# Investigation Arabic Speech Recognition Using CMU Sphinx System

Hassan Satori[1, 2], Hussein Hiyassat[3], Mostafa Harti[1, 2], and Noureddine Chenfour[1, 2]
[1]UFR Informatique et Nouvelles Technologies d'Information, Dhar Mehraz Fès, Morocco
[2]Département de Mathématiques et Informatique, Faculté des Sciences, Morocco
[3]Arab Academy for Banking and Financial Sciences, Amman, Jordan

**Abstract:** *In this paper, Arabic was investigated from the speech recognition problem point of view. We propose a novel approach to build an Arabic automated speech recognition system using Arabic environment. The system, based on the open source CMU Sphinx-4, was trained using Arabic characters.*

**Keywords:** *Speech recognition, Arabic language, HMMs, CMUSphinx-4, artificial intelligence.*

*Received September 20, 2007; accepted December 13, 2007*

## 1. Introduction

Automatic Speech Recognition (ASR) is a technology that allows a computer to identify the words that a person speaks into a microphone or telephone. It has a wide area of applications: command recognition (voice user interface with the computer), dictation, interactive voice response, it can be used to learn a foreign language. ASR can help also, handicapped people to interact with society. It is a technology which makes life easier and very promising [8].

View the importance of ASR too many systems are developed, the most popular are: Dragon Naturally Speaking, IBM via voice, Microsoft SAPI. Open source speech recognition systems are available too, such as [24, 5, 17, 16, 11, 19]. We are interested in exploring this last, which is based on Hidden Markov Models (HMMs) [8]. A Hidden Markov Model (HMM) is a statistical model where the system being modeled is assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters, from the observable parameters, based on this assumption. The extracted model parameters can then be used to perform further analysis, for example for pattern recognition applications. Its extension into foreign languages (English is the standard) represent a real research challenge area.

Although Arabic is currently one of the most widely spoken language in the world, there has been relatively little speech recognition research on Arabic compared to the other languages [14, 23, 22]. The first works on Arabic ASR has concentrated on developing recognizers for Modern Standard Arabic (MSA). The most difficult problems in developing highly accurate ASRs for Arabic are the predominance of non diacritized text material, the enormous dialectal variety, and the morphological complexity.

Kirchhoff *et al.* [23] investigate the recognition of dialectal Arabic and study the discrepancies between dialectal and formal Arabic in the speech recognition point of view. Vergyri *et al.* [22] investigate the use of morphology-based language model at different stages in a speech recognition system for conversational Arabic; they studied also the automatic diacritizing Arabic text for use in acoustic model training for ASR. In their previous papers Satori *et al.* [21, 20], introduce an Arabic voice recognition system where both training and recognizing process use Romanized characters.

Most of previous works on Arabic ASR have been concentrated on developing recognizers using Romanized characters. In this work we investigate a system using entirely Arabic environment. We have generated a pronunciation dictionary and trained acoustic model with Arabic speech data. In the next section we present a brief description of the Arabic language. In section 3, we describe the Arabic speech recognition system and our investigations to adapt the system to Arabic language. In section 4, we present experimental results. Finally, in section 5, we provide our conclusions and future directions.

## 2. Arabic Language

Arabic is a Semitic language, and it is one of the oldest languages in the world. It is the 5th widely used language nowadays [2]. Standard Arabic has 35 basic the Pharyngealized (L) which is rarely used) and six are vowels, three long and three short [17].

Vowels can be arranged depending on the degree of constriction at the articulation point and on the tongue hump position as shown in Table 1 [4].

Table 1. Arabic vowels classification.

| **Tongue Hump Position**<br>**Degree Constriction** | **Front** | **Central** | **Back** |
|---|---|---|---|
| **High** | i كسرة<br>ii ياء ممدودة | | u ضمة<br>uu واو ممدودة |
| **Low** | | a فتحة<br>aa ألف ممدودة | |

Vowels are relatively easy to identify because of their high energy resulting in instance formant patterns. The vowels are voiced sounds.

Arabic phonemes contain two distinctive classes, which are named pharyngeal and emphatic phonemes. These two classes can be found only in Semitic languages like [17, 4, 7]. The allowed syllables in Arabic language are: CV, CVV, CVC, CVVC, CVCC and CVVCC where V indicates a (long or short) vowel while C indicates a consonant. Arabic utterances can only start with a consonant [17].

The structure of the syllable in Arabic is, of course, based on the phonemic of Arabic. The peak, or nucleus, is always the most prominent element of the Arabic syllable. It must be composed of a vowel; either long or short. Literal Arabic has six syllable patterns. Table 2 shows a classification of them. Arabic has one open short syllable, one-long syllable and four closed-long ones. The Arabic syllable system has the following features:

- The syllable must begin with a consonant followed by a vowel.
- The syllable never begins with two consonants.
- No one phoneme syllable exists in Arabic.

All Arabic syllables must contain at least one vowel. Also Arabic vowels cannot be initials and can occur either between two consonants or final in a word. Arabic syllables can be classified as short or long. The CV type is a short one while all others are long. Syllables can also be classified as open or closed. An open syllable ends with a vowel, while a closed syllable ends with a consonant. For Arabic, a vowel always forms a syllable nucleus, and there are as many syllables in a word as vowels in it [6, 18, 19].

Table 2. Arabic syllable patterns.

| | **Open** | **Closed** |
|---|---|---|
| **Short** | CV | |
| **Long** | CVV | CVC,<br>CVVC, CVCC,<br>CVVCC |

# 3. Arabic Speech Recognition System

This section describes our experience to create and develop an Arabic voice recognition system using entirely Arabic environment. Both training and recognizing process use Arabic characters.

## 3.1. System Overview

All of our experiments, both training and test were based on CMUSphinx4 system, which is HMM-based, speaker-independent, continuous recognition system capable of handling large vocabularies [10]. Our approach for modeling Arabic sounds in The CMU Sphinx system consisted of generated and trained acoustic and language models with Arabic speech data. The dictionary adopted in the experiments was made using Arabic characters, Figure 1 shows an excerpt from it. Also, new scripts were added to fine tuning CMUSphinx4 for Arabic. Different parameters in the system were adjusted, some are training parameters (number of state per HMM, number of Gaussians densities) and other are decoding parameters (silence insertion probability, word insertion probability, language weight) [3]. Thus, a 5-state left-to-right architecture is adopted to model each speech unit, and each state was modeled using a mixture of 16 Gaussians. Also, best results found at: filler insertion probability= 0.1, word insertion probability =0.123, language weight =6. Those parameters were determined through preliminary experiments in which we observed the performance of the system.

### 3.1.1. Corpus Preparation

An in house corpus was created from all 10 Arabic digits. A number of 60 Moroccan speakers (35 males and 25 females) were asked to utter all digits 5 times.

Table 3. Recording system parameters used for the corpus preparation.

| **Parameter** | **Value** |
|---|---|
| Sampling Rate | 16 kHz, 16 bits |
| Wave Format | Mono, Wav |
| Corpus | Adigits |
| Speakers | 60 (35 Males + 25 Females) |

Hence, the corpus consists of 5 repetitions of every digit produced by each speaker. Depending on this, the corpus consists of 3000 tokens. Table 3 shows parameters used for the corpus preparation. During the recording session, each utterance was played back to ensure that the entire digit was included in the recorded signal. All the 3000 (10 digits, 5 repetitions, 60 speakers) tokens were used for training phases.

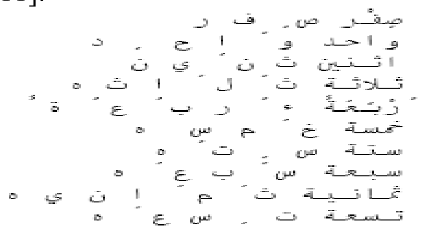
### 3.1.2. Training

For training acoustic models is necessary a set of feature files computed from the audio training data, one each for every recording in the training corpus. Each recording is transformed into a sequence of feature vectors consisting of the Mel-Frequency Cepstral Coefficients (MFCCs). The training was performed using 3000 utterances of speech data collected from Moroccan speakers.

Table 4. Phonemes symbols used in the training of HMMs.

| Transliteration | Alphabet |
|---|---|
| Alef | ء |
| Ba’ | ب |
| Ta’ | ت |
| Tha’ | ث |
| Ha’ | ح |
| Emphatic Kha’ | خ |
| Dal | د |
| Ra’ | ر |
| Ayn | ع |
| Sin | س |
| Emphatic Sad | ص |
| Lam | ل |
| Mim | م |
| Ha’ | ه |
| Waw | و |
| Ya’ | ي |
| Fatha | َ |
| Kasra | ِ |
| Between kasra and fatha | ِ |

The training process consists of: convert the audio data to a stream of feature vectors, convert the text into a sequence of linear triphone HMMs as shown in Table 4 using the pronunciation dictionary, and find the best state sequence or state alignment through the sentence HMM for the corresponding feature vector sequence. For each senone, gather all the frames in the training corpus that mapped to that senone in the above step and build a suitable statistical model for the corresponding collection of feature vectors. The circularity in this training process is resolved using the iterative Baum-Welch or forward-backward training algorithm [11].



صِفْر ص ِ ف ر
واحِد و َ ا ح ِ د
اثنين ث ِ ن َ ي ن
ثلاثة ث َ ل َ ا ث َ ه
أَرْبَعَة ء َ ر ب َ ع َ ة
خمسة خ َ م س َ ه
ستة س ِ ت َ ه
سبعة س َ ب ع َ ه
ثمانية ث َ م َ ا ن ِ ي َ ه
تسعة ت ِ س ع َ ه

Figure 1. Excerpt from pronunciation dictionary used in adigits application where each Arabic word is mapped onto its phoneme representation.

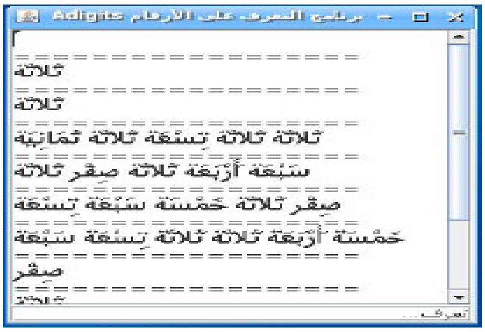
## 4. Experimental Results

In order to evaluate the performances of the application, we performed some experiments on different individuals each one of them was asked to utter 10 Arabic digits. We recorded the number of words that were correctly recognized, and then a mean recognition ratio for each tester was calculated as shown in Table 5.

Table 5. Results of adigit application test.

| | Test 1 | Test 2 | Test 3 | Mean Recognition Ratio |
|---|---|---|---|---|
| M1 | 10 | 9 | 10 | 96,67% |
| M2 | 8 | 10 | 10 | 93,33% |
| M3 | 10 | 9 | 9 | 93,33% |
| W1 | 9 | 8 | 9 | 86,66% |
| W2 | 8 | 8 | 9 | 83,33% |
| W3 | 9 | 8 | 10 | 90,00% |

Results are very satisfactory taken into account the small size of the corpus of training (personal corpus) which was used if compared with corpora used for English. We notice that, in order to reach good recognition performance it is recommended to train the system with large corpuses [11] (more than 500 different voices). We didn’t use large corpus; since our main goal is to demonstrate the possible adaptability of the system to Arabic environment as shown in Figure 2.



```
========================
ثلاثة
========================
ثلاثة
========================
ثلاثة ثلاثة تِسعَة ثلاثة ثمانِيَة
========================
سَبْعَة أَرْبَعَة ثلاثة صِفر ثلاثة
========================
صِفر ثلاثة خَمْسَة سَبْعَة تِسعَة
========================
خَمْسَة أَرْبَعَة ثلاثة ثلاثة تِسعَة سَبْعَة
========================
صِفر
========================
```

Figure 2. Automatic speech recognition system using Arabic environment.

## 5. Conclusion

To conclude, a spoken Arabic recognition system was designed to investigate the process of automatic speech recognition using Arabic environment. Both training and recognizing process use Arabic characters. Our experiment demonstrates the possible adaptability of the CMU Sphinx4 to Arabic language. We project to extend our application for wide Arabic language recognition, and for the Moroccan dialect language.

## Acknowledgements

We would like to thank people involved in the development on the CMU Sphinx 4 open source system [11].

## References

[1] Ajami Y., "Investigating Spoken Arabic Digits in Speech Recognition Setting," *in Proceedings of Information's and Computer Science*, UK, pp. 173-174, 2005.

[2] Al-Zabibi M., "An Acoustic Phonetic Approach in Automatic Arabic Speech Recognition," *Document with UMI,* the British Library, UK, 1990.

[3] CMU http: //cmusphinx .sourceforge.net./html/ cmusphinx .php, 2003.

[4] Deller J., Proakis J., and Hansen J., *Discrete Time Processing of Speech Signal*, Macmillan, NY, 1993.

[5] Deshmukh N., Ganapathiraju A., Hamaker J., Picone J., and Ordowski M., "A Public Domain Speech to Text System," *in Proceedings of 6th European Conferences on Speech Communication and Technology*, Hungary, pp. 2127-2130, 1999.

[6] El-Imam A., "An Unrestricted Vocabulary Arabic Speech Synthesis System", *Computer Journal of IEEE Transactions on Acoustic Speech and Signal Processing*, vol. 37, no. 12, pp. 1829-1845, 1989.

[7] Elshafei M., "Toward an Arabic Text to Speech System," *Computer Journal of the Arabian Science and Engineering*, vol. 4, no. 16, pp. 565-583, 1991.

[8] Haton M., Cerisara C., Fohr D., Laprie Y., and Smaili K., *Reconnaissance Automatique de la Parole du Signal a Son Interpretation*, Monographies and Books, Oxford, 2006.

[9] Hiyassat H., Nedhal Y., and Asem S., "Automatic Speech Recognition System Requirement Using Z Notation," *in Proceedings of of AMSE' 05*, France, pp. 514-523, 2005.

[10] Huang D., *Automatic Speech Recognition: The Development of the SPHINX System*, Kluwer Academic Publishers, 1989.

[11] Huang X., Acero A., and Hon H., *Spoken Language Processing: A Guide to Theory, Algorithm and System Design*, Prentice Hall, 2001.

[12] Huang X., Alleva F., Hon W., Hwang M., and Rosenfeld R., "The SPHINX-II Speech Recognition System: An Overview," *Computer Journal of Computer Speech and Language*, vol. 7, no. 2, pp. 137-148, 1993.

[13] Huang X., Ariki Y., and Jack M., "Hidden Markov Models for Speech Recognition," *Technical Report*, Edinburgh, UK, 1990.

[14] Kirchho K., Bilmes J., Henderson J., Schwartz R., Noamany M., Schone P., Ji G., Das S., Egan M., He F., Vergyri D., Liu D., and Duta N., "Novel Approaches to Arabic Speech Recognition," *Technical Report*, Ohns-Hopkins University, 2002.

[15] Lee K., Hon H., and Reddy R., "An Overview of the SPHINX Speech Recognition System," *Computer Journal of IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 38, no. 1, pp. 35-45, 1990.

[16] Li X., Zhao Y., Pi X., Liang H., and Nefian V., "Audio Visual Continuous Speech Recognition Using a Coupled Hidden Markov Model," *in Proceedings of 7th International Conferences on Spoken Language Processing,* Denver, pp. 213-216, 2002.

[17] Muhammad A., "Alaswaat Alaghawaiyah," *in Proceedings of International Conference on Signal Processing*, Jordan, pp. 646-651, 1990.

[18] Pullum G. and Ladusaw W., *Phonetic Symbol Guide*, Near New, USA, 1996.

[19] Ravishankar K., "Efficient Algorithms for Speech Recognition," *PhD Thesis*, 1996.

[20] Satori H. and Chenfour N., "Arabic Speech Recognition System based on CMUSphinx," *in Proceedings of International Symposium on Computational Intelligence*, Morocco, pp. 31-35, 2007.

[21] Satori H., Harti M., and Chenfour N., "Introduction to Arabic Speech Recognition Using Cmusphinx System," *in Proceedings of Information and Communication Technologies Interantinal Symposium (ICTIS'07)*, Morocco, pp. 139-115, 2007.

[22] Vergyri D. and Kirchhoff K., *Automatic Diacritization of Arabic for Acoustic Modelling in Speech Recognition*, Editors, Coling, Geneva, 2004.

[23] Vergyri D., Kirchhoff K., Duh K., and Stolcke A., "Morphology Based Language Modeling for Arabic Speech Recognition," *in Proceedings of Interspeech*, Germany, pp. 2245-2248, 2004.

[24] Young S., "The HTK Hidden Markov Model Toolkit: Design and Philosophy," *Technical Report TR 152,* 1994.

**Hassan Satori** is currently working in the Department of Computer Science and Mathematics, Faculty of Sciences, University Sidi Mohamed Ben Abbdallah Fez, Morocco.

**Hussein Hiyassat** is currently working in the Department of Computer Information System, Jordan, Arab Academy for Banking and Financial Sciences.

**Mostafa Harti** is currently working in the Department of Computer Science and Mathematics, Faculty of Sciences, University Sidi Mohamed Ben Abbdallah Fez, Morocco.

**Noureddine Chenfour** is currently working in the Department of Computer Science and Mathematics, Faculty of Sciences, University Sidi Mohamed Ben Abbdallah Fez, Morocco.