

Automatic Speech Recognition for Tamazight Enchained Digits

A. EL GHAZI^{1,*}, C. DAOUI¹, N. IDRISSE²

¹Laboratory of Modeling and Calculation, Faculty of Sciences and Technics, BeniMellal

²Team of Information Processing and Telecommunication, Faculty of Sciences and Technics, BeniMellal

*Corresponding author: hammadgm@yahoo.fr

Received June 30, 2013; Revised November 14, 2013; Accepted February 06, 2014

Abstract The evolution of humane-machines dialogue involved the apparition of a new security management technique. For this reason, there are a lot of systems that uses voice stamps and signal processing. In this work, we have treated a first stage of a security system that consists on password validation devoted to Tamazight dialect. In this context, an automatic speech recognition system for Tamazight enchained digits is established. We have based on construction rules of these digits to minimize a training database and to avoid the overlap between different numbers to increase a recognition rate.

Keywords: HMM (Hidden Markov Model), ASRS (Automatic Speech Recognition System), Tamazight, security systems

Cite This Article: A. EL GHAZI, C. DAOUI, and N. IDRISSE, "Automatic Speech Recognition for Tamazight Enchained Digits." *World Journal Control Science and Engineering*, vol. 2, no. 1 (2014): 1-5. doi: 10.12691/wjcse-2-1-1.

1. Introduction

Speech is the most natural means with which the persons communicate. The ASRS aims to transcribe a speech signal to a succession of words corresponding to a sentence pronounced by the speaker. This technology uses some methods of signal processing and artificial intelligence [3]. There are many applications which use speech recognition, we give as examples: robotic command, handicaps help, automatic answering machines, learning other languages and security systems. Thus, this paper focuses on Tamazight dialect and proposes an automatic speech recognition system for the Tamazight enchained digits.

For the importance of ASRS [21,22,23], many systems based on voice recognition has been developed, the more known are: the file research system in mobile phones, the oral commands dedicated to robot, automatic answering in central servers...etc. Following this flexible uses in different areas, the ASRS is a subject of several researches since the beginning of 1980's. Unfortunately, despite the incredible evolution of computers, the automatic speech recognition is not an active subject of research and the obtained results are very far from the ideal one would have expected twenty years ago.

Since the beginning of the twentieth century, there are several speech recognition systems concerning different languages, the best known: English speech recognition system [6-19] and French speech recognition system [8-21] that are integrated in mobile phone devices. On the other side, there are some speech recognition systems related to

the popular dialects such as Indian dialect and Algerian one [10].

Tamazight dialect [1] presents a popular language, more diffused in Moroccan mountainous regions, it constitutes a principal communication element for the majority of population of these regions. In this context, the realization of a speech recognition system relative to this dialect presents an advantage to expand the use of some systems like the automatic counters, mobiles phones, computers and TV emission's translation to facilitate their understanding. Therefore, in this work, we present an automatic speech recognition system for Tamazight digits from 1 to 199. This system is based on an optimal training corpus composed from the pronunciation of digits from 1 to 10 in addition to morpheme 'd' and syntagme 'id' which are the coordinators to construct the enchained digits. Taking into account of this construction rules, the Tamazight speech recognition can be expanded easily from the isolated number to the enchained digits. Our system can be exploited in dictation application and in security systems. In this sense, based on the realized system, we present a first part of a security system that is composed by the password verification and the speaker identification. The password verification system is devoted to Tamazight dialect and permits to compare the password pronounced by the user with the one already recorded in the user's database.

In this paper, each phoneme or syllable is modeled by a hidden Markov model [17], with three states. The choice of this model [15] returns to its robust modeling in automatic speech recognition.

In the remainder of this work, we will present a summarize of theoretical bases of automatic speech

recognition, then we will present the construction rules of the enchainned digits and security system’s architecture (section 3), after we will give the experimental results (section 4) and we will end with a conclusion.

2. Theoretical Bases of Automatic Speech Recognition

2.1. Signal Processing

The speech signal is subjected to a treatment by using a mathematics functions and transformations, to extract the main information that can be used in training and recognition steps. In this paper, we used the Mel Frequency Cepstral Coefficients MFCC [3-22], these parameters are the most popular coefficients used in speech treatment [24,25]. The steps of the extraction are shown in the following figure:

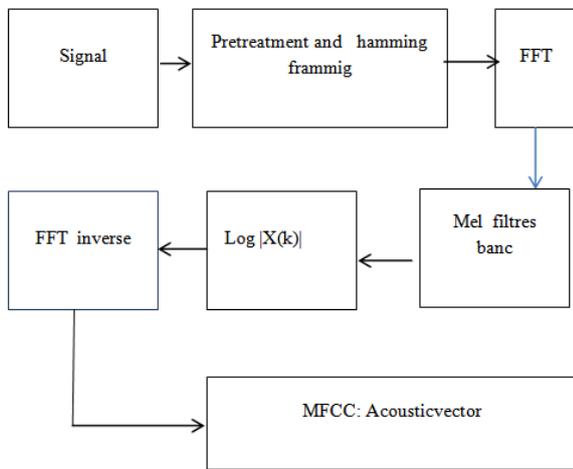


Figure 1. MFCC’s extraction steps

2.2. Hidden Markov Model

Since the beginning of 20th century, a lot of system shows the reliability of stochastic modeling in speech recognition area. The Hidden Markov Model [3] is the stochastic model more used in speech recognition system, it is based on the principal of states to model the speech unities named phonemes or syllabus. So, each state is characterized by the mean vector and covariance matrix, these parameters are adjusted in the training step by using the Baum-welch algorithm [15]. In the recognition step, we use Viterbi’s algorithm [15-22], the further permits to run through a graph of words or sentences and follow the elevated value of the observation probability.

3. The Proposed System

3.1. Transition from the Isolated Digits Recognition to a Continuous

Many speech recognition systems are concerned with isolated words, the development of systems that can recognize an enchainned word aims to cover the sentences and continuous signals. A speech recognition system for Tamazight enchainned digits based on their construction rules is formulated from the isolated numbers from 1 to 10, to which we have added the combinations. These coordinators signify the addition or multiplication of two numbers or more expressed respectively by the morphemes and syntagmes ‘d’ and ‘id’ to form the final digit. A constructions rules of Tamazight enchainned digits from the isolated numbers are detailed in next paragraphs.

3.1.1. Digits from 11 to 19

The morpheme ‘d’ is the tens addition element, for digits from 11 to 19, the rule is to commence by the units then the morpheme ‘d’ before the number ten. For example $\varsigma \circ \text{I} \ \Lambda \ \square \circ \text{U}$ (yan dmraw). The generalization of this rule can be presented by the following application:

$$f : [1,9] \rightarrow [11,19]$$

$$x \rightarrow x + 10$$

The following table presents some examples of enchainned digits:

Table 1. examples of enchainned digits

First number	morpheme	Second element	Resulted digits (Tifinagh)	Digits pronunciation
$\varsigma \circ \text{I}$	Λ	$\square \circ \text{U}$	$\varsigma \circ \text{I} \ \Lambda \ \square \circ \text{U}$	Yan d mraw (1+10) : $f(x=1)=11$
$\theta \leq \text{I}$			$\theta \leq \text{I} \ \Lambda \ \square \circ \text{U}$	Sin d mraw(2+10) : $f(x=2)=12$
$\kappa \text{Q} \circ \text{E}$			$\kappa \text{Q} \circ \text{E} \ \Lambda \ \square \circ \text{U}$	Krad d mraw(3+10) : $f(x=3)=13$
$\dagger \text{M} \circ \dagger$			$\dagger \text{M} \circ \dagger \ \Lambda \ \square \circ \text{U}$	Tza d mraw(9+10) : $f(x=9)=19$

3.1.2. Digits from 21 to 99.

For all digits from 21 to 99, the rule is to begin by the tens then the syntagme ‘id’ before the number 10, and later the number of units after the morpheme ‘d’. The following application presents the generalization of this rule:

$$f : [2,9]^* [1,9] \rightarrow [21,99]$$

$$(x, y) \rightarrow x * 10 + y$$

The following table presents some examples of this rule:

Table 2. Example of enchainned digits from 21 to 99

First number	Syntagme	Second number	Morpheme	Third number	Resulted digit (Tifinagh)	Resulted digit
$\theta \leq \text{I}$	$\leq \Lambda$	$\square \circ \text{U}$	Λ	$\varsigma \circ \text{I}$	$\theta \leq \text{I} \ \leq \Lambda \ \square \circ \text{U} \ \Lambda \ \varsigma \circ \text{I}$	Sin id mraw d yan(21)($f(2,1)=2*10+1=21$)
$\kappa \text{Q} \circ \text{E}$					$\kappa \text{Q} \circ \text{E} \ \leq \Lambda \ \square \circ \text{U} \ \Lambda \ \varsigma \circ \text{I}$	Krad id mraw d yan (31)($f(3,1)=3*10+1=31$)
$\dagger \text{M} \circ \dagger$					$\dagger \text{M} \circ \dagger \ \leq \Lambda \ \square \circ \text{U} \ \Lambda \ \varsigma \circ \text{I}$	Tza id mraw d yan (91)($f(9,1)=9*10+1=91$)

3.1.3. Digits above 100

Above the number 100, the rule consists to commence by 100 then the syntgme 'd', then we follow the rules described above. In this interval (>100) we can sum up the rule by the following application:

$$f : [0,9]^* [0,9]^* [0,9] \rightarrow [100,199]$$

$$(x, y, z) \rightarrow x * 10 + y + z + 100$$

In Table 3, we present some examples.

Table 3. examples for enchainned digits above 100

First number	Syntagme	Second number	Morpheme	Third number	Resulted digits (Tifinagh)	Resulted digits
t < C < E < E <	A	0 < 1	ε A	COU	t < C < E < E < A 0 < 1 ε A COU	Timidi d sin id mraw (120) F(2,0,0)=100+2*20+0+0=120
		5 < 1	A	COU	t < C < E < E < A R Q < E A COU	Timidi d yan d mraw(111) F(1,1,0)=100+1*10+1=111
		t < 9 < t	A	COU	t < C < E < E < A t < 9 < t A COU	Timidi d tza d mraw (119) F(1,9,0)=100+1*10+1+9+0=119

3.2. Security System’s Architecture

Currently, the automation of security systems and transaction, especially in money transfer field and modification of sensitive databases, constitute a challenge. In this paper, we exploit the automatic speech recognition to build a security system that can be integrated in automatic wicket, mobile phones and in personnel computers. Generally, the security system is composed of two parts: password validation and speaker verification.

3.2.1. Password Validation

In this step, an automatic speech recognition system is established. This system is used in the password composition. In the existing systems the digits constitute the composition corpus of the passwords namely the automatic wickets and mobile phones. In this paper, we have chosen to work with the Tamazight enchainned digits

in password composition. We took our recognition interval from 1 to 199. The users are invited to pronounce their passwords, the verification system and comparison allows to browse a database (database of users and their passwords) and compare the code with the existing one. In fact, if the code is correct, the second step will be the speaker verification.

2.3.2. Speaker Verification

The second step is the speaker verification, it permits to check his identity based on small corpus already registered. Creating voice tags for each speaker is the personal identification means. In this work, we have treated the first case of authentication, it consists on the realization of password recognition system of Tamazight dialect.

The following organization chart presents the summary of the proposed system:

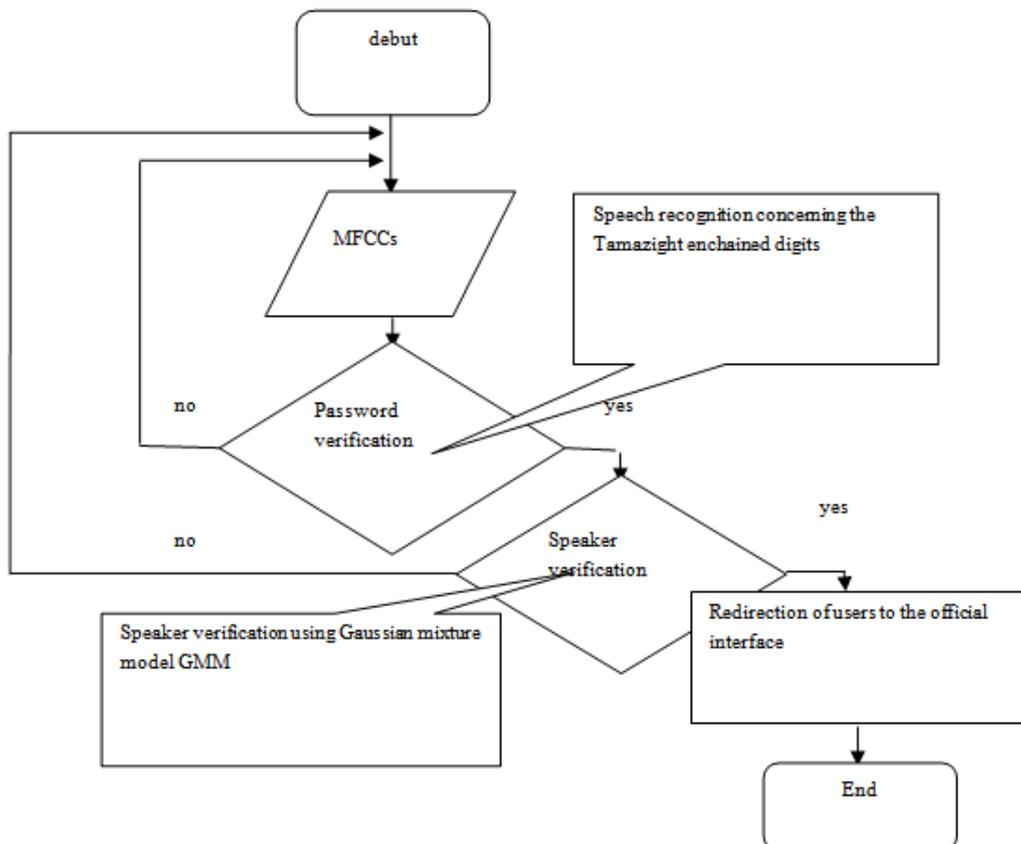


Figure 2. organigram of a security system

4. Experimental Results

4.1. Training Databases

The training database is composed by the pronunciation of Tamazight numbers from 1 to 10, the number 100 and two coordination elements 'd' and 'id' as well. The following table presents the characteristics of the training database.

Table 4. training database characteristics

Pronunciation numbers	The whole period
More than 5192 pronunciations for each digits	6,28h

4.2. Results

The training system is tested with a test database that contains 100 pronunciations, the system is evaluated by calculating the recognition rate defined by:

$$T = \frac{\text{number of recognized word}}{\text{testdatabase's size}}$$

We can also calculate the following quantities:

- T_R = reject rate

$$T_R = \frac{\text{number of not recognized word}}{\text{test database's size}}$$

- T_E = error rate

$$T_E = \frac{\text{number of false recognized word}}{\text{testdatabase's size}}$$

The obtained results are presented in the Table 5 and in the graph of the Figure 3.

Table 5. Experimental results

Test database	Evaluation rates
200 pronunciations, the corpus is composed by double digit numbers (from 11 to 99) and tree digit numbers (100 à 199).	$T=94,32\%$ $T_R=2,68\%$ $T_E=3\%$

In the following figure, we present a variation of the recognition rate according the delay of a training database. In fact, the recognition rate is proportional to the size of training database. So, increase the quality and size of training data is necessary to improve the recognition level.

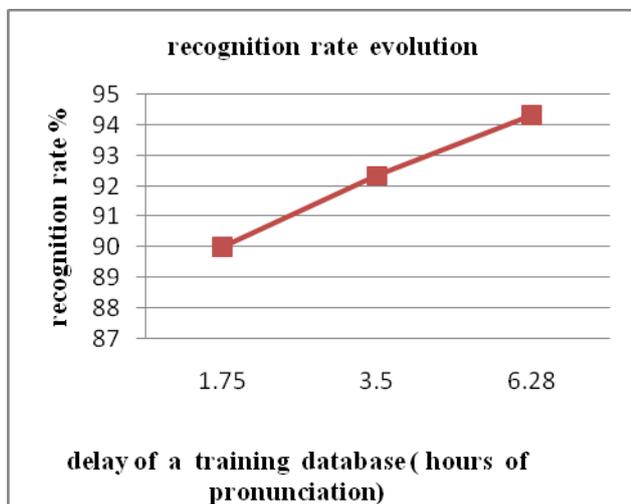


Figure 3. variation of recognition rates according the training database's delay

The Table 6 shows a comparison of the current system with a classical one. This latter, is trained with a database composed by the full pronunciation of numbers from 11 to 29. These numbers are pronounced by the same people as the current system. the comparison between these two systems is done on a test database composed by the numbers from 11 to 29.

Table 6. Comparison results

	Current system	Classical system
T	95%	86%
T_R	2%	3%
T_E	3%	11%

A speech recognition system based on the construction rules permits to minimize a training database's size by reformulating the enchainned digits from the isolated ones. While, the classical system is based on the full pronunciation of numbers from 11 to 29. In this context, we must integrate all pronunciations of numbers in this interval for training. Also, the comparison of T_E between the current system and the classical one permits to conclude that the first one reduces the overlaps between numbers.

In Figure 4, we present the graphical interface used for offline recognition. A recorded file can be recognized by using this interface, which is used to test the system using a test database.

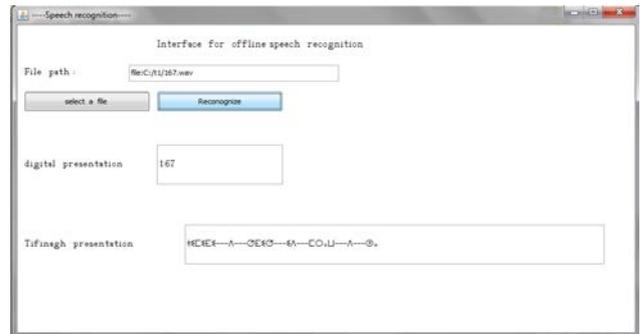


Figure 4. Java interface for offline recognition

In Figure 5, we present the online recognition, this interface is used for the live recognition, the further permits to recognize the user's password.

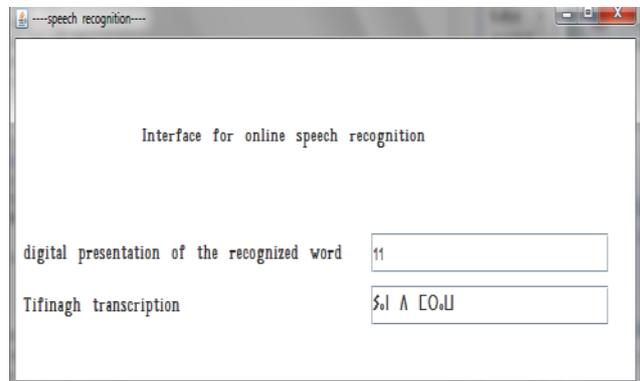


Figure 5. java interface for online recognition

The first part of the security system is presented in Figure 6, it concerns a recognition of passwords. A user pronounce his password, the further is compared with a user specific password already stored in the user's database.

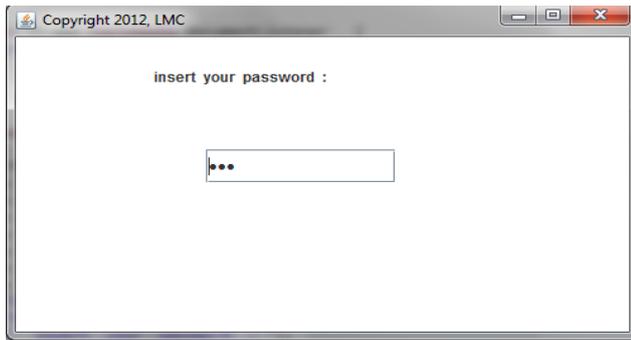


Figure 6. Java interface for passwords recognition

5. Conclusion

In this paper, we have exploited the construction rules of Tamazight digits to construct an automatic speech recognition system. This system is based on synthesis of chained digits from isolated number from 1 to 10, it appear as an important tool to minimize the training corpus, it avoid the overlap between different pronunciations. This system permits also, to expand the automatic recognition of Tamazight dialect. The obtained results are satisfactory in comparison with the training database's size. Therefore, our system constitutes a first part of a security system which we expect to improve the second part by realizing the speaker verification system.

References

- [1] AbedllahBoumalek, 'Variation syntaxique en Amazighe', publication IRCAM, 2008.
- [2] A. Chan, EvandroGouvêa& Rita Singh "Building Speech Applications Using Sphinx and Related Resources": <http://docpp.sourceforge.net>, August 2005.
- [3] A. Cornijeol and L. Miclet, "Apprentissage Artificielle-méthode et concept" editioneyrolles 1988.
- [4] S. DERRODE "Introduction au modele Markovienne pour signal et image", Institut Fresnel - UMR, jaune 2012.
- [5] Ali sadiqui& Noureddine chenfour "Reconnaissance de la parole arabe basé sur CMU Sphinx", Séria Informatica. Vol VIII fasc. 1 2010.
- [6] B. Resch "Automatic Speech Recognition with HTK" 2003.
- [7] Divejver and J. Killer, "Pattern recognition" in Pattern Recognition: a statistical approach"; Edition: Prentice Hall 1982.
- [8] G. SEMET & G. TREFFOT "La reconnaissance de la parole avec les MFCC" TIPE juin 2002.
- [9] H. Satori & M. Harti. "Système de la reconnaissance de la reconnaissance automatique de la parole", Faculté des Sciences, DharMehraz Fès, Maroc.
- [10] B. TOUNSI, ' Inférence d'identité dans le domaine forensique en utilisant un système de reconnaissance automatique du locuteur adapté au dialecte Algérien ' 2008
- [11] M. Amour, A. Bouhjar& F. Boukhris IRCAM: publication: "initiation à la langue Amazigh" 2004.
- [12] R. Gonzales and M. Thomson, "Syntactic pattern recognition" 1986.
- [13] Reweis, "Hidden Markov-Modele-Sam" 1980.
- [14] Robinerand Juang. "Fundamentales of speech recognition" 1993.
- [15] Benjamin LECOUTEUX 'Reconnaissance automatique de la parole guidée par des transcriptions a priori' Thesis in Avignon university of vaucluse countries 2008.
- [16] ChunshengFang "From Dynamic Time Warping (DTW) to Hidden Markov Model" (HMM) University of Cincinnati article 2009.
- [17] P. Galley, B. Grand & S. Rossier, "reconnaissance vocale Sphinx-4" EIA de Fribourg mai 2006.
- [18] S. Sigurdsson, Kaare Brandt Petersen and Tue Lehn-Schiøler "Mel Frequency Cepstral Coefficients: An Evaluation of Robustness of MP3Encoded Music", Informatics and Mathematical Modelling Technical University of Denmark article 2006.
- [19] S. Jamoussi, "Méthodes statistiques pour la compréhension automatique de la parole", Ecole doctorale IAEM Lorraine, 2004.
- [20] SEMET Gaetan& TREFFO, 'Grégory,'Reconnaissance de la parole avec les coefficients MFCC' TIPE june 2002.
- [21] T. AL ANI "Modèles de Markov Cachés (Hidden Markov Models (HMMs))", Laboratoire A2SI-ESIEE-Paris / LIRIS 2006.
- [22] T. Pellegrini et R. Duée "Suivi de la voix parlée grâce aux modèles de Markov Caché", lieu: IRCAM 1 place Igor Stravinsky 75004 PARIS june 2003.
- [23] RICHARD DUFOUR, 't ranscription automatique de laparole spontanée', thesis in Maine University, 2010.
- [24] A. EL GHAZI, C. DAOUI, N. IDRISSE, 'Speech recognition sytem concerning the morrocan dialect', IJEST Vol. 4 No.03 March 2012.
- [25] A. El Ghazi, C. Daoui, N. Idrissi, M. Fakir, B. Bouikhalene, 'Speech Recognition System Based On Hidden Markov Model Concerning the Moroccan Dialect DARIIJA', Global Journal of Computer Science and Technology Volume 11 Issue 15 Version 1.0 September 2011.