The Segmentation of Arabic word signal based on Eigenvalues and Eigenvectors principles

ISSN–1817–2695

Abdul-Hussein M. Abdullah and Esra'a J. Harfash
Department of Computer Science, College of Science, University of Basrah, IRAQ
((Received 6/5/2009, Accepted 2/3/2010))

Abstract

In this paper, we use the principles of eigenvalues and eigenvectors for the first time in the field of Arabic word segmentation. In the beginning, the segmentation algorithm calculates eigenvalues and eigenvectors for each input speech signal, and selects the largest values. The values of final vector are used for linear separation to consonant regions and vowel regions by using threshold value which calculated from the final vector itself. The segmentation points are extracted efficiently and determine the beginning and the end points of each phoneme of the input word signal in high accuracy. Experimental results for a number of isolated Arabic words for a number of speakers (males and females) given in the paper show that the algorithm determines the beginning and the end points for each phoneme in the input speech signals efficiently.

Key words: Segmentation , Speech segmentation, Automatic Speech Recognition (ASR), Automatic segmentation, Arabic speech segmentation **.**

1. Introduction

In the field of Automatic Speech Recognition (ASR), the Automatic segmentation process has drawn some attention in recent years; efficient segmentation algorithm has been an interesting and challenging problem for large best speech segmentation.

The segmentation is the process of determining the boundaries of the sound segmentations which are the locations that divide the segments in the input speech, where the common problem in many areas of speech processing is the identification of the presence or absence of a voice component in a given signal, especially the determination of the beginning and the ending boundaries of voice segments [1]. This task is very important in many applications of (ASR) , then must be achieved in high accuracy with a small amount of error.

The Automatic segmentation is the heart of the acoustic-phonetic recognizer. It deals with the problem of finding best parameters that ensure determination the appropriate boundaries of segments. The success of the segmentation process lays on the correct treatment of the data of the input speech signal, and then we need to spend a big part of the effort on preparing the method able to give high efficient parameters.

There are various control strategies that are used to limit the range of segmentation points. For example, for individual word recognition, the constraint that a word contains at least two phonetic units and no more than six phonetic units means that the control strategy needs to consider solutions between 1 and 5 internal segmentation points.

Generally, speech is produced in the vocal tract which can be modeled as a tube of varying diameter extending from the vocal chords to the lips. The vocal chords produce a periodic pressure wave which travels along the vocal tract. The vocal tract behaves as a resonator in which some frequencies are amplified whereas others are eliminated from the final speech signal. The energy the pressure wave contains increased until it reaches a steady state where the variation is very little, then decreased until the pressure wave is released through the mouth and nose. According to this behavior of the wave through the vocal tract then the segmentation can be achieved [2].

In this paper, we present new segmentation algorithm to detect the boundaries segments by applying the principles of eigenvalue and eigenvector on the input Arabic word signal and then using the resulted data to determine the boundaries of each phoneme effectively.

2. The Arabic Phonemes

Arabic is currently the sixth most widely spoken language in the world with an estimated number of 300 million speakers. Arabic is an official language in more than 22 countries. Since it is also the language of religious instruction in Islam, many more speakers have at least a passive knowledge of the language.

Spoken Arabic language contains 29 consonants and 6 vowels. Three vowels are characterized, as short while the other three are long vowels. In fact, the long and short vowels are very similar but they differ in duration. The long vowels are double in duration of that of the short vowels. The difference between the long and the short vowels is quantitative.

The syllable in Arabic is based on the constructive components that are contained in its structure. The successive constructive 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 as the "nucleus" of the syllable. The remaining components are referred to as "margined factors". Acoustically, the nucleus is represented by formant structures and it has more intensity than the marginal.

The three short vowels and their long counterparts always form the syllable nucleus. All consonants always represent the marginal phonemes of the syllable structure. The marginal phonemes may be represented by either the initiation or the termination of the syllable. The initiation is always a single consonant whereas the termination may be single consonant, two consonants, or zero consonant. There are five syllable patterns. In their representation C stands for consonant, V stands for short vowel and VV stands for long vowel. The five patterns are CV, CVC, CVV, CVVC and CVCC [3,4].

3. The related works

There are many of researches on segmentation process; some of the existing work is surveyed in this section.

Abdul-Hussein has used two empirical parameters which depend on the extracted features of the sound signal in order to gain best segmentation [2].

Esra'a has proposed a segmentation technique to reduce the error ratio in determine the correct segment points; by depending on group of statistical operators which apply on the signal after apply Fourier transformation on this signal [5].

Adell and Bonafonte have presented a new approach to solve the problem of phone segmentation by preparing databases for concatenation Text-to-Speech synthesis, and this approach based on a Regression Tree to perform Boundary Specific Correction of the HMM segmentation [6] .

Almpanidis and Kotropoulos have modeled Speech samples with two-sided generalized Gamma distributed ,they are employed the Bayesian Information Criterion for identifying the phoneme boundaries in noisy speech [7].

Petrushin has described and compared two time domain algorithms for segmentation voiced speech into quasi-periodical units that correspond to pitch periods [8].

4. Eigenvectors and Eigenvalues

An eigenvector of a matrix is a vector such that, if multiplied with the matrix, the result is always an integer multiple of that vector. This integer value is the corresponding eigenvalue of the eigenvector. This relationship can be described by the equation $\boldsymbol{M} \times \boldsymbol{u} = \boldsymbol{\lambda} \times \boldsymbol{u}$, where $\boldsymbol{u}$ is an eigenvector of the matrix $\boldsymbol{M}$ and $\boldsymbol{\lambda}$ is the corresponding eigenvalue. Eigenvectors possess following properties: [9,10,11]

- They can be determined only for square matrices
- There are $\boldsymbol{n}$ eigenvectors (and corresponding eigenvalues) in an $n \times n$ matrix.
- All eigenvectors are perpendicular, i.e. at right angle with each other.

5. Automatic Segmentation of Speech Signal

Our goal is to propose an accurate and efficient algorithm for Arabic speech segmentation applications. Figure (1) shows the block diagram of the proposed algorithm.

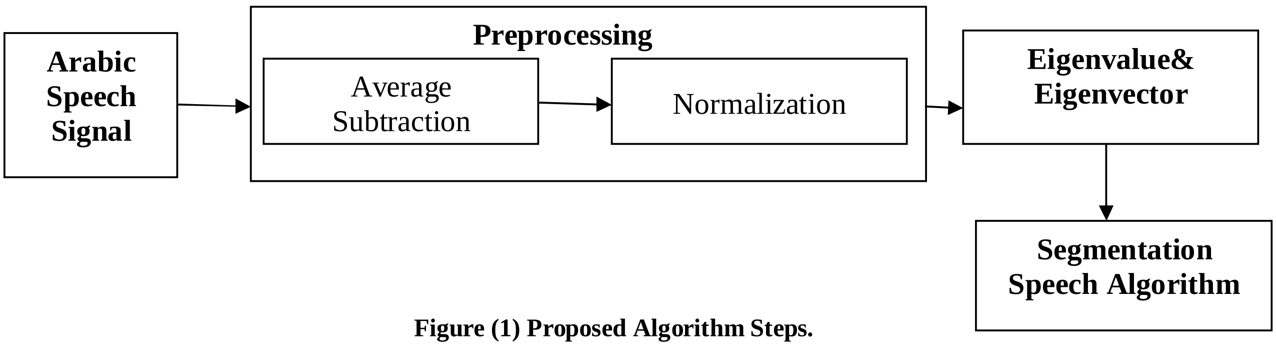

Figure (1) Proposed Algorithm Steps.

The following steps describe the proposed algorithm:

1. **Input speech signal:** A speech signal is recorded in room environment, by using sound forge system to record Arabic speech words. Every sample has length is 8 bits integer; with sampling rate is 8 KHz.

2. **Speech Preprocessing**: Before segmenting the signal into frames, the signal subtraction, which calculates the difference between the original signal samples and the signal average is applied as below:

$$Avg = \frac{1}{n}\sum_{i=0}^{n-1} x(i)$$

$$Y(i) = x(i) - Avg \ , \quad for\ i = 0,1,......n-1$$

Where **x, y** are the original signal and the resulting signal after the average subtraction and **n** is the number of samples. The second stage, the normalization, which divides each sample by the root of the summation of the square samples of the signal as follows:

$$X(i) = \frac{x(i)}{\sqrt{\frac{1}{n}\sum_{i=0}^{n-1} x^2(i)}}$$

In order to extract the N-sample interval from the speech wave an appropriate time window must multiply the speech wave. Hamming window has been applied to each frame.

3. **Eigenvalue Calculation**: After preprocessing, the eigenvalues and eigenvector should be calculated. The description of the exact algorithm for the determination of eigenvectors and eigenvalues is omitted here, as it belongs to the most math programming libraries.

4. **The segmentation Process:** According to the syllabic Arabic speech construction mentioned in section (2), any Arabic word consists of a combination of one or more syllables. For example, the word **/sabâh/** (**CVCVVC**) includes the syllable **/sa/** (**CV**) and the syllable **/bâh/** (**CVVC**). We use the following steps to determine the end point of each phoneme:

   a. Calculate $\lambda_i = (Col1_i + Col2_i)/2$

   Where *Col1* is the resulting vector from eigenvalues, and *Col2* is the resulted vector from eigenvector.

   b. Calculate the threshold **T** as:

$$T = \frac{\sum_{i=0}^{n-1} \lambda_i}{n}$$

c. Identify the regions which satisfy the condition:

$((\lambda i - T) < 0)$ and $((\lambda i_{+1} - T) \geq 0)$

Then select the frame which corresponds to the value $\boldsymbol{\lambda_i}$ as the end point (transition point from consonant to vowel).

d. Identify the regions which satisfy the condition:

$((\lambda i - T) \geq 0)$ and $((\lambda i_{+1} - T) < 0)$

Then select the frame which corresponds to the value $\boldsymbol{\lambda_i}$ as the end point (transition point from vowel to consonant) .

6. Results:

a. According to the previous steps of our algorithm, The first example is word /ʃ**â**ʃ/ **"شاش"** , Figure (2) shows the **λ** with its graph and the final result with the resulting segment points show in Table (1) and Table (2), that satisfy the above conditions:

Table (1) the segmentation points for signal word /ʃâʃ/ "شاش"

sequence	$\boldsymbol{\Lambda}_i$	$\boldsymbol{\lambda}_i$*-T*	$\boldsymbol{\lambda}_{i+1}$	$\boldsymbol{\lambda}_{i+1}$*-T*	*Segmentation points*
1	**0.63993**	**-0.36123**	**1.8306**	**0.82947**	**8**
2	**1.2986**	**0.29741**	**0.66504**	**-0.33612**	**21**

Then the edge of each phoneme can be obtained for the input word /ʃ**â**ʃ/ " شاش" as shown in Table(2):

Table (2) the boundaries of each phonemes in word /ʃâʃ/ "شاش" consequently.

Phoneme	Beginning at frame	Ending at frame
شـ	**1**	**8**
ا	**9**	**21**
ش	**22**	**32**

	Col1	*Col2*	λ
1.	*0.0086374*	*0.047835*	*0.028236*
2.	*0.016239*	*0.052091*	*0.034165*
3.	*0.031519*	*0.049704*	*0.040612*
4.	*0.046661*	*0.066531*	*0.056596*
5.	*0.045803*	*0.075746*	*0.060774*
6.	*0.030445*	*0.059742*	*0.045094*
7.	*0.030253*	*0.063267*	*0.04676*
8.	*1.2207*	*0.059105*	*0.63993*
9.	*1.7779*	*1.8833*	*1.8306*
10.	*1.9626*	*1.7975*	*1.8801*
11.	*2.2982*	*2.5333*	*2.4158*
12.	*2.3952*	*2.5352*	*2.4652*
13.	*2.7307*	*2.3822*	*2.5565*
14.	*2.4994*	*2.4293*	*2.4644*
15.	*2.8873*	*2.2312*	*2.5593*
16.	*2.5974*	*2.4114*	*2.5044*
17.	*2.4464*	*2.7724*	*2.6094*
18.	*2.4086*	*2.4028*	*2.4057*
19.	*2.3057*	*2.2273*	*2.2665*
20.	*1.9661*	*2.1654*	*2.0658*
21.	*1.0819*	*1.5152*	*1.2986*
22.	*0.5024*	*0.82768*	*0.66504*
23.	*0.33321*	*0.57387*	*0.45354*
24.	*0.20348*	*0.41309*	*0.30829*
25.	*0.037867*	*0.11318*	*0.075523*
26.	*0.028607*	*0.05916*	*0.043884*
27.	*0.035335*	*0.052089*	*0.043712*
28.	*0.022443*	*0.088232*	*0.055337*
29.	*0.018135*	*0.054133*	*0.036134*
30.	*0.012501*	*0.044406*	*0.028454*
31.	*0.0091488*	*0.03219*	*0.020669*
32.	*0.0091488*	*0.055009*	*0.032079*

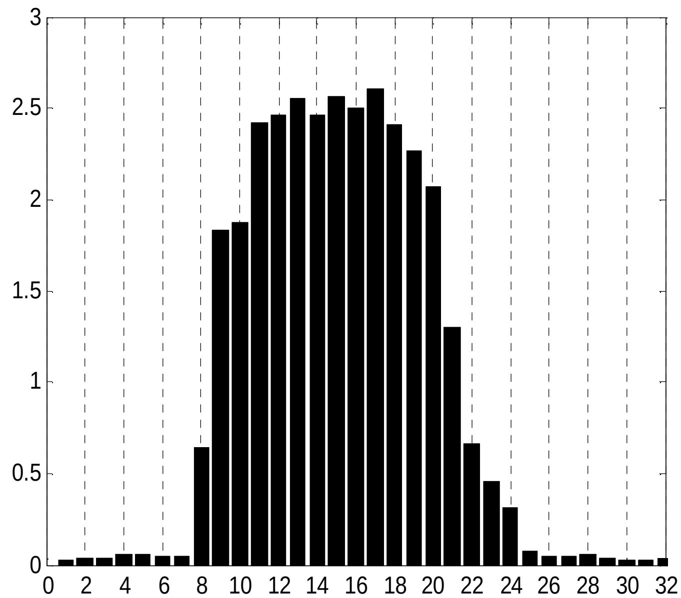

Figure (2) the vector λ and its graph for the word signal "شاش"

b. The second example shows the application of our segmentation algorithm to the isolated word /**has an**/ to obtain the segment boundaries. Figure (3) shows the λ with its graph and the final result in Table (3) and Table (4).

Table (3) the segmentation points for signal word /hasan/ "حَسَن"

sequence	λ_i	λ_i-*T*	λ_{i+1}	λ_{i+1}-*T*	*Segmentation point*
1	**0.16221**	**-0.7422**	**1.1447**	**0.24026**	**9**
2	**1.5297**	**0.62531**	**0.13436**	**-0.77005**	**15**
3	**0.064318**	**-0.84009**	**1.3681**	**0.46366**	**20**
4	**1.1995**	**0.29514**	**0.89601**	**-0.0083982**	**29**

Table (4) the boundaries of each phonemes in word /hasan/ " حَسَن".

Phoneme	Beginning at frame	Ending at frame
حـ	**1**	**9**
الفتحة	**10**	**15**
سـ	**16**	**20**
الفتحة	**21**	**29**
ن	**30**	**42**

	Col1	*Col2*	λ
1.	*0.007549*	*0.052039*	*0.029794*
2.	*0.010882*	*0.035309*	*0.023095*
3.	*0.009898*	*0.047953*	*0.028926*
4.	*0.015937*	*0.053062*	*0.0345*
5.	*0.019145*	*0.036746*	*0.027946*
6.	*0.019601*	*0.080267*	*0.049934*
7.	*0.033972*	*0.067564*	*0.050768*
8.	*0.056315*	*0.059625*	*0.05797*
9.	*0.16482*	*0.15959*	*0.16221*
10.	*1.8859*	*0.40344*	*1.1447*
11.	*2.5775*	*2.6317*	*2.6046*
12.	*3.3126*	*2.4993*	*2.906*
13.	*3.3045*	*2.457*	*2.8807*
14.	*3.8932*	*2.2523*	*3.0727*
15.	*1.8599*	*1.1995*	*1.5297*
16.	*0.049239*	*0.21948*	*0.13436*
17.	*0.018294*	*0.13215*	*0.075224*
18.	*0.018117*	*0.11395*	*0.066034*
19.	*0.017803*	*0.086174*	*0.051988*
20.	*0.019132*	*0.1095*	*0.064318*
21.	*2.6357*	*0.10042*	*1.3681*
22.	*4.068*	*2.2505*	*3.1592*
23.	*3.5917*	*2.309*	*2.9504*
24.	*2.5596*	*2.5256*	*2.5426*
25.	*1.6938*	*1.8006*	*1.7472*
26.	*1.7869*	*1.3298*	*1.5584*
27.	*1.9084*	*1.1543*	*1.5314*
28.	*1.7473*	*1.4607*	*1.604*
29.	*1.1314*	*1.2677*	*1.1995*
30.	*0.9592*	*0.83282*	*0.89601*
31.	*0.68203*	*0.90305*	*0.79254*
32.	*0.44709*	*0.8685*	*0.65779*
33.	*0.40718*	*0.74486*	*0.57602*
34.	*0.27541*	*0.75003*	*0.51272*
35.	*0.23695*	*0.65787*	*0.44741*
36.	*0.15662*	*0.56208*	*0.35935*
37.	*0.14119*	*0.44318*	*0.29218*
38.	*0.12533*	*0.48468*	*0.305*
39.	*0.062943*	*0.34021*	*0.20158*
40.	*0.053763*	*0.30528*	*0.17952*
41.	*0.017551*	*0.14577*	*0.081659*
42.	*0.017551*	*0.036705*	*0.027128*

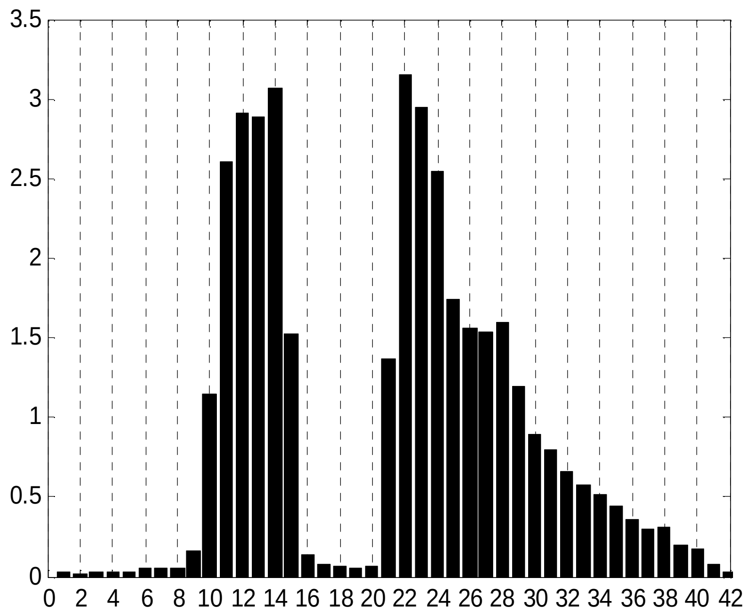

Figure (3) the vector λ and its graph for the word signal "حَسَن".

7. Conclusions

In this work, the principles of eigenvalue and eigenvectors are suggested for first time to use in a segmentation algorithm to purpose of determining the segment points. The results are shown that this segmentation algorithm has ability for find the boundaries segments with more high efficiency, because of it has a strong ability that related with the information that included in the eigenvalue and eignvector which have efficient indicators about the segmentation points.

We apply our new algorithm at 50 voices of different words, and we find that the successful percent is 80%, in compare with another algorithm that depends on use a group of statistical operators, which used with the same 50 voice and the successful percent was 72% [5].

References

[1] G. Almpanidis and C. Kotropoulos, " ***Voice Activity Detection using Generalized Gamma Distribution*",** Department of Informatics, Aristotle University of Thessaloniki, Box 451 Thessaloniki, GR-54124, Greece, [2005].

[2] A. M. Abdullah, " ***Hybrid Approach for Arabic Speech Signal Recogniti on***", Ph.D. Thesis, Department of Computer Science, College of Science, University of Basrah, Iraq, [2000].

[3] عمر ، أحمد مختار ،"***دراسة الصوت اللغوي***"،1976 ، جامعة الكويت.

[4] بركة ،بسام ،"***علم الاصوات العام:أصوات اللغة العربية***" ، 1988، مركز الانماء.

[5] E. J. Harfash, "***Using group of statistical operator toward optimal segmentation of Arabic word signal to its basic phonemes***", Journal of Basrah Research (sciences), Vol. 34, No. 4, [2008].

[6] J. Adell and A. Bonafonte, " ***Towards phone segmentation for concatenation speech synthesis*** ", Department of Signal Theory and Communication, TALP Research Center, Technical University of Catalonia (UPC), Barcelona (Spain), [2004].

[7] A. George and K. Constantine, " ***Voice Activity Detection using Generalized Gamma Distribution*** ", Department of Information, University of Aristotle, Box 451 Thessaloniki,GR-54124, Greece, [2003].

[8] A. P. Valery, "***Adaptive Algorithms for Pitch-synchronous Speech signal Segmentation***", Accenture Technology Labs, 161 N. Clark St., Chicago, IL 60601, USA, [2003].

[9] D. Pissarenko, " ***Eigenface-based facial recognition***", December 1, [2002].

[10] B. Zhou and J. H. L. Hansen, " ***A Novel Algorithm For Rapid Speaker Adaptation Based On Structural Maximum Likelihood Eigenspace Mapping*",** The Center for Spoken Language Research, University of Colorado at Boulder, Boulder, CO, 80302, U.S.A, [2001].

[11] K. Patrick, B. Gilles, and D. Pierre, " ***Eigenvoice Modeling with Sparse Training Data***" ,IEEE Transactions on Speech and Audio Processing, Vol. 13, No. 3, May, [2005].

تقطيع أشارة الكلمة العربية بألاعتماد على مبادئ القيم والمتجهات الذاتية

عبدالحسين محسن عبد الله إسراء جاسم حرفش
قسم علوم الحاسبات-كلية العلوم-جامعة البصرة

الخلاصة

في هذا البحث أستخدمنا مبادئ القيم والمتجهات الذاتية لاول مرة في مجال تقطيع صوت الكلمة العربية . في البدء تحسب خوارزمية التقطيع القيم الذاتية والمتجهات الذاتية لكل أشارة كلام داخلة ، ثم تنتقى القيم الاكبر. أن قيم المتجه النهائي الناتج تستعمل لاجل الفصل الخطي الى مناطق صوت ساكنة ومناطق صوت معتلة بأستعمال قيمة عتبة محسوبة من المتجه النهائي نفسه ،أن نقاط القطع تستخرج هنا بشكل كفوء، ومن خلالها تحدد نقاط البداية والنهاية لكل فونيم من أشارة الكلمة الداخلة وبدقة عالية. أن النتائج التجريبية لعدد من الكلمات العربية المعزولة ولعدد من المتكلمين (ذكور واناث) أظهرت أن هذه الخوارزمية لها القدرة على تحديد نقاط البدء والنهاية لكل فونيم في أشارة الكلام الداخل بكفاءة عالية جداً.