

Syllabic Segmentation Algorithm of Arabic Word

Faten B. Sofya^{*}

Mahdi F.M. Al-Obadi^{**}

ABSTRACT

An experimental evaluation of an automatic algorithm for segmentation of speech into Arabic syllabic units is reported here. The parameter used for segmentation is the energy and statistical parameter of acoustic signal and depending on the characteristics of the phonology in Arabic language and the fairly close correspondence between the acoustics and phonetics of these syllables. Speech data of four speakers(2 males, 2 females), consisting of mono-syllabic and poly-syllabic words were used to test the segmentation algorithm. It is shown that the final output result of this test is valid for a bout 85%.

(2 2)

. 85%

^{*}Computer Science Dept. - College Of Computer Science and Mathematics-
University of Mosul

^{**}Mathematical Dept. - College Of Computer Science and Mathematics-
University of Mosul

Received: 6/6 /2006 _____ Accepted: 17/4 / 2008

1. Introduction

Arabic language has been alive for the last 4000 years. It is the language of the Holy Quran that is the main preserving reservoir of the language. It is worth noticing that the Arabic letters are distinguished by their beautiful pattern. It is also one of the main languages written from right towards left. Other languages are Persian Hebrew. Urdu. ...,etc. The Arabic letters are cursively written with most of them construct from connecting portions of successive curve fragments with various curvatures, [5].

Segmentation process, and recognition is one the three main phases in general speech systems development. Speech segmentation is the process of partitioning an entire speech into some isolated sub-words with optional boundaries.

The goal of this work is to segment each word into syllable, speech segmentation into syllabic units have several advantages for speech synthesis and recognition. For speech recognition, it involves a recognition vocabulary based on a small group of syllabic units, which can also form the basis for synthesis, also its important because co-articulation effects are already included in these units and also because they represent perceptual units. Syllables may be defined linguistically in terms of the inherent sonority of the each sound. Peaks of sonority generally correspond to peaks of syllabicity (syllable nucleus) but one cannot empirically extract sonority information exist in parallel such as prosody (stress and pitch).

The degree of vocal obstruction in the production of sounds determines the degree of sonority of sounds. The syllable may be defined to be the distance between two minima of sonority Peaks of sonority of ten correspond to vowels. The maximum number of syllables forming a word can never exceed seven, and no more two syllables are allowed in a word of the open type [1, 8].

2. Arabic Properties

The standard Arabic language has basically 35 phonemes, which are 6 vowels and 29 consonant. Standard Arabic language phoneme can be classified into two categories, vowel and consonant as follows: [7]

- Vowels

In Arabic, there are three short vowels /i/, /u/, and /a/, () respectively, which contrast phonemically with their long counterparts /ii/, /uu/ and /aa/ (), which are explained in table(1). The vowels are voiced sound and can be easily identified because of their high energy. The difference between short and long vowel is approximately double or more [7].

Table (1): Arabic Vowels Classification

Tongue hump position Degree constriction	Front	Central	Back
High	/i/ ; /ii/		/u/ ; /uu/
Low		/a/ ; /aa/	

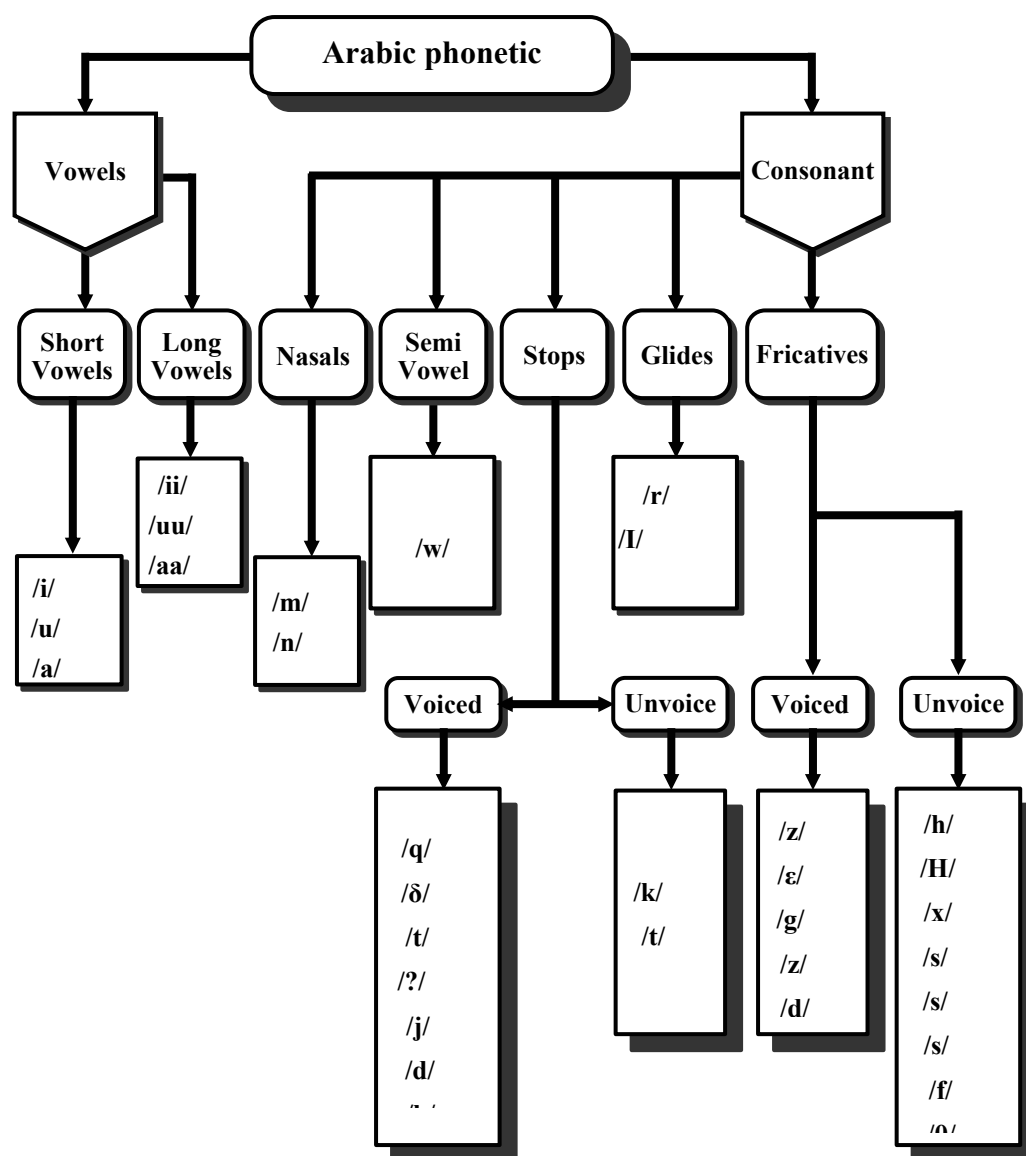
- Consonants

Consonants have less acoustic energy than vowels. The characteristics that form a vowel are relatively more prominent and stable than those of the consonants. The consonants vary individually making it easier to deal with them in groups i.e. nasal, stops, etc. whenever a pharyngealized consonant occurs within a syllable the whole syllable, phonetically, is pharyngealized, each consonant has this one allophone. Also the phenomenon of pharyngealized is not confined to the syllable boundary but may or not have an influence on the neighboring

syllable. This puts the immediate consonantal phonemes preceding and following, in free variation pharyngealized or non pharyngealized. Table(2) shows the classification of Arabic constant.

Voiced consonants positions are in free variation voiced or voiceless, and voiceless consonants intervocalically, are also in free variation voiced or voiceless, [2, 9].

Table (2): Classification of the standard Arabic consonants system



3. Syllables in Arabic

The syllable in Arabic is based on the contrastive components that are contained in its structure. The successive contrastive elements within a syllable boundary are made up of the segmental phonemes of the language. Each syllable has a main part that stands out and has prominence. This part is referred to here as the “nucleus” of the syllable. The remaining components are referred to as ‘marginal factors’. The vowel always forms the syllable nucleus while the consonant represents the marginal phonemes in the syllable structure. The number of syllable in utterance will be identical to the number of vowels [6,7]. There are five syllable patterns as follows:

- | | |
|--------|---------|
| 1. CV | 4. CVVC |
| 2. CVC | 5. CVCC |
| 3. CVV | |

The first four patterns occur initially, medially and finally and the fifth pattern occurs only finally, or in isolation. Syllables can be classified as:

1. Short or long.
2. Closed or open.

In syllable structure, no more than two consecutive consonants are allowed. Arabic utterances can only start with a consonant a special case when two consecutive consonants are the same, C_1 , C_1 , are called consonantal gemination. A vowel can only take place between two consonants or in final position. Geminated consonants CC, can only take place between two vowels, [3].

4. Arabic Segmentation Algorithm

The overall aim of the Arabic syllabic segmentation algorithm is to evaluate the maxima and minima points of the energy curve and statistical measure.

The proposed segmentation algorithm proceeds by performing five steps, namely digital filtering the speech samples, computing the utterance energy, maxima /minima points picking algorithm of the energy, maxima /minima point of statistical parameters and maxima /minima point significance testing. The significant minia point represents the boundary of syllables, figure (1) illustrates outline of segmentation process.

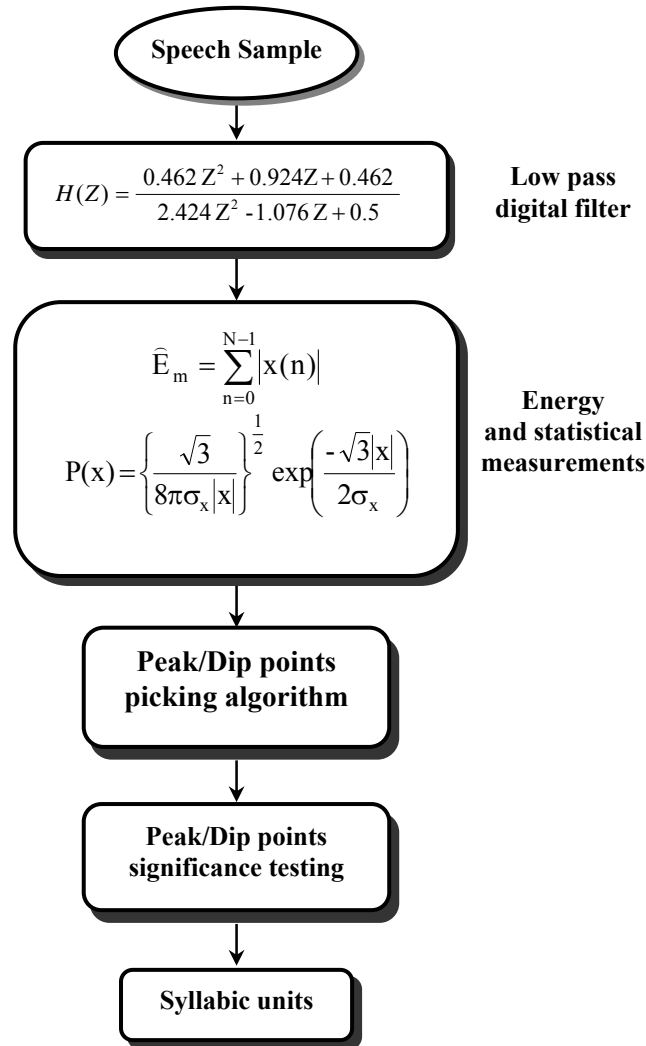


Fig.(1): Flowchart of the Arabic syllabic segmentation algorithm**- Filtering**

In order to smooth the energy of the speech signal, the speech samples of the time signal are low-pass filtered by a recursive digital filter [1] .

$$H(Z) = \frac{0.462 Z^2 + 0.924Z + 0.462}{2.424 Z^2 - 1.076 Z + 0.5} \quad (1)$$

- Speech energy profile

The average magnitude function is used, equation (2). This form of energy definition has the advantage of producing a smooth energy function. The energy $E(k)$ of the speech is calculated in frames of 256 speech samples in length. Windowed by Hamming window $w(n)$ of size 256 and overlapped by half frame length. The end point detector based on energy is found to detect the presence of speech in a background of noise.

$$E_n = \sum_{m=-\infty}^n [x(m) \times w(n - m)]^2 \quad (2)$$

- Maxima / Minima points

Now picking all the maxima and minima points of the energy. The maxima energy point is located when the slop of energy curve change sign from positive to negative and vice versa for the minima energy point. The result is the formation of two array containing the location of all the maxima ($E_{max}(E_n)$) and minima point ($E_{min}(E_n)$).

- Statistical parameters

Gamma probability density function(Gpdf) in equation (3) of the energy is found after normalize energy, so the mean is zero and variance is unity. We can find the maxima($E_{max}(Gpdf)$) /

minima(Emin(Gpdf)) point of Gamma probability by using the same method which is described above [4].

$$P(x) = \left\{ \frac{\sqrt{3}}{8\pi\sigma_x|x|} \right\}^{\frac{1}{2}} \exp\left(\frac{-\sqrt{3}|x|}{2\sigma_x} \right) \quad ; \quad -\infty < x < \infty \quad (3)$$

4.5 Minima point significant

In order to arrive at potentially syllabic nuclei and boundaries, some form of significance testing that to be applied. The proposed method depending on two equations

$$AA = [\max(Gpdf) - \min(Gpdf)] / \text{threshold1} \quad (2a)$$

$$BB = [E_{\min}(E_n) - E_{\min}(Gpdf)] * \text{threshold2}. \quad (2b)$$

Where threshold1 is empirically found approximately (11-15)% of max (Gpdf), and threshold2 is empirically found between (19-25)% of max(Gpdf). The significant minima point is that in BB which is less than AA, which represents the starting and ending frame of each syllabic

Finally, these maxima and minima points must satisfy two conditions, depending on the rules of phonological Arabic system :-

First: between any two consecutive minima point it must contain at least one maxima point.

Second, the difference between any two minima is greater than some threshold (represent the minimal number of frame corresponding to smallest syllabic).

The distance between any consecutive minima represents the length of syllable in the word. Figure (2), (3) and (4) show the boundary of segmentation algorithm of three examples, with the explanation of segmentation algorithm.

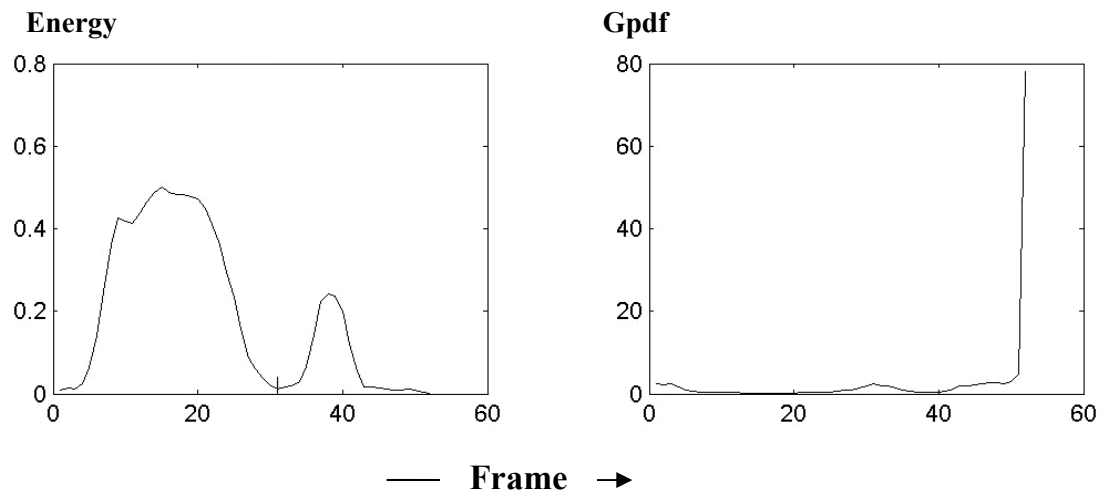


Fig. (2): segmentation of word “ of ” in male voice

Emin = 2, 3, 11, 31, 48, 50

Emax = 9, 15, 38, 49

Significant min. frame = 2, 31, 48
(boundary of syllabic)

Significant max. frame = 15, 38
(number of vowel in word)

Speech segmented = 256, 3968, 6144

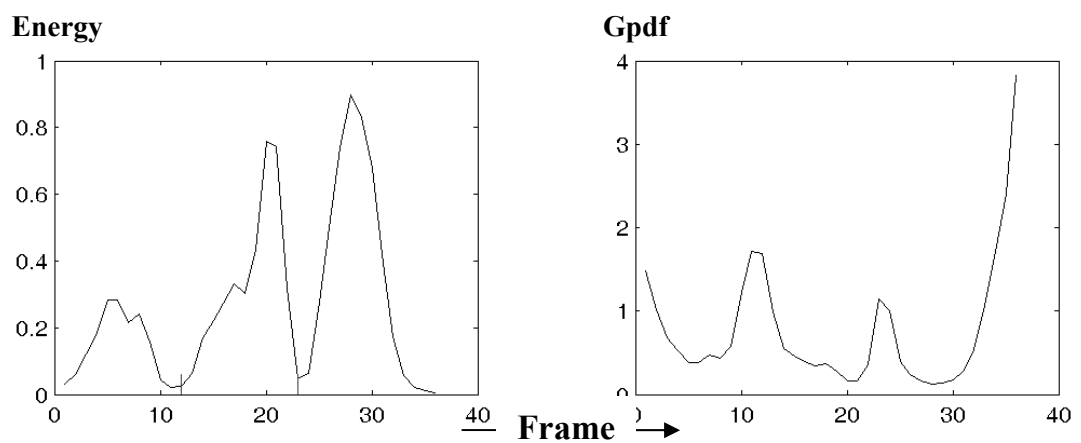


Fig. (3): segmentation of word “ of ” in male voice

Emin = 2, 7, 12, 18, 23, 34

E_{max} = 5, 8, 17, 20, 28

Significant minimum frame = 2, 12, 23, 34

Significant maximum frame = 5, 20, 28

Speech segmented = 256, 1536, 2944, 4352

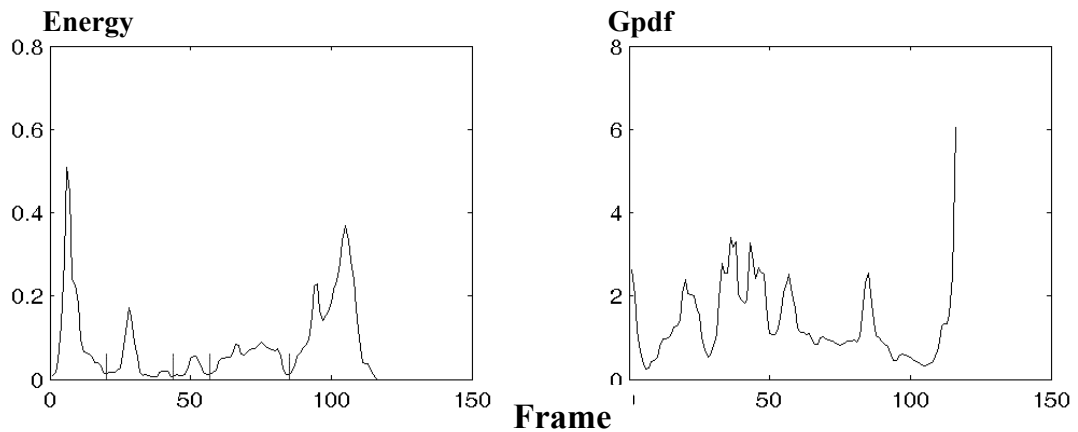


Fig. (4): segmentation of word “ ” in female voice

Emin = 2, 21, 33, 36, 38, 44, 46, 56, 64, 69, 80, 84, 97, 112, 114

E_{max} = 6, 28, 34, 37, 41, 45, 51, 63, 66, 75, 81, 95, 105, 113

Significant minimum frame = 2, 21, 44, 56, 84, 114

Significant maximum frame = 6, 28, 51, 66, 75, 95, 105

Speech segmented = 256, 2688, 5632, 7168, 10752, 14592

5. Conclusion

The performance of the Arabic syllabic segmentation described above was evaluated by four speakers (2-males, 2-females), each speaker uses 125 words to be segmented which contain different number of syllables (between one and five), the results of all speakers are shown in table (3) .

Table (3): Results of segmentation process

	True	False	Percentage %
Male 1	108	17	86.4
Male 2	113	12	90.4
Female 1	100	25	80
Female 2	106	19	84.8

Reference

- [1] Abdulhadi O. (1994), ' *Syllabic Segmentation Algorithm for Arabic Spoken Isolated Words* ', the Arabian journal for science and engineering, vol.19, no.2A.
- [2] Hisham R. and Nedhal A S.(2003), ' *Non-Uniform Speech Segmentation for Arabic Speech Synthesis* ', second international conference on signal system decision and information technology, Tunisia .
- [3] Khaled A. G. and Fahad A. O. (2002), ' *Syllable Based Labeling for Continuous Arabic Speech Recognition* ', kag@acm.org and flotaiby@hotmail.com .
- [4] Lawrence Rabiner and Ronald W. S. (1978) , ' *Digital Processing of Speech Signals* ' ,First Edition, Bell Laboratories .
- [5] Mahir R H. (2002), ' *Font Architecture of Arabic Alphabet* ', Electronic Computers Journal-39.
- [6] Ossama E. (1993), ' *Using Prosody in Automatic Segmentation of Speech* ', <http://www.cs.sc.edu/~essa> .
- [7] Salman H. A.(1970), ' *Arabic Phonology an Acoustical and Physiological Investigation* ', Indiana University, Mouton & Co. N. V. , publishers, The Hague.
- [8] Simon K., Todd S., Stephen I., Paul T. and Alex S. (1998), ' *Speech Recognition Via Phonetically Featured* ', <http://www.cstr.ed.ac.uk>.
- [9] Walid A. M. and Helal M. Y. (2002), ' *A Proposed Algorithm For Speech Segmentation Using The Vector Quantization Technique* ', Electronic Computers Journal vol.39.