Conférence Internationale en Automatique & Traitement de Signal (ATS-2017)
Proceedings of Engineering and Technology – PET
Vol.23pp.28-32

Phonetics Arabic Database Automatically Segmented: PADAS

Mohamed Khalil Krichi [#1], Cherif Adnan *[2]

[#] *Electronics Applied Research Laboratory,*
Polytechnic Central Private School of Tunis 3 street Mohamed V, Elkram Tunis Tunisia
[1]krichi_moha@yahoo.fr
* *Faculty of Sciences of Tunis/ Laboratory of Signal Processing*
Physics Department University of Tunis-Manar TUNIS, 1060, TUNISIA
[2]*adnan2fr @yahoo.fr*

Abstract— This work describes a construction of PADAS "Phonetics Arabic Database Automatically segmented" based on a data-driven Markov process. The use of a segmentation database is necessary in speech synthesis and recognizing speech. Manual segmentation is accurate but inconsistent, since it is often produced by more than one label and require time and money. The MAUS segmentation and labeling exist for German speech and other languages but not in Arabic. It is necessary to modify MAUS for establish a segmental database for Arab. The speech corpus contains a total of 600 sentences recorded by 3 (1 female and 2 male) Arabic native speakers from Tunisia, 200 sentences for each.

Keywords— HTK; MAUS; Phonetic; Database; Automatic Segmentation.

I. INTRODUCTION

Many researches such as automatic speech recognition or speech synthesis are now based on database e.g. English [1, 2, 3 and 4]. For obtaining a good result, the database must be balanced, segmented and reduce the noise (noise in step of record). The target of this work is to product a robust speaker-independent continuous Arabic. These recordings contain all the phonemes of Arabic language. This database are rich characteristic and balanced.

This database of speech recordings must be based on a proper written set of sentences and phrases created by experts. Therefore, it is crucial to create a high quality written (text) set of the sentences and phrases before recording them. Any work based on the learning step requires a database to learn the system and then evaluate it. They are a several international databases in field of speech such as TIMIT which was developed by DARPA Committee for American English. And we also find other databases in different known languages, such as French and German, and unknown, as Vietnamese and Turkish.

For Arabic, we have not found a standard database, but we still found a few references. KACST [5] database developed by the Institute of King Abdul -Aziz in Saudi Arabia.

A. KACST

In 1997, KACST created a database for Arabic language sounds. In This database, there are 663 phonetically words. These words are phonetically rich and containing all Arabic phonemes.

In the domain of signal process, ASR and text-to-speech synthesis applications a data base is necessary.

In 2003, KACST produced a technical report of the project "Database of Arabic Sounds: Sentences". The sentences of Arabic Database have been written using the said 663 phonetically rich words. The database consists of 367 sentences; 2 to 9 words per sentence.

The purpose is to produce Arabic sentences and phrases that are balanced and phonetically rich based on the previously created list of 663 phonetically rich words [6].

B. ALGASD

ALGERIAN ARABIC SPEECH DATABASE (ALGASD) [7] created for the treatment of Algeria Arabic speech taking into account the different accents from different regions of the country. Unavailability and lack of resources for a database audio prompted us to build our own database to make the recognition of numbers and operations of a standard calculator in Arabic for a single user. We made 27 recordings of 28 vocabulary words.

Database is the most important tool for multiple domains as speech synthesis or speech recognition. to provide database a interesting and contains all the acoustic units must have all the possible linguistic combinations .The quality of the final result of the synthesis is directly dependent on the quality of recordings made during the development of the acoustic units therefore a filtering step dictionary is mandatory.

The implementation stages can be summarized as follows:

The choice of dictionary (set of sentences contains several examples of phonemes.).

ISSN 2356-5608

Conférence Internationale en Automatique & Traitement de Signal (ATS-2017)
Proceedings of Engineering and Technology – PET
Vol.23pp.28-32

- Sound recording expressions.
- Noise reduction.
- Segmentation

II. ARABIC LANGUAGE

Today, the first language in the world is Arabic language [8].The Arabic language is a derivational and inflectional language. The original Arabic is the language spoken by the Arabs. In addition, it is the sacred language of the Koran and Islam. Because the spread of Islam and the spread of the Qur'an, the language became a liturgical language. It is spoken in 22 countries, while the number of speakers is more than 280 million [9].

A. *Alphabet Consonants*

The Arabic alphabet consists of twenty eight consonants (see Table 1) basic, but there are authors who treat the letter (alif) as the twenty –ninth consonant. The (alif) behaves as a long vowel never found as consonant of the root.

Vowels play an important role in the Arabic words, not only because they remove the ambiguity, but also because they give the grammatical function of a word regardless of its position in the sentence. Indeed, vowels are not as consonants, they are rarely noted. They are written only to clarify ambiguities in the editions of the Koran or in the academic literature. In other words, vowels have a dual function: one morphological or semantic and the other are syntactic. Arabic has two sets of vowels, the short one and the other long.

B. *Short Vowels*

The short vowels (ُ ,ِ ,َ) are added below or above consonants. When the consonant has no vowel, it will mark an absence of vowel represented in Arabic by a silent vowel (ْ).

Long Vowels

Long vowels are long letters, they are formed by a brief vowels and one of the following letters (ي , و , ا)

The Diacritics

Short vowels are represented by symbols called diacritics (see Figure 5). Three in number, these symbols are transcribed as follows:

- The Fetha [a] is symbolized by a small line on the consonant (مَ / ma /)
- Damma the [u] is symbolized by a hook above the consonant (مُ/ mu /)
- The kasra [i] is symbolized by a small line below the consonant (مِ/ mi /)
- A small round o symbolizing Sukun is displayed on a consonant when it is not linked to any vowel.

C. *The Tanwin*

The sign of tanwin is added to the end of words undetermined. It is related to exclusion with Article determination placed at the beginning of a word. Symbols tanwin are three in number and are formed by splitting diacritics above, which results in the addition of the phoneme / n / phonetically:

[an]: (عً/ AIan /)

[un] : (عٌ / AIun /)

[in]: (عٍ/ AIin /)

D. *The Chadda*

The sign of the chadda can be placed over all the consonants non initial position. The consonant which is then analyzed receives a sequence of two consonants identical:

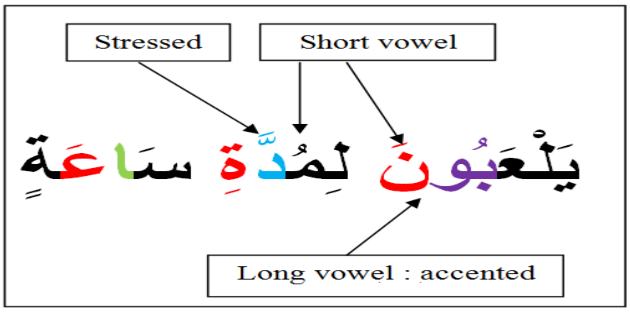

Fig 1 EXAMPLE OF A DATABASE USED / JALAIABUUNA LIMUDDATI SAAITIN / ("THEY PLAY FOR AN HOUR").

TABLE 1: ARABIC CONSONANT AND VOWELS AND THEIR SAMPA CODE.

Graphemes	symbol	Graphemes	symbol	Graphemes	symbol	Graphemes	symbol
ء	?	غ	g	ر	r	ي	j
ز	z	ف	f	ء	b	َ	a
س	s	ق	q	ب	t	ً	a:
ش	S	ك	k	ت	T	ِ	i
ص	s’	ل	l	ث	Z	ٍ	i:
ض	d’	م	m	ج	X	ْ	u
ط	t’	ن	n	ح	x	ُ	u:
ظ	D’	د	d	ه	h		
ع	?’	ذ	D	و	w		

III. BALANCED SELECTION OF ARABIC WORDS

In the Arabic language, the number of the syllables isn’t much and the structure syllabic is easily detectable.

In the beginning for each syllable there are a consonant followed by a vowel.

The Arabic words are composed at least by one syllable.

Arabic syllables can be classified either according to the length of the syllable or according to the end of the syllable. Short vowels are denoted by (V) and long vowels are denoted

ISSN 2356-5608

by (VV). Every vowel is placed in the second place of the syllable. These characteristic make the process of syllabification easier. There are five type of syllable. Short syllables occur only in CV form, because it is ending with a vowel so it is open. Medium syllable can be in the form of open CVV, or closed CVC. A long syllable has two closed forms CVVC, and CVCC. ; most contain two or more syllables. The longest word is combined of five syllables. Table II illustrates Arabic syllables. Some of the Arabic words are spelled together forming new long words with 6 syllables like (يَأْكُلُونَهُ), or 7 syllables like (يَسْتَقْبِلُونَهُ).

TABLE 2: THE DIFFERENT ARABIC TYPES SYLLABLES.

Syllable	Arabic Example		English meaning
CV	لِ	li	to
CVV	فِي	fii	in
CVC	قُلْ	qul	say
CVCC	بَحْرْ	bahr	sea
CVVC	مَالْ	maAl	money
CVVCC	زَارّ	zaArr	visit

A. *Corpus Description*

The sentences are checked and monitored for phonetic balanced distribution for a set of phonetically rich and well balanced words. Some of these sentences and phrases can be removed and / or replaced by others in order to achieve adequate phonetic distribution [10] .The corpus, we used to build our database is composed of 200 phrases, with an average of 5 words per sentence. These sentences contain 1000 words, 2600 syllables, 7445 phonemes including 2302 short vowels and long vowels. These sentences were read at an average speed (from 10 to 12 phonemes/second) by Tunisian speakers, two male and a female. They were sampled at 16 KHz with 16 bits per sample.

B. *Corpus Analysis*

We conducted a statistical analysis of our corpus. Table 3 shows the results of this study. We can note the following:

- The short vowel [a] and the long vowel [a:] appear with a frequency of 17%, followed by vowels [i] and [i:] with an occurrence frequency of 14.3%. The vowels [u] and [u:] represent 7%.
- The occurrence of the vowel (short and long) is about 37%.
- The most frequent Arabic consonants are: [?] (15%), [n] (6.66%), [l] (6.63%), [m] (5.59%), etc.

TABLE 3: OCCURRENCE FREQUENCY (%) OF ARABIC CONSONANTS AND VOWELS.

Consonant and vowels	Phoneme Repetitions	%
?	523	13,34%
b	102	2,60%
t	92	2,35%
T	70	1,79%
x	19	0,48%
/X	20	0,51%
G	35	0,89%
d	39	0,99%
D	40	1,02%
r	102	2,60%
z	35	0,89%
s	48	1,22%
S	73	1,86%
s’	18	0,46%
d’	24	0,61%
t’	19	0,48%
D’	24	0,61%
?’	61	1,56%
g	23	0,59%
f	61	1,56%
q	61	1,56%
k	80	2,04%
l	260	6,63%
m	219	5,59%
n	261	6,66%
h	123	3,14%
w	51	1,30%
j	80	2,04%
a	400	10,20%
a:	254	6,48%
i	400	10,20%
i:	28	0,71%
u	252	6,43%
u:	23	0,59%
total	3920	100%

C. *Noise Reduction*

Signal degradation by noise is a pervasive problem [11].

In the field of signal processing, noise suppression is a major problem. The wavelet transform is the main method used in the degradation signals. A whole series of different scaling functions and wavelet (or scaling and wavelet coefficients) offer many possible settings and regulatory variables.

ISSN 2356-5608

Conférence Internationale en Automatique & Traitement de Signal (ATS-2017)
Proceedings of Engineering and Technology – PET
Vol.23pp.28-32

[12]The audio recordings were noisy with a continuous background noise. Our goal is to reduce this undesirable component. Figures 2 and 3 shows the time signal before and after filtering for a particular audio file. We note in particular that the zone of silence highlighted is closer to zero in the filtered signal in the original signal release.

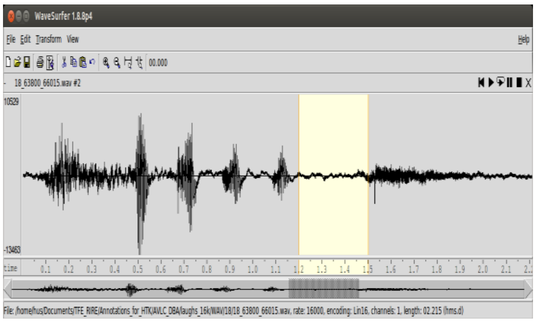

Fig 2 EXAMPLE FOR ORIGINAL SPEECH OF DATABASE FILE.

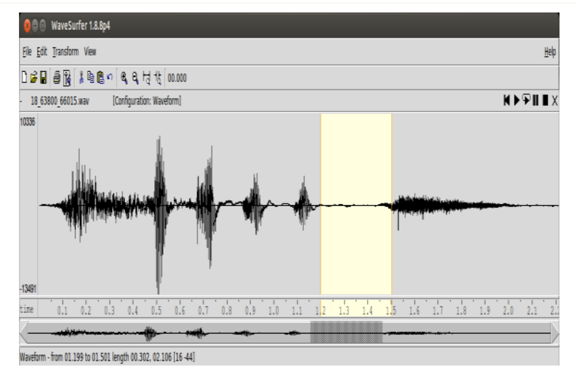

Fig 3 Example for the Same Speech Denoising.

D. *Automatic Segmentation*

Nowadays, the Practical applications of automatic S&L are implemented as a statistical search for a S&L $\hat{k}$ in a space Ψ of all possible S&Ls, which can be formulated as [13,14]:

$$\hat{k} = \arg\max_{k \in \psi} P(k \mid O) = \arg\max_{k \in \psi} \frac{P(k)p(O \mid k)}{P(O)} \qquad (1)$$

Where, O is the acoustic observation on the corresponding speech signal. The MAUS system models $P(k)$ for each recording O. Each path from the start node to the end node represents a possible $k \in \psi$ and accumulates to the probability $P(k)p(O \mid k)$ which is determined by HMMs for each phonemic segment and a simple Viterbi search through the graph yields the maximal $P(k)p(O \mid k)$.

The 'Munich Automatic Segmentation' (MAUS) system developed by Department of Phonetics, University of Munich, For more details about the MAUS method refer to [15], [16] and [17].

The purpose is analyzing a spoken utterance. Indeed, the MAUS system accepts in input a speech wave and another file with an orthographic transcription. In the text file, every sentence divided in single words. Thereafter a text-to-phoneme algorithm is used. This algorithm based on rule a combination of lexicon lookup.

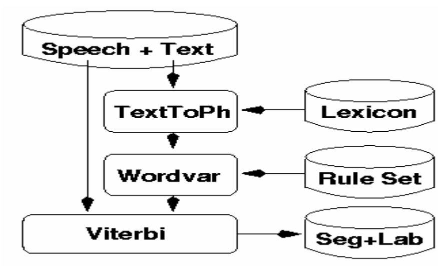

Fig 4 PROCESSING IN MAUS.

E. *Corpus Labeling and Segmentation*

MAUS system use in input a "wav" file and a "text" file. Every file text contains the phonetic transcription. This transcription describes the «wav" file. The file resultant is a file ".par"

. The file transcription is composed of the list of sentence phonemes with their prosodic characteristics.

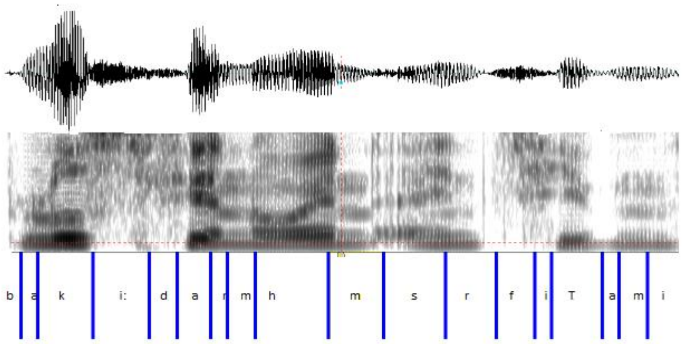

Fig 5 Example MAUS segmentation and labeling taken from the Arabic corpus with SAMPA code.

IV. RESULT AND EVALUATION

A. *Result*

ISSN 2356-5608

A collection of documents collected defined the database result of this work. PADAS database defines as follow:
The files (.wav), four files of transcription (txt, word, phn, textGrid) exist for each sentence of the corpus, which contains respectively:

- The text of the marked sentence (.txt) ;
- The associated time aligned word transcription (.word) ;
- The associated time aligned phonetic transcription (.phn) ;
- Temporal description of each phoneme; start time and end time (.textGrid).

```
File type = "ooTextFile"
Object class = "TextGrid"

xmin = 0
xmax = 2.010000
tiers? <exists>
size = 1
item []:
    item [1]:
        class = "IntervalTier"
        name = "MAU"
        xmin = 0
        xmax = 2.010000
        intervals: size = 18
        intervals [1]:
            xmin = 0.000000
            xmax = 0.320000
```

Fig 5 Temporal description of each phoneme; start time and end time.

B. *Evaluation*

In total, 600 sentences were segmented, 400 sentences for the two speakers (male, 200 sentences for every one), 200 sentences for the third speaker (female). For each segmented 200 sentences, we randomly selected 10 sentences for segmented manually. To do this, we need 6 students in our research laboratory, two for each 10 sentences. The results are summarized in the following table:

TABLE 4: EVALUATION RESULT

speaker	Manual segmentation	Automatic segmentation
First male speaker	99%	94%
second male speaker	99%	94.4%
female speaker	99%	94.1%

V. CONCLUSION

In this paper, it describe our work towards developing the PADAS «Phonetic Arabic Database Automatically Segmented» based on balanced speech corpus and rich phonetic, which is automatic segmented with the MAUS system. This work includes creating the rich phonetic and balanced speech corpus; building an Arabic phonetic dictionary, reducing noise by wavelet method and an evaluation of the automatic segmentation. The current release of our database contains 1 female and 2 male voices. The purpose of this work is to build a database to be used in all area of Speech processing. This variety is useful when used in speech synthesis or speech recognition.

REFERENCES

[1] A. Black and K. Tokuda, “The Blizzard Challenge Evaluating Corpus-Based Speech Synthesis on Common Datasets,” in Proceeding of Interspeech, Portugal, pp. 77-80, 2005.
[2] S. D’Arcy and M. Russell, “Experiments with the ABI (Accents of the British Isles) Speech Corpus,” in Proceedings of Interspeech 08, Australia, pp. 293-296, 2008.
[3] J. Garofolo, L. Lamel , W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue, “TIMIT Acoustic-Phonetic Continuous Speech Corpus,” Technical Document, Trustees of the University of Pennsylvania, Philadelphia, 1993.
[4] K. Tokuda, H. Zen, and A.W. Black, “An HMM-based speech synthesis system applied to English”, in IEEE Speech Synthesis Workshop, 2002.
[5] M. Alghamdi, A. Alhamid, and M. Aldasuqi, “Database of Arabic Sounds: Sentences,” Technical Report, Saudi Arabia, 2003.
[6] M.A. Mansour “Kacst arabic phonetics database”. Riyadh, Kingdom of Saudi Arabia. 2004.
[7] G.Droua-Hamdani “Algerian Arabic speech database (algasd)”. December 2010.
[8] R. Gordon, “Ethnologue: Languages of the World, Texas: Dallas”, SIL International, 2005.
[9] A. Omar “Dirasat Al–Swat Al–Lugawi”.Cairo: Alam Al– Kutub 1985.
[10] L. Pineda, G.mez M., D. Vaufreydaz and J. Serignat “Experiments on the Construction of a Phonetically Balanced Corpus from the Web,” in Proceedings of 5th International Conference on Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science, Korea, pp. 416-419, 2004.
[11] L. Hadjileontiadis and S. Panas. “Separation of discontinuous adventitious sounds from vesicular sounds using a wavelet based filter”, IEEE Trans. Biomed. Eng., vol. 44, n° 7, pp. 876-886, 1997.
[12] S. Mallat. “A wavelet tour of signal processing”. Academic Press, 1999.
[13] F. Schiel, and J. Harrington: “Phonemic Segmentation and Labelling using the MAUS Technique”. Workshop 'New Tools and Methods for Very-Large-Scale Phonetics Research', University of Pennsylvania, January 28-31, 2011.
[14] F. Schiel, “MAUS Goes Iterative”. Proc. of the IV. International Conference on Language Resources and Evaluation, Lisbon, Portugal, pp. 1015-1018. 2004.
[15] J.L. Fleiss “Measuring nominal scale agreement among many raters”. Psychological Bulletin, Vol. 76, No. 5 pp. 378-382. 1971.
[16] S. Burger, K. Weilhammer, F. Schiel, H. G. Tillmann, “Verbmobil Data Collection and Annotation”. Foundations of Speech-to-Speech Translation (Ed.Wahlster W), Springer, Berlin, Heidelberg. 2000.
[17] F. Schiel, C. Heinrich, and S. Barf`ußer “Alcohol Language Corpus”. Language Resources,2011.

ISSN 2356-5608