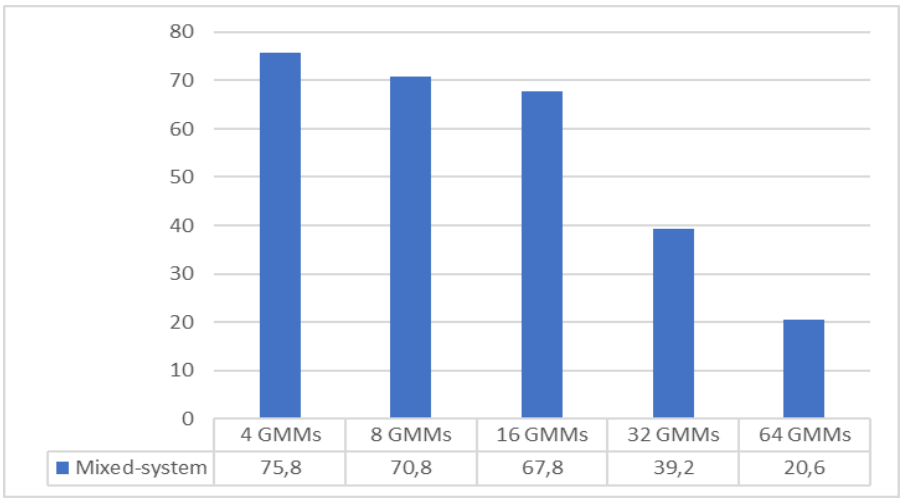

Fig. 6. The recognition rates of mixed-system based on 5 HMMs and different GMMs values

We observe the drop of recognition rates with the augmentation of GMMs value. Also we have observed that the substitution words increase with drop of recognition rates and the augmentation of GMMs values. Figure 7 shows the evolution of substitution words with different GMMs values.

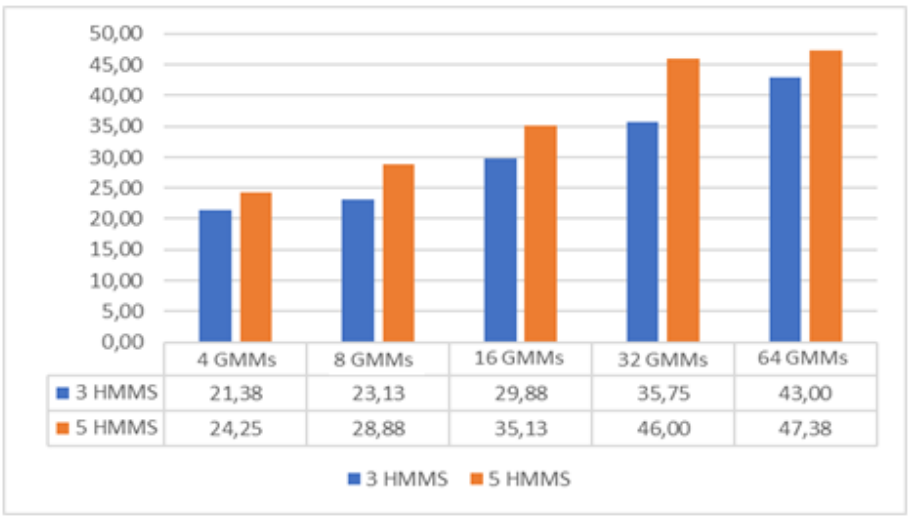

Fig. 7. The evolution of substitution words based on 3 and 5 HMMs and different GMMs values.

As a comparison, our results are lower than of those [11] because in their implementation, they use only one language. On the other hand, our results are higher than of those [21].

6 Conclusion

In this paper, we have presented a mixed speech recognition system that includes four dialects, which are Moroccan Amazigh, Algerian Kabyle, Moroccan Darija and Algerian Darija dialects. We have designed a mixed system based on the combination of hidden Markov models and Gaussian mixture models. We have proposed an effective approach to improve the accuracy of a mixed system based on the pronunciation dictionary. Our best obtained result is 78,8 % found with 3 HMMs and 4 GMMs and Our technique made a difference of around 29.1% from an ordinary method.

In the future, we will try to improve our obtained results by using large data and more ASR parameters. In addition, we will try to develop a mixed speech recognition system based on several dialects and languages.

References

1. Alotaibi, Y., Mamun, K., & Ghulam, M. (2009, July). Noise Effect on Arabic Alphadigits in Automatic Speech Recognition. In IPCV (pp. 679-682).
2. Zealouk, O., Satori, H., Laaidi, N., Hamidi, M., & Satori, K. (2020). Noise effect on Amazigh digits in speech recognition system. International Journal of Speech Technology, 1-8.
3. Hamidi, M., Satori, H., Zealouk, O., & Satori, K. (2020). Amazigh digits through interactive speech recognition system in noisy environment. International Journal of Speech Technology, 23(1), 101-109.
4. Barkani, F., Satori, H., Hamidi, M., Zealouk, O., & Laaidi, N. (2020, April). Amazigh Speech Recognition Embedded System. In 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET) (pp. 1-5). IEEE.
5. Zealouk, O., Satori, H., Hamidi, M., & Satori, K. (2020). Pathological Detection Using HMM Speech Recognition-Based Amazigh Digits. In Embedded Systems and Artificial Intelligence (pp. 281-289). Springer, Singapore.
6. Hamidi, M., Satori, H., Zealouk, O., & Satori, K. (2020). Interactive Voice Application-Based Amazigh Speech Recognition. In Embedded Systems and Artificial Intelligence (pp. 271-279). Springer, Singapore.
7. Addarrazi, I., Satori, H., & Satori, K. (2020). Lip Movement Modeling Based on DCT and HMM for Visual Speech Recognition System. In Embedded Systems and Artificial Intelligence (pp. 399-407). Springer, Singapore.
8. Telmem, M., & Ghanou, Y. (2020). A Comparative Study of HMMs and CNN Acoustic Model in Amazigh Recognition System. In Embedded Systems and Artificial Intelligence (pp. 533-540). Springer, Singapore.

9. Addarrazi, I., Satori, H., & Satori, K. (2017, April). Amazigh audiovisual speech recognition system design. In 2017 Intelligent Systems and Computer Vision (ISCV) (pp. 1-5). IEEE.
10. Satori, H., Zealouk, O., Satori, K., & Elhaoussi, F. (2017). Voice comparison between smokers and non-smokers using HMM speech recognition system. International Journal of Speech Technology, 20(4), 771-777.
11. Satori, H., & Elhaoussi, F. (2014). Investigation Amazigh speech recognition using CMU tools. International Journal of Speech Technology, 17(3), 235-243.
12. Meftah, A. H., Qamhan, M., Alotaibi, Y., & Selouani, S. A. (2020, February). Emotional Speech Recognition Using Rhythm Metrics and a New Arabic Corpus. In 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA) (pp. 57-62). IEEE.
13. Eljawad, L., Aljamaeen, R., Alsmadi, M., Almarashdeh, I., Abouelmagd, H., Alsmadi, S., ... & Alazzam, M. (2019). Arabic Voice Recognition Using Fuzzy Logic and Neural Network. ELJAWAD, L., ALJAMAEEN, R., ALSMADI, MK, AL-MARASHDEH, I., ABOUELMAGD, H., ALSMADI, S., HADDAD, F., ALKHASAWNEH, RA, ALZUGHOUL, M. & ALAZZAM, MB, 651-662.
14. Elharati, H. A., Alshaari, M., & Këpuska, V. Z. (2020). Arabic Speech Recognition System Based on MFCC and HMMs. Journal of Computer and Communications, 8(03), 28.
15. Alshayeji, M., & Sultan, S. (2019). Diacritics effect on arabic speech recognition. Arabian Journal for Science and Engineering, 44(11), 9043-9056.
16. Alsharhan, E., & Ramsay, A. (2019). Improved Arabic speech recognition system through the automatic generation of fine-grained phonetic transcriptions. Information Processing & Management, 56(2), 343-353.
17. Frihia, H., & Bahi, H. (2017). HMM/SVM segmentation and labelling of Arabic speech for speech recognition applications. International Journal of Speech Technology, 20(3), 563-573.
18. Khelifa, M. O., Elhadj, Y. M., Abdellah, Y., & Belkasmi, M. (2017). Constructing accurate and robust HMM/GMM models for an Arabic speech recognition system. International Journal of Speech Technology, 20(4), 937-949.
19. Satori, H., Hiyassat, H., Haiti, M., & Chenfour, N. (2009). Investigation Arabic Speech Recognition Using CMU Sphinx System. International Arab Journal of Information Technology (IAJIT), 6(2).
20. Wahyuni, E. S. (2017, November). Arabic speech recognition using MFCC feature extraction and ANN classification. In 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE) (pp. 22-25). IEEE.
21. Lounnas, K., Satori, H., Teffahi, H., Abbas, M., & Lichouri, M. (2020, April). CLIASR: A Combined Automatic Speech Recognition and Language Identification System. In 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET) (pp. 1-5). IEEE.
22. Ezzine, A., Satori, H., Hamidi, M., & Satori, K. (2020, June). Moroccan Dialect Speech Recognition System Based on CMU SphinxTools. In 2020 International Conference on Intelligent Systems and Computer Vision (ISCV) (pp. 1-5). IEEE.
23. Huang, X., Acero, A., Hon, H. W., & Foreword By-Reddy, R. (2001). Spoken language processing: A guide to theory, algorithm, and system development. Prentice hall PTR.
24. Hamidi, M., Satori, H., Zealouk, O., & Satori, K. (2019). Speech Coding Effect on Amazigh Alphabet Speech Recognition Performance. Journal of Advanced Research in Dynamical and Control Systems, vol. 11, no 2, p. 1392-1400.

25. Zealouk, O., Satori, H., Hamidi, M., Laaidi, N., & Satori, K. (2018). Vocal parameters analysis of smoker using Amazigh language. International Journal of Speech Technology, vol. 21, no 1, p. 85-91.
26. Hamidi, M., Satori, H., Zealouk, O., Satori, K., & Laaidi, N. (2018, October). Interactive Voice Response Server Voice Network Administration Using Hidden Markov Model Speech Recognition System. In 2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4) (pp. 16-21). IEEE.
27. Zealouk, O., Hamidi, M., Satori, H., & Satori, K. (2020). Amazigh Digits Speech Recognition System Under Noise Car Environment. In Embedded Systems and Artificial Intelligence (pp. 421-428). Springer, Singapore.
28. Ilham, A., Hassan, S., & Khalid, S. (2018, October). Building a first Amazigh database for automatic audiovisual speech recognition system. In Proceedings of the 2nd International Conference on Smart Digital Environment (pp. 94-99).
29. Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77(no. 2), 257–286.
30. Shariah, M. A. A., Ainon, R. N., Zainuddin, R., & Khalifa, O. O. (2007, November). Human computer interaction using isolated-words speech recognition technology. In 2007 International Conference on Intelligent and Advanced Systems (pp. 1173-1178). IEEE.