Investigation Amazigh speech recognition using CMU tools

Hassan Satori · Fatima ElHaoussi

Abstract   The aim of this paper is to describe the development of a speaker-independent continuous automatic Amazigh speech recognition system. The designed system is based on the Carnegie Mellon University Sphinx tools. In the training and testing phase an in house Amazigh_Alphadigits corpus was used. This corpus was collected in the framework of this work and consists of speech and their transcription of 60 Berber Moroccan speakers (30 males and 30 females) native of Tarifit Berber. The system obtained best performance of 92.89 % when trained using 16 Gaussian Mixture models.

1 Introduction

Automatic Speech Recognition (ASR) means a process that inputs human speech and tries to convert to its corresponding set of words based on a specific algorithm (Al-Zabibi 1990). ASR has attracted a huge amount of interest in the last decades, mainly due to a wide area of applications involving such technology. ASR can be used in industrial and civil areas such as; hands free operations, mobile voice application, human-computer interaction, automatic translation, automated telephone services and can help handicapped people to control equipments or navigate on internet. It is a technology which makes life easier and very promising (Haton et al. 2006). Recently, ASR systems have played important roles in several areas and different ASR systems are found in literature. ASR systems are highly dependent on: the language spoken (English, French), the way to recognize speech (isolated words, continuous speech) and the speaker characteristics (speaker dependent, speaker independent). These categories of systems are used depending on the type of users' application (Huang et al. 2001; Alotaibi and Shahshavari 1998; Satori et al. 2007, 2009; Abushariah et al. 2012).

In general, spoken alphabets and digits for different languages were targeted by ASR researchers. A speaker-independent spoken English alphabet recognition system was developed by Cole et al. (1990). That system was trained on one token of each letter from 120 speakers. Their performance was 95 % when tested on a new set of 30 speakers, but it was increased to 96 % when tested on a second token of each letter from the original 120 speakers. An artificial neural networks based speech recognition system was designed and tested with automatic Arabic digits recognition by Ajami Alotaibi (2005). The system was an isolated word speech recognizer and it was implemented both as a multi-speaker (i.e., the same set of speakers were used in both the training and testing phases) mode and speaker-independent (i.e., speakers used for training are different from those used for testing) mode. This recognition system achieved 99.5 % correct digit recognition in the multi-speaker mode, and 94.5 % in speaker-independent mode for clean speech. The Arabic speech recognition system developed by Hyassat and Zitar (2006) used CMU Sphinx4 engine based on

H. Satori (✉) · F. ElHaoussi
Department of Computer Science and Mathematics,
Faculty Polydisciplinary of Nador, Mohammed Premier
University, 300, Selouane 62700, Nador, Morocco
e-mail: hsatori@yahoo.com

H. Satori
Department of Computer Science and Mathematics,
Faculty of Sciences Dhar Mahraz,
Sidi Mohamed Ben Abbdallah University, B.P. 1796,
Fez, Morocco

HMM, which obtained a word recognition rate of 92.21 % for about 35 min of training and 7 min of testing speech data. Their system was trained using different Gaussian mixture models and they obtained best performance with eight Gaussians. Recently, Silva et al. (2012) investigated the speech recognition system digit in Portuguese language using Line Spectral Frequencies (LSF). They demonstrate that LFS provides best results in compared to those obtained by using Mel-Frequency Cepstrum Coefficients (MFCC).

The domain of ASR technologies is an expensive process and requires a considerable amount of resources. That is why only a small part of the world's languages can benefit of this kind of technologies and related tools (Le and Besacier 2009). The Amazigh language is considered as an African resource poor or less-resourced language (Boukous 1995; Greenberg 1966). To the best of our knowledge, there is a few speech recognition research works on less-resourced language as Amazigh. In this paper we describe our experience to design Amazigh speech recognition system based on HMM.

The paper is organized as follows: Sect. 2 presents a brief description of the Amazigh language. In Sect. 3, we describe the Amazigh speech recognition system and our investigations to adapt the system to Amazigh language. Section 4 investigates the experimental results. Finally, in Sect. 5, we provide our conclusions and future directions.

2 Amazigh language

The Amazigh language, known as Berber or Tamazight, is a branch of Hamito-Semitic (Afro-Asiatic) languages. It is, spoken in a vast geographical area of North Africa. Amazigh covers the northern part of Africa which extends from the Red Sea to the Canary Isles and from the Niger and Mali (Tuareg) in the Sahara to the Mediterranean Sea (Boukous 1995; Greenberg 1966); Amazigh Languages 2013; Galand 1988).

In Morocco, the Amazigh language is spoken by some 28 % of the population, grouped in three main regional varieties, depending on the area and the communities: *Tarifit* spoken in northern Morocco, *Tamazight* in Central Morocco and South-East, and *Tachelhit* spoken in southern Morocco (Ouakrim 1995; Chaker 1984).

Since 2003, Tifinaghe-IRCAM has become the official graphic system for writing Amazigh in Morocco. This system contains: (Outahajala and Zenkouar 2011; Boukous 2009; Fadoua and Siham 2012).

– 27 consonants including: the labials (ⵎ, ⵀ, ⵍ), the dentals (ⵜ, ⵏ, ⴹ, ⴻ, ⵉ, ⵔ, ⵇ, ⵞ), the alveolars (ⵙ, ⵥ, ⵣ, ⵤ), the palatals (ⵛ, ⵌ), the velar (ⴽ, ⵅ), the labiovelars (ⴽʺ, ⵅʺ),

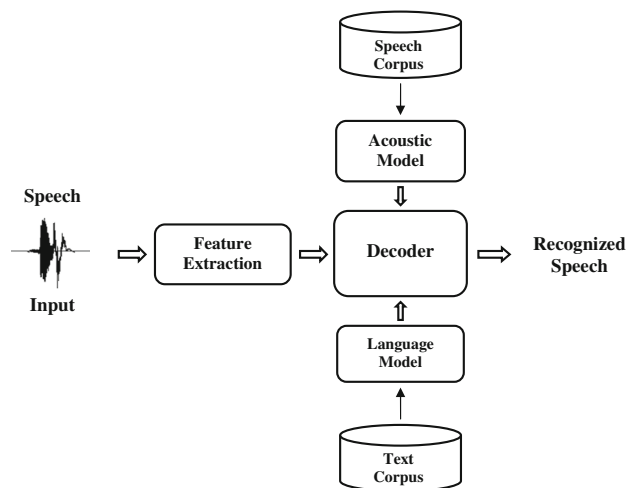**Fig. 1** Block diagram of ASR system

the uvulars (ⵇ, ⵅ, ⵖ), the pharyngeals (ⵂ, ⵃ) and the laryngeal (ⵁ, ⵀ);

– 2 semi-consonants : ⵢ and ⵡ .;

– 4 vowels: three full vowels: ⴰ, ⵉ and ⵓ and neutral ⴻ vowel (or schwa) which has a rather special status in Amazigh phonology.

3 Amazigh speech recognition system

This section describes our experience to create and develop an Amazigh voice recognition system using CMU Tools [cmu sphix web]. Figure 1 illustrates the main components that are usually found in a typical ASR system.

3.1 System overview

All of our experiments, both training and recognizing were based on CMU Sphinx system, which is HMM-based, speaker-independent, continuous recognition system capable of handling large vocabularies (CMU Sphinx Open Source Speech Recognition Engines 2013; Huang 1989; Lee 1989). Our approach for modeling Amazigh sounds in The CMU Sphinx system consisted of generated and trained acoustic and language models with Amazigh speech data. The dictionary adopted in the experiments was made using 43 Amazigh words and their transcriptions. The allowed syllables in Amazigh language are: V, CV, VC, CVC, C, CC and CCC where V indicates a vowel while C indicates a consonant (Ridouane 2003). Table 1 represents the Amazigh 10 first digits and 34 alphabets along with the way of how to pronounce them, type of syllable, their transcription using English Arabic and Amazigh scripts, and number of syllables in every spoken word.

Table 1 Ten first digits and 33 Amazigh alphabets with their syllables and their transcription in English, Arabic and Amazigh letters

Amazigh Alphadigits English Transcription	Tifinaghe Transcription [IRCAM]	Arabic Transcription	Syllabes	No. Of Syllabe	Arabic correspondence	Database Amazigh_Alphadigits Word code
AMYA	ⵓⵛⵉⵓ	أميا	VC-CV	2	صِفْر	AM00
YEN	ⵉⵏⵉ	يان	CVC	1	واحد	AM 01
SIN	ⵙⵉⵏ	سين	CVC	1	اثنان	AM02
KRAD	ⴽⵇⵓⴻ	كراض	VC-CVC	2	ثلاثة	AM03
KOZ	ⴽⴽⵅⵣ	كوز	CVC	1	أَرْبَعَةً	AM04
SMMUS	ⵙⵛⵛⵄⵙ	سموس	CC-CVC	2	خمسة	AM05
SDES	ⵚⴻⵉⵚ	سضيس	CCVC	1	ستة	AM06
SA	ⵙⵓ	سا	CV	1	سبعة	AM07
TAM	ⵜⵓⵛ	تام	CVC	1	ثمانية	AM08
TZA	ⵜⵣⵓ	تزا	CC-CV	2	تسعة	AM09
YA	ⵓ	يا	CV	1	ا	AM10
YAB	ⴱ	ياب	CVC	1	ب	AM11
YAG	ⵅ	ياݣ	CVC	1	-	AM12
YAGG	ⵆ	ياݣ	CVC	1	-	AM13
YAD	ⵏ	ياد	CVC	1	د	AM14
YADD	ⴹ	ياض	CVCC	1	ض	AM15
YEY	ⵢ	ياي	CVC	1	ي	AM16
YAF	ⴼ	ياف	CVC	1	ف	AM17
YAK	ⴽ	ياك	CVC	1	ك	AM18
YAKK	ⴿ	ياكك	CVCC	1	-	AM19
YAH	ⵀ	ياه	CVC	1	ه	AM20
YAHH	ⵃ	ياح	CVC	1	ح	AM21
YAAA	ⵄ	ياع	CVC	1	ع	AM22
YAKH	ⵅ	ياخ	CVC	1	خ	AM23
YAQQ	ⵇ	ياق	CVC	1	ق	AM24
YI	ⵉ	يي	CV	1	ي	AM25
YAJ	ⵊ	ياج	CVC	1	ج	AM26
YAL	ⵍ	يال	CVC	1	ل	AM27
YAM	ⵎ	يام	CVC	1	م	AM28
YAN	ⵏ	يان	CVC	1	ن	AM29
YO	ⵓ	يو	CV	1	ُ	AM30
YAR	ⵔ	يار	CVC	1	ر	AM31
YARR	ⵕ	يارً	CVC	1	-	AM32
YAGH	ⵖ	ياغ	CVC	1	غ	AM33
YAS	ⵙ	ياس	CVC	1	س	AM34
YASS	ⵚ	ياص	CVC	1	ص	AM35
YASH	ⵛ	ياش	CVC	1	ش	AM36
YAT	ⵜ	يات	CVC	1	ت	AM37
YATT	ⵟ	ياط	CVC	1	ط	AM38
YAW	ⵡ	باو	CVC	1	و	AM39
YAY	ⵢ	باي	CVC	1	ى	AM40
YAZ	ⵣ	ياز	CVC	1	ز	AM41
YAZZ	ⵥ	ياز	CVCC	1	-	AM42

Table 2 System parameters

Parameter	Value
Sampling rate	16 kHz
Number of bits	16 bits
Number of Channels	1, Mono
Audio data file format	.wav
Corpus	Amazigh_Alphadigits 10 digits and 33 amazigh alphabets
Accent	Moroccan Tarifit Berber
Number of speakers	60
Number of token per speaker	430
Speakers' gender	30 males and 30 females
Speakers' age	12–45 years-old
Recording time	3 h 20 min
Token number	25,800
Size of raw speech	~1 Gb
Number of words	43 (33 alphabets + 10 digits)
Number of repetitions per word	10
Condition of noise	Normal life
Preemphased	$1-0.97z^{-1}$
Window type	Hamming, 25.6 ms
Frames overlap	10 ms

3.2 Speech database preparation

The database *Amazigh_Alphadigits* was created in the framework of this work and it contains a corpus of speech and their transcription of 60 Berber Moroccan speakers.[1] The corpus consists of spoken 10 Amazigh firsts' digits (0–9) and 33 Amazigh alphabets collected from 60 Moroccan speakers native of Tarifit Berber (30 males and 30 females) aged between 12 and 45 years-old. The audio files were generated by speakers pronouncing the digits in numerical order and alphabets in alphabetical order. Thus, the task of labeling speech signals after segmentation is easy. The sampling rate of the recording is 16 kHz, with 16 bits resolution. Table 2 shows more speech corpus technical details.

During the recording sessions, speakers were asked to utter the 10 digits and 33 alphabets sequentially starting with digits followed by alphabets. Audio recordings for a single speaker were saved into one ".wav" file and sometimes up to four ".wav" files depending on number of sessions the

speaker spent to finish recording. It is time consuming to save every single recording once uttered.

Hence, the corpus consists of 10 repetitions of every digit and alphabets produced by each speaker. Depending on this, the corpus consists of 25,800 tokens. During the recording session, the waveform for each utterance was visualized back to ensure that the entire word was included in the recorded signal see Fig. 2. Therefore, there was a need to segment manually these bigger ".wav" files into smaller ones each having a single recording of a single word and manual classification of those ".wav" files into the corresponding directories was done. Wrongly pronounced utterances were ignored and only correct utterances are kept in the database.

3.3 Training

Training is the process of learning the Acoustic Model and Language Model to construct the knowledge base used by the system. The knowledge base contains: Acoustic Model, Language Model and Pronunciation Dictionary.

3.3.1 Feature extraction

The purpose of this sub-system (see fig. 1) is to extract speech features which play a crucial role in speech recognition system performance.

As seen in Table 2, the parameters used in our system, were 16 KHz sampling rate with a 16 Kbit sample, 25.6 ms Hamming Window with consecutives frames overlap by 10 ms and Mel-Frequency Cepstral Coefficients (MFCC).

3.3.2 Acoustic model

The acoustic model provides a mapping between the observed features of basic speech units (phonemes) provided by the front-end of the system and the Hidden Markov Models (HMMs) (Huang et al. 1990). In the HMM based technique words in the target vocabulary are modeled as a sequence phonemes, while each phoneme is modeled as a sequence of HMM states. The basic HMM model used in this work is a 5-states HMMs architecture for each Amazigh phoneme, three emitting sates and two non emitting ones as entry and exit which join models of HMM units together in the ASR engine, as shown in Fig. 3. Each emitting state consists of Gaussian mixtures trained on 13 dimensional Coefficients MFCC, their delta and delta delta vectors, which are extracted from the signal.

In this study, the acoustic model was performed using speech signal from the Amazigh Alphadigits training database. Table 4 shows a description database subsets used in the training. Every recording in the training corpus are transformed into a sequence of feature vectors. For each recording, a set of features files are computed using the front-end pro-

---

[1] The Amazigh Speech Corpus was collected by students during two periods of three month: (Mars to Mai 2011 and 2012), within the framework of the graduate programs of the faculty Polydisciplinary of Nador, Morocco.
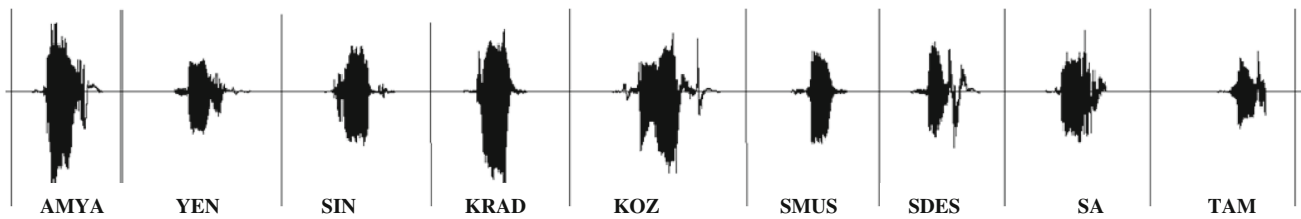
Fig. 2 Waveform for the ten firsts Amazigh digits recording session (speaker moel) before cutting it manually into separated single utterance.
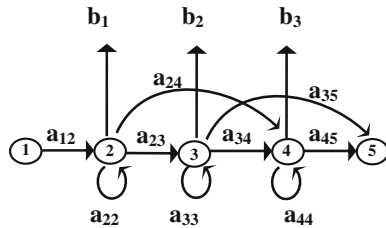

Fig. 3 The 5-states HMM model.

Table 3 The phonetic dictionary list used in the training

AMYA	a m y a	YAAA	y a aa
YEN	y e n	YAKH	y a kh
YEN(2)	y a n	YAQQ	y a qq
SIN	s i n	YI	y i
KRAD	k r a dd	YAJ	y a j
KOZ	k o z	YAL	y a l
SMUS	s m u s	YAM	y a m
SDES	s d e s	YAN	y a n
SDES(2)	s d e es	YO	y o
SA	s a	YAR	y a r
TAM	t a m	YARR	y a rr
TAM(2)	tt a m	YAGH	y a gh
		YAS	y a s
YA	y a	YASS	y a ss
YAB	y a b	YASH	y a sh
YAG	y a g	YAT	y a t
YAGG	y a gg	YATT	y a tt
YAD	y a d	YAW	y a w
YADD	y a dd	YAY	y a y
YEY	y e y	YAZ	y a z
YAF	y a f	YAZZ	y a zz
YAK	y a k		
YAK	y a kk		
YAH	y a h		
YAHH	y a hh		

vided by Sphinxtrain (CMU lmtool 2013). In this stage, the engine looks into the phonetic dictionary (see Table 3) which maps every used Amazigh word to a sequence of phonemes. During the training, all 44 Amazigh phonemes are used by

means of a phone list (Satori et al. 2009). The Amazigh phonemes are further refined into Context-Dependent (CD) tri-phones and added to the HMM set.

3.3.3 Language model

The n-gram language model used by the ASR system guide the search for correct word sequence by predicating the likelihood of the nth word, using the $n-1$ preceding words. The common feasible n-gram models are uni-gram, bi-gram and tri-gram. Creating of a language model consists of computing the word uni-gram counts, which are then converted into a task vocabulary, with word frequencies, generating the b-grams and tri-grams from the training text based on this vocabulary. In this work, The CMU-Cambridge statistical language modeling toolkit is used to generate Language model of our system (CMU lmtool 2013).

3.3.4 Pronunciation dictionary

The pronunciation dictionary called also lexicon it contains all Amazigh words we want to train followed by their pronunciation. Table 3, shows the phonetic dictionary list used in the training of our system. The alternate transcriptions marked with parenthesis like (2) stand for second pronunciation. The pronunciation dictionary serves as an intermediary between the Acoustic Model and Language Model.

4 Experimental results

In order to evaluate the performances of the system we had performed three experiments (Exp.1–Exp.3) and obtained results shown in the tables below. The experiments included the training and testing the system using different subsets of the Amazigh_*Alphadigits* corpus. In the first experiment the system was trained and tested by using only the Amazigh digits, second experiment working on the Amazigh alphabets, third the combination of the Amazigh digits and alphabets together. Table 4 shows the database subsets used in the three experiments their descriptions, Number of words, Number of speakers, Token number.

Table 4 Amazigh_Alphadigits database subsets description

Database subset name	Parameter	Value
Amazigh_alphadigits1	Number of words	43 (33 alphabets + 10 digits)
	Number of speakers	42
	Speakers' gender	21 males and 21 females
	Token number	18,060
	Recording time	2 h 20 min
	Used for	Training experience 3
Amazigh_alphadigits2	Number of words	43 (33 alphabets + 10 digits)
	Number of speakers	18
	Speakers' gender	9 males and 9 females
	Token number	7,740
	Recording time	1 h 0 min
	Used for	Testing experience 3
Am_digits1	Number of words	10 (10 digits)
	Number of speakers	42
	Speakers' gender	21 males and 21 females
	Token number	4,200
	Recording time	0 h 33 min
	Used for	Training experience 2
Am_digits2	Number of words	10 (10 digits)
	Number of speakers	18
	Speakers' gender	9 males and 9 females
	Token number	1,800
	Recording time	0 h 14 min
	Used for	Testing experience 2
Am_alphabets1	Number of words	33 (33 alphabets)
	Number of speakers	42
	Speakers' gender	21 males and 21 females
	Token number	13,860
	Recording time	1 h 48 min
	Used for	Training experience 1
Am_alphabets2	Number of words	33 (33 alphabets)
	Number of speakers	18
	Speakers' gender	9 males and 9 females
	Token number	5,940
	Recording time	0 h 46 min
	Used for	Testing experience 1

In all experiments corpus subsets were disjointed and partioned to training 70 % and testing 30 % in order to assure the speaker independent aspect. Also, the system was trained using different Gaussian mixture models. The numbers of

Table 5 Amazigh digits recognition rate for different GMMs when trained with Amazigh_Alphadigits1 and tested with Amazigh_Alphadigits2 (Experiment 1)

Amazigh digits	Recognition rate (%) for different Gaussian mixtures		
	4 GMM	8 GMM	16 GMM
AMYA	93.33	93.33	93.89
YEN	92.78	92.22	92.78
SIN	90.00	90.00	93.89
KRAD	94.44	94.44	95.56
KOZ	88.89	89.44	89.44
SMUS	91.11	92.22	92.78
SDES	90.56	93.33	93.33
SA	88.89	89.44	90.00
TAM	90.56	90.56	93.33
TZA	90.00	90.00	93.89
Total recognition rate	91.01 %	91.50 %	92.89 %

Gaussian mixtures per model were 4, 8, and 16. Each of 43 digits and alphabets were considered separately.

In the case of the first experiment, Table 5 shows the accuracy rate of the system in addition to the system total accuracy rate using all digits. Depending on the testing corpus subset, the system had to recognise 1,800 token for the all 10 digits. The system performances are 91.01, 91.50 and 92.89 % was found for using 4, 8 and 16 Gaussian mixture distributions, respectively. It is found that 16 GMMs obtained the best recognition rate of 92.89 %.

By considering the digits recognition analysis, the most frequently misrecognized Amazigh digits are SA and KOZ. Both of these words are monosyllabic and the modelling process of such one syllable words is more difficult than two or more syllable words.

In the second experiment, the systems try to recognize 5,940 samples of all 33 amazigh alphabets. Table 6 shows the accuracy rate of the system. The performances are 87.90, 88.50 and 89.28 % for using 4, 8 and 16 GMMs, respectively. Also, in the case of alphabets the best results was found with 16 GMMs. The most frequently misrecognized Amazigh alphabets are YA and YO.

In the last experiment, all alphabets and digits are combined in order to use a maximum available dataset. In testing, the system was programmed to test total of 7,740 alphabets and digits tokens. The system correct rates in this case were 88.07, 88.88 and 89.07 % for 4, 8 and 16 GMMs, respectively. This confirms our previous observation that 16 GMMs performs better as compared to 4 and 8 GMMs. Also, it is noted that he system performance was better for alphabets but lower for digits (Tables 7, 8).

Table 6 Amazigh alphabets recognition rate for different GMMs when trained with Amazigh_Alphadigits1 and tested with Amazigh_Alphadigits2 (Experiment 2)

Tifinaghe Alphabet	Recognition rate (%) for different Gaussian mixtures		
	4 GMM	**8 GMM**	**16 GMM**
ⴰ	79.44	79.44	79.44
ⴱ	83.33	83.33	83.89
ⵝ	89.44	89.44	90.00
ⵅ	85.56	85.56	86.11
ⴷ	86.67	87.22	87.22
ⴹ	87.22	87.22	87.22
ⴺ	86.67	88.33	88.00
ⵖ	90.00	90.56	91.11
ⴼ	88.33	88.33	90.00
ⴽ	86.67	87.78	90.00
ⴸ	90.56	90.56	91.11
ⵈ	85.56	85.56	87.22
ⵂ	87.22	87.22	88.33
ⵃ	86.11	86.11	86.11
ⵣ	92.78	90.56	91.11
ⵞ	86.67	88.33	88.89
ⵉ	90.00	92.22	93.33
ⵏ	86.67	86.67	88.33
ⵛ	90.56	90.56	93.33
ⵍ	90.00	90.00	91.11
ⵊ	78.89	78.89	79.44
ⵔ	91.11	91.11	91.11
ⵇ	90.56	90.56	91.11
ⵢ	90.56	93.33	91.11
ⵙ	91.11	91.11	91.11
ⵯ	87.22	87.22	88.33
ⴻ	91.67	93.33	93.33
ⵜ	88.33	88.33	88.33
ⴴ	86.11	87.78	90.00
ⵋ	87.22	88.33	88.33
ⵚ	86.11	88.33	90.00
ⵗ	91.67	91.67	92.22
ⵄ	90.56	90.56	90.56
Total Recognition rate	**87.90%**	**88.35%**	**89.00%**

Table 7 Amazigh alphadigits recognition rate for different GMMs when trained with Amazigh_Alphadigits1 and tested with Amazigh_Alphadigits2 (Experiment 3)

Amazigh Alphadigits	Recognition rate (%) for 16 GMM	Amazigh Alphadigits	Recognition rate (%) for 16 GMM
AMYA	92.78	ⵔ	87.22
YEN	92.22	ⵅ	86.11
SIN	93.33	ⵒ	90.56
KRAD	94.44	ⵞ	88.33
KOZ	89.44	ⵉ	92.22
SMUS	92.78	ⵎ	86.67
SDES	93.33	ⵛ	90.56
SA	83.33	ⵏ	90.00
TAM	92.22	ⵗ	78.89
TZA	90.56	ⵔ	91.11
ⵀ	79.44	ⵕ	90.56
ⴱ	83.33	ⵢ	93.33
ⵝ	89.44	ⵙ	91.11
ⵅ	85.56	ⵟ	87.22
ⵠ	87.22	ⵛ	93.33
ⴹ	87.22	ⵜ	88.33
ⵣ	88.33	ⴳ	87.78
ⵃ	90.56	ⵓ	88.33
ⵘ	88.33	ⵚ	88.33
ⵖ	87.78	ⵡ	91.67
ⵯ	90.56	ⵥ	90.56
ⵋ	85.56		
Total Recognition rate	**89.07%**		

Table 8 System overall recognition rate for experiments 1–3

	Overall system recognition rate for different Gaussian mixtures		
	4 GMM (%)	8 GMM (%)	16 GMM (%)
Exp1.	91.01	91.50	92.89
Exp2.	87.90	88.35	89.00
Exp3.	88.07	88.88	89.07

5 Conclusion

In this paper, we investigated the speaker independent alphadigits ASR system using a database of sounds corresponding to digits and alphabets spoken in Moroccan Amazigh language. This system was implemented by using Carnegie Mellon University Sphinx tools based on HMMs. This work includes creating the speech database amazigh alphadigits, which consist of many subsets covering all

digits and alphabets of Berber language used in the training and testing phase of the system. Recognition results show that our Amazigh ASR system is speaker independent and its performance is comparable to that reported (Hyassat and Zitar 2006) Arabic recognition results.

References

Abushariah, M. A. A. M., Ainon, R. N., Zainuddin, R., Elshafei, M., & Khalifa, O. O. (2012). Arabic speaker-independent continuous automatic speech recognition based on a phonetically rich and balanced speech corpus. *International Arab Journal of Information Technology*, *9*(1), 84–93.

Ajami Alotaibi, Y. (2005). Investigating spoken Arabic digits in speech recognition setting. *Information and Computer Science, 173*, 115–139.

Alotaibi, Y. A., & Shahshavari, M. M. (1998). *Speech recognition—What it takes for a computer to understand your commands*. IEEE Potentials.

Al-Zabibi, M. (1990) An acoustic-phonetic approach in automatic Arabic Speech Recognition. The British Library in Association with UMI.

Amazigh Languages. (2013). *Encyclopædia Britannica Online*. Retrieved 23 June, 2013, from http://www.britannica.com/EBchecked/topic/61496/Amazigh-languages.

Boukous, A. (1995). Société, langues et cultures au Maroc: Enjeux symboliques (No. 8). Faculté des Lettres et des Sciences Humaines.

Boukous, A. (2009). *Phonologie de l'amazighe*. Rabat: Institut royal de la culture amazighe.

Chaker, S. (1984). *Textes en linguistique berbère: introduction au domaine berbère*. Paris: Ed. du C.N.R.S.

CMU lmtool. (2013). Retrieved June 23, 2013, from http://www.speech.cs.cmu.edu/tools/lmtool-new.html.

CMU Sphinx Open Source Speech Recognition Engines. (2013). Retrieved February 10, 2013, from http://www.cmusphinx.sourceforge.net/html/cmusphinx.php.

Cole, R., Fanty, M., Muthusamy, Y., & Gopalakrishnan, M. (1990). Speaker-independent recognition of spoken English letters. In *International joint conference on neural networks (IJCNN)* (Vol. 2, pp. 45–51).

Fadoua, A. A., & Siham, B. (2012). Natural language processing for Amazigh language: Challenges and future directions. *Language Technology for Normalisation of Less-Resourced Languages, 19*.

Galand, L. (1988). Le berbère. In J. Perrot (Ed.), *Les langues dans le monde ancien et moderne. Part 3: Les langues chamito-sémitiques* (pp. 207–242). Paris: CNRS.

Greenberg, J. H. (1966). *The languages of Africa*. Mouton: The Hague.

Haton, M.-C., Cerisara, C., Fohr, D., & Laprie, Y., & Smaili, K. (2006). *Reconnaissance automatique de la parole du signal a son interpretation*. Paris: Universciens Dunod.

Huang, X., Acero, A., & Hon, H. (2001). *Spoken language processing a guide to theory, algorithm and system design*. Upper Saddle River: Prentice Hall.

Huang, X. D. (1989). The SPHINX-II Speech Recognition System: An overview. *Computer Speech and Language*, *7*(2), 137–148.

Huang, X. D., Ariki, Y., & Jack, M. A. (1990). *Hidden Markov models for speech recognition*. Edinburgh: Edinburgh University Press.

Hyassat, H., & Zitar, R. A. (2006). Arabic speech recognition using SPHINX engine. *International Journal of Speech Technology*, *9*(3–4), 133–150.

Le, V. B., & Besacier, L. (2009). Automatic speech recognition for under-resourced languages: Application to Vietnamese language. *IEEE Transactions on Audio, Speech, and Language Processing*, *17*(8), 1471–1482.

Lee, K. F. (1989). *Automatic Speech Recognition the development of the SPHINX system*. Boston: Kluwer.

Ouakrim, O. (1995). *Fonética y fonología del Bereber*. Survey: University of Autònoma de Barcelona.

Outahajala, M., Zenkouar, L., & Rosso, P. (2011). Building an annotated corpus for Amazighe. In *Proceedings of 4th international conference on Amazigh and ICT*, Rabat, Morocco.

Ridouane, R. (2003). Suites de consonnes en berbère: phonétique et phonologie. Doctoral Dissertation, Université de la Sorbonne nouvelle-Paris III.

Satori, H., Harti, M., & Chenfour, N. (2007). Arabic Speech Recognition system based on CMUSphinx. In *Proceedings of ISCIII2007, 3rd international symposium on computational intelligence and intelligent informatics*, Agadir, Morocco, pp. 31–35.

Satori, H., Hiyassat, H., Harti, M., & Chenfour, N. (2009). Investigation Arabic Speech Recognition using CMU Sphinx System. *The International Arab Journal of Information Technology, 6*(2), 186–190.

Silva, D. F., de Souza, V. M., Batista, G. E., & Giusti, R. (2012). Spoken digit recognition in Portuguese using line spectral frequencies. In *Advances in artificial intelligence—IBERAMIA 2012* (pp. 241–250). Berlin: Springer.