International Journal of Engineering Sciences Paradigms and Researches (IJESPR)
(Vol. 29, Issue 01) and (Publishing Month: April 2016)
(An Indexed, Referred and Impact Factor Journal)
ISSN: 2319-6564
www.ijesonline.com

Phonemes based Speech Word Segmentation using K-Means

Abdul-Hussein M. Abdullah[1] and Esra Jasem Harfash[2]

[1, 2]Department of Computer Science, College of Science, University of Basrah, IRAQ
[1]Abdo60_2004@yahoo.com and [2]Esra_jasem_2011@yahoo.com

Publishing Date: April 25, 2016

Abstract

Phoneme Speech segmentation is an important task in more speech sound processing applications. In this work uses k-mean algorithm to segment speech word sound to its phonemes .We are run k-means to separate between vowel regions and another the consonant regions in the input word signal. Then the segments points between vowel phoneme and consonant phonemes can be determined, and each phoneme can extracted easily. We are measure the performance of k-means by using different data type of waveform of word: Time domain, FFT, and wavelet transformation. We apply our system on 100 different words, then the accuracy of determine succeed segment points is 80.33%

Keywords: *Audio segmentation, Automatic Speech Segmentation, Clustering, K-Means Algorithm.*

1. Introduction

Automatic speech segmentation has many advantages in more applications in speech processing, e.g., in automatic speech recognition and automatic annotation of speech corpora[1].The good of the segmentation affects the recognition performance in several ways: Speaker adaptation and speaker clustering methods assume that a segment is spoken by a single speaker. The language model performs better if segment boundaries correspond to boundaries of sentence-like units [2].

This development to the speech systems created a demand for new and better speech databases (using new voices, new dialects, new special features to consider, etc.), often with phonetic level annotation information (and others). This trend re-enforces the importance of automatic segmentation and annotation tools because of the drastic time and cost reduction in the development of speech corpora even when some little human action is needed[3].

Speech can be represented phonetically by a limit set of symbols called the phonemes of the language, the number of which depends upon the language and the refinement of the analysis. For most languages the number of phonemes lies between 32 and 64. Each phoneme is distinguished by its own unique pattern and different phonemes are distinguishable in terms of their formant frequencies[4].Speakers of a language can easily dissect its continuous sounds into words. With more difficulty, they can split words into component sounds segments (phonemes).

The phoneme segmentation is an approach to isolating component word sounds to its 'distinctive unit sounds' or phonemes. Then the Automatic Speech Segmentation is the process of taking the phonetic transcription of an audio speech segment and determining where in time particular phonemes occur in the speech segment, by using appropriate algorithms in computer[5].

One good approach that can be used in segmentation process is the principle of clustering. Among the formulations of partitional clustering based on the minimization of an objective function, k-means algorithm is the most widely used clustering and studied. Where each data object must be describable in terms of numerical coordinates. This algorithm partitions the data points (objects) to C groups (clusters), so as to minimize the sum of the (squared) distances between the data points and the center (mean) of the clusters [6,7].In this paper, a tool for automatic phoneme segmentation using k- means algorithm.

2. K-Means Clustering

Clustering is an unsupervised classification that is the partitioning of a data set in a set of

meaningful subsets. Each object in dataset shares some common property- often proximity according to some defined distance measure. Among various types of clustering techniques, K-Means is one of the most popular algorithms. The objective of K-means algorithm is to make the distances of objects in the same cluster as small as possible.

K-means is a prototype-based, simple partition clustering technique which attempts to find a user-specified k number of clusters. These clusters are represented by their centroids. It Divide n object into this K clusters, to create relatively high similarity in the cluster and, relatively low similarity between clusters. And minimize the total distance between the values in each cluster to the cluster center. A cluster centroid is typically the mean of the points in the cluster. This algorithm is simple to implement and run, relatively fast, easy to adapt, and common in practice[8,9].

The basic steps of k-means clustering are simple. In the beginning we determine number of cluster and we assume the centroid or center of these clusters. We can take any random objects as the initial centroids or the first k objects in sequence can also serve as the initial centroids. The k-means algorithm will do the two steps below until convergence:

1. Each instance Xi is assigned to its closest cluster.
2. Each cluster center Cj is updated to be the mean of its constituent instances. Where and the K selected initial cluster means.

This algorithm aims at minimizing an *objective function*, (in this case a squared error function). The objective function, where is a chosen distance measure between a data point and the cluster Centre , is an indicator of the distance of the *n* data points from their respective cluster centers[10,11,12].

The main steps of k-means clustering algorithm can be describe as follows[9,13]:

1. Randomly select k data object from dataset D as initial cluster centers.
2. Repeat:
    a. Calculate the distance between each data object d_i, where (1<=i<=n) and all k cluster centers c_j, where (1<=j<=k), and assign data object d_i to the nearest cluster.
    b. For each cluster c_j, recalculate the cluster centers.
    c. Until no change in the cluster Centre.

3. Segmentation Framework

The k-mean has been used here to determine the segmentation points in the sound word signal. And does so by grouping *frames* of the signal X into two groups, one of them for vowels and the other for consonants. In the following, the steps that is followed to determine the segments points:

1. Input: Each input word speech signal X is recorded in the environment of the room. The sampling rate is 8KHz and eachsampleof8 bits length.
2. Preprocessing: this stage includes the following stapes:
    a. Normalize the speech signal as follows:

$$X_i = \frac{x_i}{\sqrt{\frac{1}{n}\sum_{i=0}^{n-1} x_i^2}} \quad (1)$$

where x(i) is sample in the sound signal , n is the overall number of samples.

    b. Divide the speech signal *X* into *N* blocks (*frame*$_1$, *frame*$_2$, ...,*frame*$_N$) where each block of length *M* samples, using Hamming window.

3. Run the k-means:
    a. Generate the initial values of the centers C_1 and C_2 randomly, where the length of C_i is M .
    b. Calculate the distance between the each $frame_i$ and the centers C_1 and C_2 separately:

$$D_{1i} = \sqrt[2]{(frame_{i1} - C_{11})^2 + (frame_{i2} - C_{12})^2 + \cdots + (frame_{iM} - C_{1M})^2} \quad (2)$$

$$D_{2i} = \sqrt[2]{(frame_{i1} - C_{21})^2 + (frame_{i2} - C_{22})^2 + \cdots + (frame_{iM} - C_{2M})^2} \quad (3)$$

Where i=1 to N.

    c. Select the minimum distance value between (D_{2i}, D_{1i}) to identify which cluster the $frame_i$ belong to it.
    d. According to a new distribution of the frames on the two groups, the values of the centers (C1, C2) is recalculated as follows:

$$C_1 = \sum_{j=1}^{z}\sum_{i=1}^{P} A_{ij} / p \quad (4)$$

$$C_2 = \sum_{j=1}^{z}\sum_{i=1}^{q} B_{ij} / q \qquad (5)$$

Where p is the number of frames in the cluster1 and q is the number of frames in the cluster2.

e. Repeat b, c and d respectively, until the model stable.

In this work, the k-means algorithm implemented on *three types of features* extracted from word speech signal, are:

- Feature set extracted in time domain of sound signal.
- Fast Fourier Transform Coefficients.
- Wavelet Transform coefficients.

4. Discussion and the experimental results

Through the experiments ,we are tested to determine the appropriate type of features that more efficient in giving good separation between vowels and consonant regions frames .The following Discussion show the overall results of phonemes segmentation after run the k-means .

(A) In time Domain

According to this type of data, the following cases of experiments carried out:

Case 1: The input to the k-means is the N of frames, each frame with full M samples.
Case 2: Take the average of each frame, then the input to k-means is N frames with one value. For the word signal in Fig. (1), Fig. (2) show the distributed of frames of this word between the vowel and consonant regions ,and table 1 shows the measurement accuracy of these above two case:

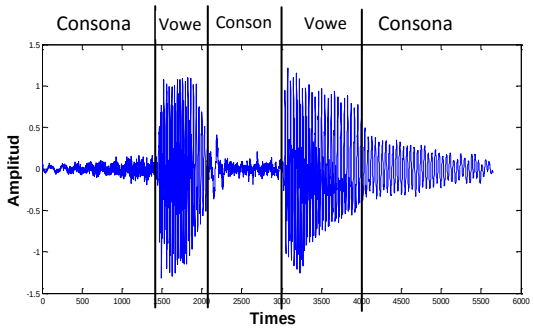

Figure 1 The input signal of word "/hasan/"

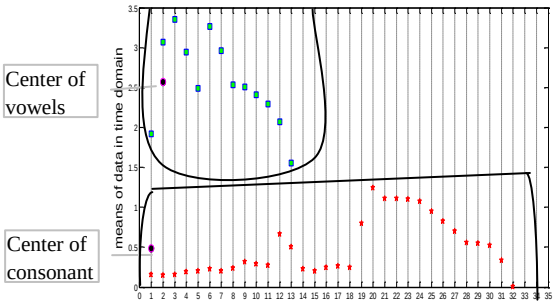

Figure 2 The results of distribute of clusters of Data in time domain

Table 1

Case	Case1	Case2
TD	50.21	90.89

(B) In Fourier transform

After apply the Fast Fourier transform calculation on each frame, the output for each is frames M / 2 coefficients. These coefficients are the input data adopted here .And the following cases have been tried with these transactions:

Case 1:, The input to the k-means is N of points each point is the mass of a length M/2coefficients.
Case 2: Take the average of each frame of FFT coefficients, then the input to the k-means is the N of points each point along a single value of mean coefficients.
Case 3: Reduce the number of coefficients of FFT to the length of (M / 2) / r, where $r=2^{no}$ and *no* is an integer, and r should be less than or equal to (M / 2).
This reduction process performed by taking the largest value out of each r coefficient, as follows:

$$\frac{E_{i1}, E_{i2}, E_{i3}, \dots E_{ir},}{max1} \frac{E_{ir+1}, E_{ir+2}, E_{ir+3}, \dots E_{i2r},}{max2} \dots\dots\dots \frac{E_{i1}, E_{i2}, E_{i3}, \dots E_{iM/2}}{\max\left({}^{m/2}/_{r}\right)}$$

Then the input here to the k-means is the N of points each point along the mass (M / 2) / r. Case 4:Also we take the Average coefficients in case 3.

Figure (3) show the result distribute of frames of the word in figure(1) depending on this type of features, and table 2 show the measurement accuracy:

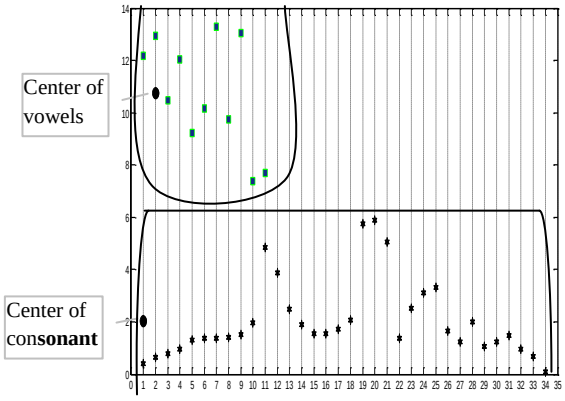

Figure 3 results of distribute of clusters of Fast Fourier Transform coefficients

Table2

Case	Case1	Casee2	Casee3	Case4
FT	47.14	77.85	68.39	80.18

(C) In Wavelet transformation

We are use Discrete Wavelet transformation DWT to performs a 4-level one-dimensional wavelet decomposition with respect to the wavelet db3, where DWT computes the approximation coefficients vector cA and detail coefficients vector cD, obtained by a wavelet decomposition as show in Figure(4).

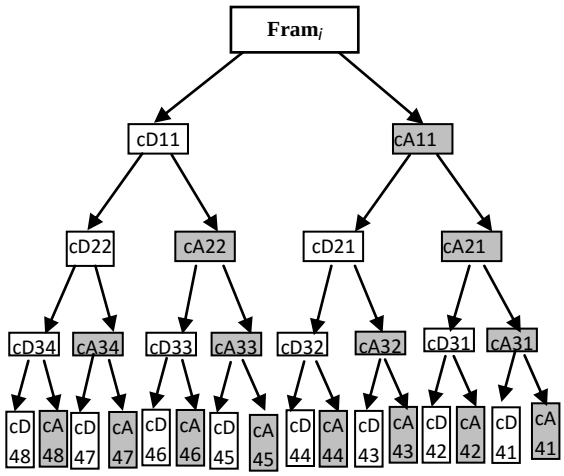

Figure 4 The wavelet transformation to four level

The following the experiments that give the best result on the wavelet coefficients:
Case 1: Take the mean of cA11, cA21, cA31 and cA41, the input to the K-means Nx4 vector values.

Case2:Take the mean of cA11,cD11 ,cA21,cD21,cA31,cD31 , cA41,cD41, then the input is Nx8.
Case 3*:* Take the mean of each nodes in the net in fig(4), the result is Nx30 vector.
Figure (5) and table 3 shows the result with this type of features.

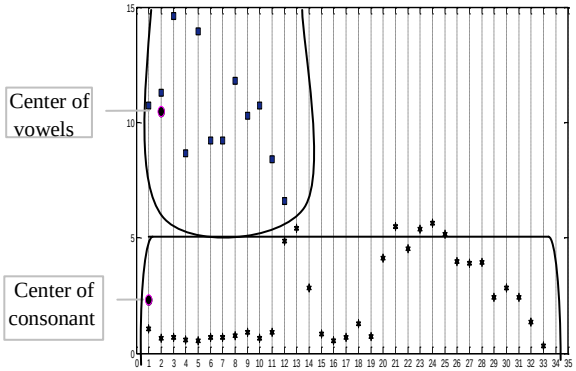

Figure 5 The results of distribute of clusters of Wavelet Transform coefficients

Table 3

Case	Case1	Case2	Case3
WT	94.036	92.786	86.724

5. Conclusion

We presented a approach of phoneme segmentation Depending on several types of features using K-means algorithm, regarding the use of natural speech recorded in real situations, and we are found the following:

1. This method gives good ability to separate the vowel phonemes (in one cluster) and the consonant phonemes in another cluster.
2. The ability of k-means model is increased when the dimension of each input data point X_i is smallest to small or one point by take the mean (or may standard deviation or variance ...ect) to this input data, where the features are became more clear .
3. In all the cases, as we see in above section 4, the performance always is good and acceptable, but the best result, we are obtain when dealing with wavelet coefficients.

References

[1] O. Johannes, U. Kalervo, and T. Altosaar,"An Improved Speech Segmentation Quality Measure: the R-value", Department of Signal Processing and Acoustics, Helsinki University of Technology, Finland, 2008.

[2] F. Kubala, T. Anastasakos, H. Jin, L. Nguyen, and R. Schwartz, "Transcribing radio news" in Proc. ICSLP, Philadelphia, PA, USA, Oct., pp. 598–601, 1996.

[3] Luis Pinto," AUTOMATIC PHONETIC SEGMENTATION AND LABELLING OF SPONTANEOUS SPEECH", Zaragoza, Del 8 al 10, November, journal of Technology, Habla, 2006

[4] M. Sarma and K. K. Sarma, "Segmentation of Assamese phonemes using SOM", Conference Paper January, 2012.

[5] B. Bigi, "Automatic Speech Segmentation of French: Corpus Adaptation, "LPL - Aix-en-Provence – France, 2012.

[6] J. Burkardt ,"K-means Clustering", Advanced Research Computing, Interdisciplinary Center for Applied Mathematics, Virginia Tech, September, 2009.

[7]. M. B. Al- Zoubi, A. Hudaib, A. Huneiti and B. Hammo ," New Efficient Strategy to Accelerate k-means clustering Algorithm", American Journal of Applied Science 5 (9): 1247-1250, ISSN 1546-9239, 2008 .

[8] R. Yadav and A. Sharma," Advanced Methods to Improve Performance of K-Means Algorithm: A Review", Global Journal of Computer Science and Technology Volume 12 Issue 9 Version 1.0 April, 2012.

[9] H.S. Behera, A. Ghosh,and S. K. Mishra ," A New Improved Hybridized K-MEANS Clustering Algorithm with Improved PCA Optimized with PSO for High Dimensional Data Set" , International Journal of Soft Computing and Engineering (IJSCE) , Volume-2, Issue-2, May ,2012.

[10] K. Teknomo, "Numerical example of k-means clustering", CNV media, 2006.

[11] K. Wagstaff ,C. Cardie , S. Rogers ,and S. Schroedl , "Constrained K-means Clustering with Background Knowledge " , Proceedings of the Eighteenth International Conference on Machine Learning, p. 577-584, 2001.

[12] R. C. de Amorim,"Learning feature weights for K-Means clustering using the Minkowskimetric", Department of Computer Science and Information Systems Birkbeck, University of London ,April,2011.

[13] O. Nagaraju, B.Kotaiah , R.A. Khan and M.RamiReddy ," Implementing and compiling clustering using Mac Queens alias K-means apriori algorithm" , International Journal of Database Management Systems (IJDMS) Vol.4, No.2, April 2012.