DYNAMIC THRESHOLDING ON SPEECH SEGMENTATION

Md. Mijanur Rahman[1], Md. Al-Amin Bhuiyan[2]

[1]*Assistant Professor, Dept. of Computer Science & Engineering, Jatiya Kabi Kazi Nazrul Islam University, Bangladesh*
[2]*Professor, Dept. of Computer Science & Engineering, Jahangirnagar University, Bangladesh*
mijan_cse@yahoo.com, alamin_bhuiyan@yahoo.com

Abstract

Word is the preferred and natural unit of speech, because word units have well defined acoustic representation. This paper presents several dynamic thresholding approaches for segmenting continuous Bangla speech sentences into words/sub-words. We have proposed three efficient methods for speech segmentation: two of them are usually used in pattern classification (i.e., k-means and FCM clustering) and one of them is used in image segmentation (i.e., Otsu's thresholding method). We also used new approaches blocking black area and boundary detection techniques to properly detect word boundaries in continuous speech and label the entire speech sentence into a sequence of words/sub-words. K-Means and FCM clustering methods produce better segmentation results than that of Otsu's Method. All the algorithms and methods used in this research are implemented in MATLAB and the proposed system achieved the average segmentation accuracy of 94% approximately.

Keywords: Blocking Black Area, Clustering, Dynamic Thresholding, Fuzzy Logic and Speech Segmentation.

--***--

1. INTRODUCTION

Automated segmentation of speech has been a subject of study for over 30 years [1] and it plays a central role in many speech processing and ASR applications. It is important to various automated speech processing algorithms: speech recognition, speech corpus collection, speaker verification, etc., in the research field of natural language processing [2,3]. The traditional or manual segmentation approach is impractical for very large data bases. It is extremely laborious and tedious. This is why automated methods are widely utilized. Common approach to speech signal zone identification is using a threshold value. In many papers speech segmentation is done by using wavelet [4], fuzzy methods [5], artificial neural networks [6] and hidden Markov models [7]. This paper proposed several pattern classification and image segmentation methods for segmenting continuous Bangla speech into words/sub-words. To do so, we have used the modified version of k-means, fuzzy c-means and Otsu's algorithms with blocking black area method and shape identification.

The paper is organized as follows: Section 1 describes the introduction of speech processing and the organization of this paper. In Section 2, we will discuss about speech segmentation and types of segmentation. Section 3 will describe pattern clustering and different clustering algorithms. In Section 4, Otsu's thresholding method will be discussed. The methodological steps of the proposed system will be described in Section 5. Sections 6 and 7 will describe the experimental results and conclusion, respectively.

2. SPEECH SEGMENTATION

Speech recognition system requires segmentation of speech signal into discrete, non-overlapping acoustic units [8][9]. It can be the segment, phone, syllable, word, sentence or dialog turn level. Word is the preferred and natural units of speech because word units have well defined acoustic representation. So, we have been chosen word as our basic unit. Automatic speech segmentation methods can be classified in many ways, but one very common classification is the division to blind and aided segmentation algorithms [10]. A central difference between aided and blind methods is in how much the segmentation algorithm uses previously obtained data or external knowledge to process the expected speech. The term blind segmentation refers to methods where there is no pre-existing or external knowledge regarding linguistic properties, such as orthography or the full phonetic annotation, of the signal to be segmented. On the other hand, Aided segmentation algorithms use some sort of external linguistic knowledge of the speech stream to segment it into corresponding segments of the desired type. Due to the lack of external information, the first phase of blind segmentation relies entirely on the acoustical features present in the signal. The second phase is usually built on a front-end processing of the speech signal using MFCC, LP-coefficients, or pure FFT spectrum [11]. In our research works, we have used FFT spectrum (such as, spectrogram) features of speech signal.

3. CLUSTERING

Clustering is the process of assigning a set of objects into a set of disjoint groups called clusters so that objects in each same cluster are more similar to each other than objects from

different clusters. Clustering techniques are applied in many application areas such as pattern recognition [12], data mining [13], machine learning [14], etc. Clustering algorithms can be broadly classified as Hard, Fuzzy, Possibilistic, and Probabilistic [15], each of which has its own special characteristics. There are a lot of applications of clustering methods, range from unsupervised learning of neural network, Pattern recognitions, Classification analysis, Artificial intelligent, image processing, machine vision, etc. Several clustering methods have been widely studied and successfully applied in image segmentation [16–22].

In this paper, the hard and fuzzy clustering schemes have been introduced to get the optimal threshold in speech segmentation. The conventional hard clustering method restricts each point of the data set to exclusively just one cluster. So the membership values to be either 0 or 1. The K-Means clustering algorithm is widely used as hard clustering method. On the other hand, the fuzzy clustering method allows an object to belong to several clusters at the same time, but with the different degrees of membership between 0 and 1. The fuzzy C-means algorithm is the most popular fuzzy clustering method used in pattern classification.

3.1 K-Means Clustering Algorithm

K-Means Clustering is one of the unsupervised learning algorithms, first used by James MacQueen in 1967 [23] and first proposed by Stuart Lloyd in 1957 as a technique for pulse-code modulation [24]. It is also referred to as Lloyd's algorithm [25]. Here, the modified algorithm is used to compute the threshold from speech spectrogram. The algorithm classifies a given data set into a certain number of clusters (i.e., k clusters) based on distance measures and to define k-centers, one for each cluster and finally computes the threshold. This algorithm is an iterative process until no new moves of data, as shown in Figure-1. Finally, this algorithm aims at minimizing an objective function, also known as squared error function and given by:

$$J = \sum_{i=1}^{k} \sum_{j=1}^{c_i} \left(\left\| x_i^{(j)} - v_j \right\| \right)^2$$

Where:

$\left\| x_i^{(j)} - v_j \right\|$ is the Euclidean distance between a data point (x_i) and the cluster center, (v_j);
$'c_i'$ is the number of data points in i^{th} cluster; and
$'k'$ is the number of cluster centers.

The algorithmic steps for calculating a threshold using k-means clustering is given below:
Let $x = \{x_1, x_2, x_3,…,x_n\}$ be the set of data values in speech spectrogram.

1) Assign the number of clusters, $k=3$ and define the 3 cluster centers.
    Let $v = \{v_1, v_2, v_3\}$ be the set of 3 cluster centers.
2) Calculate the distance (d_{ij}) between i^{th} data and j^{th} cluster center.
3) Assign each data to the closest cluster with minimum distance.
4) Recalculate the new cluster center using:

$$v_i = \frac{1}{c_i} \sum_{j=1}^{c_i} x_j$$

    where, $'c_i'$ represents the number of data points in i^{th} cluster.
5) Recalculate the distance between each data and new obtained cluster centers.
6) If no data was reassigned then stop, otherwise repeat from step (3).
7) Calculate the desired *threshold* from the average of 3 cluster centers.
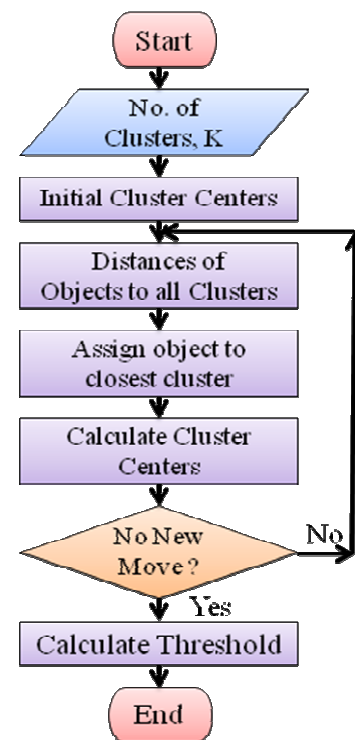
Figure1. K-Means Clustering Algorithm - Calculating the desired threshold.

3.2 FCM Clustering Algorithm

The algorithm is developed by J C Bezdek in 1973 for pattern classification [26], and first reported by J C Dunn in 1974 [27], later J C Bezdek used this algorithm for pattern recognition [28]. FCM is a generalization of K-Means. While K-Means assigns each object to one and only one cluster,

FCM allows clusters to be fuzzy sets, so that each object belongs to all clusters with varying degrees of membership and the following restriction: The sum of all membership degrees for any given data point is equal to 1. It works by assigning membership to each object on the basis of distance between the cluster center and the object. More the object is near to the cluster center more is its membership. The algorithm is an iteration process, as shown in Figure-2. After each iteration, membership and cluster centers are updated, and finally compute a threshold from the cluster centroids.

The algorithmic step for calculating a threshold using fuzzy c-means clustering is given below:

Let $x = \{x_1, x_2, x_3, ..., x_n\}$ be the set of data values in speech spectrogram.

1) Assign the number of cluster, $c=3$ and define the cluster centers.
   Let $v = \{v_1, v_2, v_3\}$ be the set of 3 cluster centers.
2) Calculate the Euclidean distance (d_{ij}) between i^{th} data and j^{th} cluster center.
3) Update the fuzzy membership function (μ_{ij}) using:

$$\mu_{ij} = 1 \bigg/ \sum_{k=1}^{c} \left(d_{ij} / d_{ik} \right)^{2/(m-1)}, \forall i = 1,2,...n; \ \forall j = 1,2,...c$$

4) Compute the fuzzy centers $'v_j'$ using:

$$v_j = \left(\sum_{i=1}^{n} (\mu_{ij})^m x_i \right) \bigg/ \left(\sum_{i=1}^{n} (\mu_{ij})^m \right), \forall j = 1,2,3,...,c \ \text{wh}$$

ere, m is the fuzziness index $[1,\infty]$.

5) Repeat from step (2) until the minimum $'J'$ value is achieved or $\left\| U^{(k+1)} - U^{(k)} \right\| < e$.
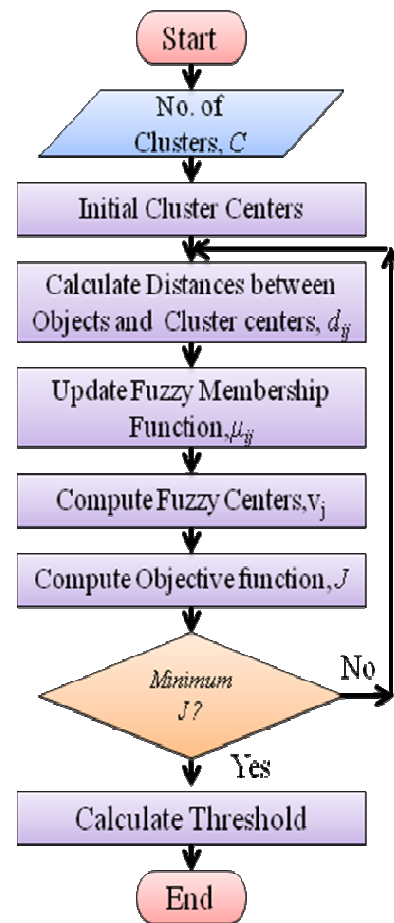
   where:
   k is the iteration step;
   e is the termination criterion between $[0, 1]$;
   $U = (\mu_{ij})_{n*c}$ is the fuzzy membership matrix; and
   J is the objective function, obtained by:

$$J = \sum_{j=1}^{c} \sum_{i=1}^{n} (\mu_{ij})^m d_{ij}$$

6) Calculate the desired *threshold* from the average of 3 cluster centroids.

Figure2. FCM Clustering Algorithm - Calculating the desired threshold.

4. OTSU'S THRESHOLDING METHOD

Thresholding is the simplest method of image segmentation; can be used to create binary image from gray scale image. In order to convert the image into a binary representation, first we converted the image into a gray scale representation and then performed particular threshold analysis [29]. Otsu's method is a simple and effective automatic thresholding method, invented by Nobuyuki Otsu in 1979 [30], also known as binarization algorithm. It is used to automatically perform histogram shape-based image thresholding; i.e. the reduction of a grayscale image into a binary image. The algorithm assumes that the image is composed of two basic classes: Foreground and Background [30]. It then computes an optimal threshold value that minimizes the weighted within class variance; also maximizes the between class variance of these two classes. Otsu's threshold is used in many applications from medical imaging to low level computer vision. Our main aim is to use Otsu's threshold in speech segmentation. Figure-3 shows the step-wise Otsu's algorithm.

4.1 Mathematical Formulation

The mathematical formulation of Otsu's method for computing the optimum threshold is given below:

Let $P(i)$ represents the image histogram of speech spectrogram. The two class probabilities $w_1(t)$ and $w_2(t)$ at level t are computed by:

$$w_1(t) = \sum_{i=1}^{t} P(i) \text{ And } w_2(t) = \sum_{i=t+1}^{I} P(i)$$

The class means, $\mu_1(t)$ and $\mu_2(t)$ are:

$$\mu_1(t) = \sum_{i=1}^{t} \frac{iP(i)}{w_1(t)} \text{ and } \mu_2(t) = \sum_{i=t+1}^{I} \frac{iP(i)}{w_2(t)}$$

Individual class variances:

$$\sigma_1^2(t) = \sum_{i=1}^{t} [i - \mu_1(t)]^2 \frac{P(i)}{w_1(t)} \quad \text{And}$$

$$\sigma_2^2(t) = \sum_{i=t+1}^{I} [i - \mu_2(t)]^2 \frac{P(i)}{w_2(t)}$$

The **within class variance (σ_w)** is defined as a weighted sum of variances of the two classes and given by:

$$\sigma_w^2(t) = w_1(t)\sigma_1^2(t) + w_2(t)\sigma_2^2(t)$$

The **between class variance (σ_b)** is defined as a difference of total variance and within class variance and given by:

$$\sigma_b^2(t) = \sigma^2(t) - \sigma_w^2(t) = w_1(t)w(t)[\mu_1(t) - \mu_2(t)]^2$$

These two variances σ_w and σ_b are calculated for all possible thresholds, $t = 0 \dots I$ (max. intensity). Otsu finds the best threshold that *minimizes the weighted **within class variance** (σ_w), also maximizes the weighted **between class variance** (σ_b).* Finally, the pixel luminance less than or equal to threshold is replaced by 0 (black) and greater than threshold is replaced by 1 (white) to obtain the binary or B/W image.
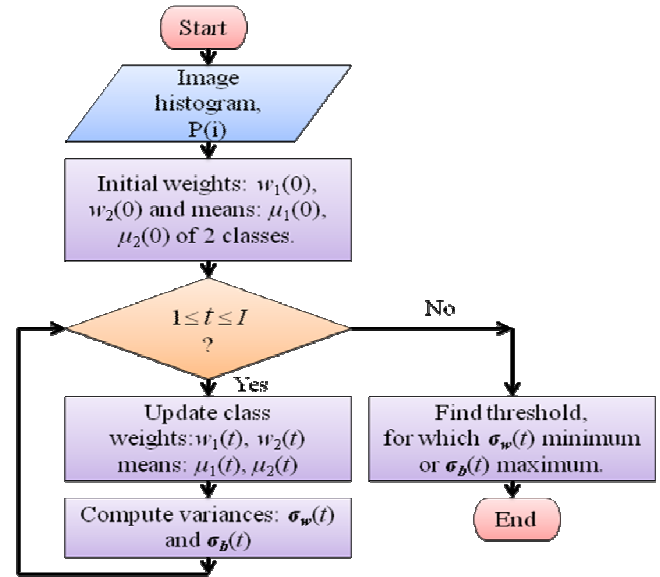
Figure3. Otsu's Thresholding Algorithm – Converting binary image.

5. THE PROPOSED APPROACH

We proposed several clustering methods (*k-means* and *fuzzy c-means* algorithms) and *Otsu's thresholding* methods, with *blocking black area* and *boundary detection* techniques, in continuous Bangla speech segmentation. The proposed segmentation system is shown in Figure-4 and will discuss in the following sub-sections.
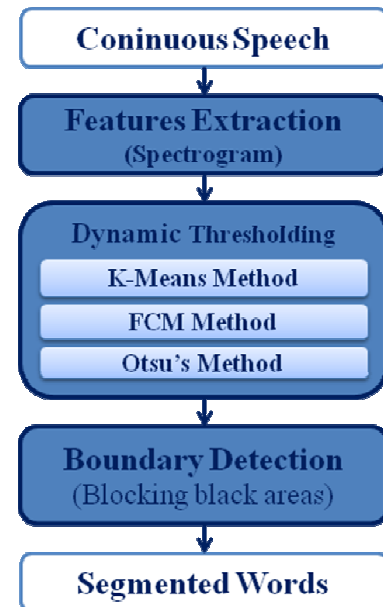
Figure4. Proposed Speech Segmentation System

5.1 Speech Spectrogram

A spectrogram is a time-varying spectral representation of signal. It shows how the spectral density of an audio signal varies over time [31]. Spectrograms can be used to identify spoken words phonetically. They are used extensively in the development of the fields of music, sonar (sound navigation and ranging), radar, and speech processing [32], etc.

The spectrogram is calculated by using the short-time Fourirer Transform (STFT) of audio signal. The STFT of a signal frame $x[n]$ can be expressed by:

$$STFT\{x[n]\}(m,\omega) \equiv X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]\omega[n-m]e^{-j\omega n}$$

Where: $x[n]$ = signal; $\omega[n]$ = window function; m is discrete and ω is continuous.

Thus, the spectrogram of the signal $x[n]$ can be estimated by computing the squares magnitude of the STFT of the signal [33], i.e.,

$$Spectrogram\,(n,\omega) = \left| STFT(n,\omega) \right|^2 .$$

We used MATLAB's 'spectrogram' function that converts speech signal into spectrogram image. The spectrogram of the speech sentence ('আমাদের জাতীয় কবি কাজী নজরুল ইসলাম') is shown in Figure-5.
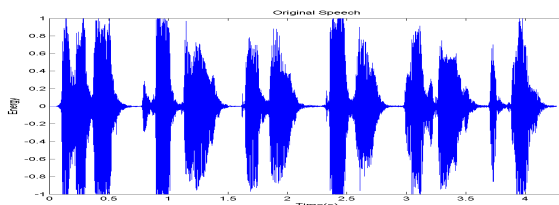
(a) Original Speech Signal
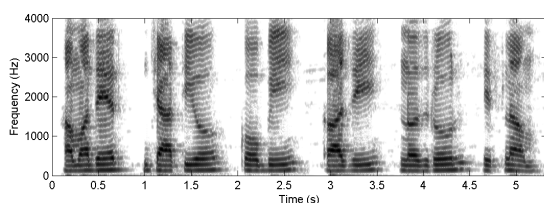
(b) Spectrogram Image

Figure5. Spectrogram of speech signal 'আমাদের জাতীয় কবি কাজী নজরুল ইসলাম'.

5.2 Dynamic Thresholding

First, we used MATLAB's 'rgb2gray' function to convert spectrogram RGB image to gray scale image with pixel values ranging from 0 to 255; where 0 represents a fully black pixel and 255 represents white pixel. Then we used 'im2bw' function with a threshold to convert gray scale image to binary image. But the problem for us is how we choose our threshold. In our research, we proposed three efficient approaches, such as, K-Means, FCM and Otsu's thresholding algorithm to compute the desired threshold, valued between 0-1.

We used MATLAB's 3-class 'kmeans' function and 3-class 'fcm' function. Both functions return the *centroids* of three clusters. The average of 1st and 2nd largest *centroids* is calculated as the desired threshold value. We also used MATLAB's 'graythresh' function that uses Otsu's thresholding method, returns a level or threshold value for which the intra-class variance of the black and white pixels is minimum. The output image replaces all pixels in the input image with luminance greater than or equal to the threshold with the value of 1 (fully white) and less than threshold with 0 (fully black). Figure-6 shows the thresholded spectrogram image of the speech sentence ('আমাদের জাতীয় কবি কাজী নজরুল ইসলাম').

5.3 Blocking Black Area Method

Here, we used a new approach '*Blocking Black Area*' method in the thresholded spectrogram image that produces rectangular black boxes in the voiced regions of the speech sentence, as shown in Figure-7. The method works as follows:

- Summing the column-wise intensity values of thresholded spectrogram image.
- Find the image columns with fewer white pixels based on summing value and replace all pixels on this column with luminance 0 (black).
- Find the image columns with fewer black pixels based on summing value and replace all pixels on this column with luminance 1 (white).
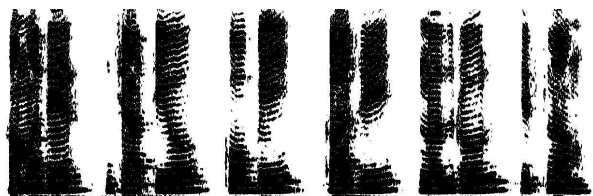

(a) K-Means thresholded Image

(b) FCM thresholded Image

(c) Otsu's thresholded Image

Figure6. Thresholded Spectrogram Images of Speech signal 'আমাদের জাতীয় কবি কাজী নজরুল ইসলাম'

(a) Thresholded Spectrogram Image from Speech Signal

(b) After applying Blocking Black Area Method

Figure7. Effect of applying Blocking Black Area Method – Producing rectangle black boxes in voiced regions of speech signal

5.4 Boundary Detection

We used MATLAB 'regionprops' function to measure the properties of each connected object in the binary image. We also used different shape measurements (such as 'Area', 'BoundingBox', 'Centroid') to identify each rectangular object that represents speech words/sub-words. The 'Extrema' measurement, which is a vector of [top-left top-right right-top right-bottom bottom-right bottom-left left-bottom left-top], is used to detect the start (bottom-left) and end (bottom-right) points of each rectangular object, as shown in Figure-8.
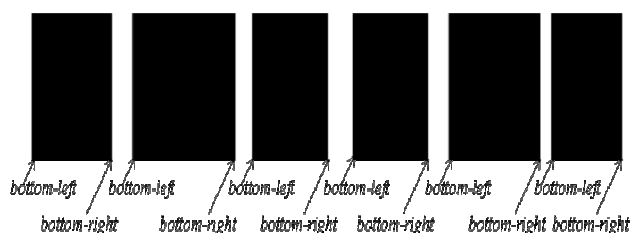
Figure8. Star and End point Detection of rectangular object.

5.5 Final Speech Segments

Each rectangular black box represents a speech word. After detecting the start and end points of each black box, the word boundaries of the original speech sentence are marked automatically by these two points and finally we have cut the word segments from the speech sentence. Figure-9 shows that 6 (six) black boxes represent 6 (six) word segments in the speech sentence '**আমাদের জাতীয় কবি কাজী নজরুল ইসলাম**'.
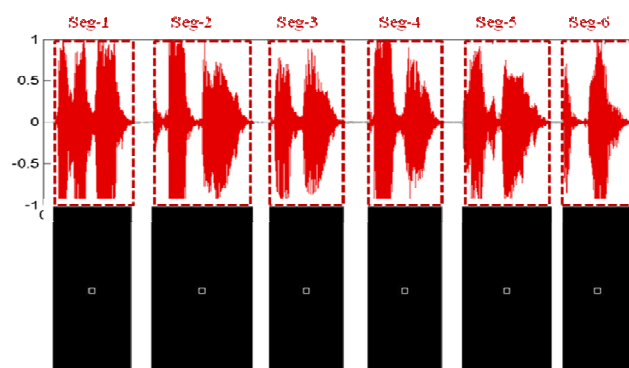
Figure9. 6 (six) detected word segments in the speech sentence 'আমাদের জাতীয় কবি কাজী নজরুল ইসলাম'.

6. EXPERIMENTAL RESULTS

The proposed system has been implemented in Windows environment. All the methods and algorithms, discussed in this paper, have been implemented in MATLAB 7.12.0 (R2011a) version. For the experiments there were compared 3 types of segmentation algorithms: k-means, fuzzy c-means clustering and otsu's thresholding method. To evaluate the performance of the proposed methods, different experiments were carried out. In our experiments, various speech sentences in Bangla language have been recorded, analyzed and segmented by the proposed method. Table-1 shows the details segmentation results for ten Bangla speech sentences and reveals that the average segmentation accuracy rates are 94.11%; and 95.55% for k-means, 96.19% for FCM clustering algorithms, and 90.58% for Otsu's method.

Table1. The details segmentation results

Speaker ID	No. of Sentences	No. of words present	No. of properly segmented words (#) and accuracy (%)						Avg. Accuracy
			K-Means		FCM		OTSU		
			#	%	#	%	#	%	%
S1	100	656	621	94.67	629	95.88	517	78.81	89.79
S2	100	656	621	94.67	622	94.82	601	91.62	93.70
S3	100	656	622	94.82	624	95.12	612	93.29	94.41
S4	100	656	631	96.18	643	98.02	619	94.36	96.19
S5	100	656	639	97.41	637	97.10	622	94.82	96.44
Total	**500**	**3280**	**3134**	**95.55**	**3155**	**96.19**	**2971**	**90.58**	**94.11**

7. CONCLUSIONS AND FURTHER RESEARCH

We have proposed three efficient methods for continuous Bangla speech segmentation: two of them are usually used in pattern classification (i.e., k-means and fuzzy c-means clustering) and one of them is used in image segmentation (i.e., Otsu's thresholding method). We also used new approaches blocking black area and boundary detection methods to properly detect word boundaries. From the experimental results it is concluded that K-Means and FCM clustering methods produce better segmentation results than that of Otsu's Method. It was observed that some of the words were not properly segmented, this is due to different causes: (i) the utterance of words differs depending on their position in the sentence; (ii) the pauses between the words are not identical in all causes; and (iii) the non-uniform articulation of speech. Also, the speech signal is very much sensitive to the properties of speaker and environments. Experimental results showed that the proposed approach is effective in continuous speech segmentation. The major goal of future research is to search a mechanism that can be employed to enable pattern discovery by learning. We have a continuous effort to develop a continuous speech recognition system using neuro-fuzzy system.

REFERENCES

[1]   Okko Rasanen, "Speech Segmentation and Clustering Methods for a New Speech Recognition Architecture", M.Sc Thesis, Dept. of Electrical and Communication Engineering, Helsinki University of Technology, Espoo, November 2007.

[2]   Cherif A., Bouafif L., and Dabbabi T. "Pitch Detection and Formant Analysis of Arabic Speech processing", Applied Acoustics, pp. 1129-1140, 2001.

[3]   Sharma M., and Mammone R., "Subwordbased Text Dependent Speaker Verification System with User-selectable Password", Proceedings of ICASSP, pp. 93-96, 1996.

[4]   Y. Hioka and N. Hamada, "Voice activity detection with array signal processing in the wavelet domain", IEICE Trans. Fund. Elec. Commun. Comput. Sci., vol. 86, no. 11, pp. 2802–2811, 2003.

[5]   F. Beritelli and S. Casale, "Robust voiced/unvoiced speech classification using fuzzy rules", in IEEE Worksh. Speech Cod. Telecommun. Proc., Pocono Manor, pp. 5–6, USA, 1997.

[6]   Y. Qi and B. R. Hunt, "Voiced-unvoiced-silence classifications of speech using hybrid features and a network classifier", IEEE Trans. Speech Audio Proces., vol. 1, no. 2, pp. 250–255, 1993.

[7]   S. Basu, "A linked-HMM model for robust voicing and speech detection", Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'03), Hong Kong, China, Vol. 1, pp. 816–819, 2003.

[8]   Thangarajan R, Natarajan A.M., "Syllable Based Continuous Speech Recognition for Tamil", In South Asian Language Review VOL.XVIII. No.1, 2008.

[9]   Kvale K. "Segmentation and Labeling of Speech", PhD Dissertation, The Norwegian Institute of Technology, 1993.

[10] Md. Mijanur Rahman and Md. Al-Amin Bhuiyan, "Continuous Bangla Speech Segmentation using Short-term Speech Features Extraction Approaches", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 3, No. 11, 2012.

[11] Sai Jayram A K V, Ramasubramanian V and Sreenivas T V, "Robust parameters for automatic segmentation of speech", Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '02), Vol. 1, pp. 513-516, 2002.

[12] A. Webb, "Statistical Pattern Recognition", New Jersey, John Wiley & Sons, 2002.

[13] P. N. Tan, M. Steinbach, V. Kumar, "Introduction to Data Mining", Boston, Addison-Wesley, 2005.

[14] E. Alpaydin, "Introduction to Machine Learning", Cambridge, the MIT Press, 2004.

[15] R.J Hathway, and J.C. Bczdek, "Optimization of Clustering Criteria by Reformulation", IEEE transactions on Fuzzy Systems, pp. 241-245, 1995.

[16] Fu, S.K. Mui, J.K. "A Survey on Image Segmentation", Pattern Recognition, Vol. 13, pp.3–16, 1981.

[17] Haralick, R.M. Shapiro, L.G., "Image Segmentation Techniques", Computer Vision Graphics Image Process. Vol. 29, pp.100–132, 1985.

[18] Pal, N. Pal, S., "A Review on Image Segmentation Techniques", Pattern Recognition, Vol. 26, pp.1277–1294, 1993.

[19] Bezdek, J.C. Hall, L.O. Clarke, L.P. "Review of MR Image Segmentation Techniques Using Pattern Recognition". Med. Phys., Vol. 20, pp.1033–1048, 1993.

[20] Kwon, M. J. Han, Y. J Shin, I.H. Park, H.W. "Hierarchical Fuzzy Segmentation of Brain MR Images", International Journal of Imaging Systems and Technology, Vol. 13, pp.115–125, 2003.

[21] Tolias, Y.A. Panas, S.M., "Image Segmentation by a Fuzzy Clustering Algorithm Using Adaptive Spatially Constrained Membership Functions", IEEE Trans. Systems, Man, Cybernet., Vol. 28, pp. 359–369, 1998.

[22] Pham, D. L. Prince, J. L., "Adaptive Fuzzy Segmentation of Magnetic Resonance Images", IEEE Trans. Medical Imaging, Vol. 18, pp. 737–752, 1999.

[23] MacQueen, J. B., "Some Methods for classification and Analysis of Multivariate Observations", Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. pp. 281–297, 1967.

[24] Lloyd, S. P., "Least square quantization in PCM", Bell Telephone Laboratories Paper, 1957. Published

in journal later: Lloyd, S. P., "Least squares quantization in PCM", IEEE Transactions on Information Theory. 28 (2): 129–137, 1982.

[25] E.W. Forgy, "Cluster analysis of multivariate data: efficiency versus interpretability of classifications", Biometrics 21: 768–769, 1965.

[26] J. C. Bezdek, "Fuzzy Mathematics in Pattern Classification", PhD Thesis, Cornell University, Ithaca, NY, 1973.

[27] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact, Well Separated Clusters", J. Cyber., 3, 32-57, 1974.

[28] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum, NY, 1981.

[29] Gonzalez, Rafael C. & Woods, Richard E, "Thresholding", In Digital Image Processing, pp. 595–611. Pearson Education, 2002.

[30] Nobuyuki Otsu, "A threshold selection method from gray-level histograms", IEEE Trans. Sys., Man., Cyber. **9** (1): 62–66, 1979.

[31] Haykin, S., "Advances in Spectrum Analysis and Array Processing", Vol.1, Prentice Hall, 1991.

[32] JL Flanagan, "Speech Analysis, Synthesis and Perception", Springer- Verlag, New York, 1972.

[33] National Instruments, "STFT Spectrogram VI": http://zone.ni.com/reference/en-XX/help/371361E-01/lvanls/stft_spectrogram_core/#details

BIOGRAPHIES

Mr. Md. Mijanur Rahman is working as Assistant Professor of the Dept. of Computer Science and Engineering at Jatiya Kabi Kazi Nazrul Islam University, Trishal, Mymensingh, Bangladesh. Mr. Rahman obtained his B. Sc. (Hons) and M. Sc degrees, both with first class first in Computer Science and Engineering from Islamic University, Kushtia, Bangladesh. Now he is continuing his PhD program in the department of Computer Science and Engineering at Jahangirnagar University, Savar, Dhaka, Bangladesh.

His teaching and research interest lies in the areas such as Digital Speech Processing, Pattern Recognition, Neural Networks, Fuzzy Logic, Artificial Intelligence, etc. He has got many research articles published in both national and international journals.

Dr. Md. Al-Amin Bhuiyan is serving as a professor of the Dept. of Computer Science and Engineering at Jahangirnagar University, Savar, Dhaka, Bangladesh. Dr. Bhuiyan obtained his B. Sc. (Hons) and M. Sc degrees both with first class in Applied Physics & Electronics from the University of Dhaka, Bangladesh. He obtained his PhD degree in Information & Communication Engineering from Osaka City University, Japan.

His teaching and research interest lies in the areas such as Image Processing, Computer Vision, Computer Graphics, Pattern Recognition, Soft Computing, Artificial Intelligence, Robotics, etc. He has got many articles published in both national and international journals.