ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

Automatic Speech Recognition using ELM and KNN Classifiers

M.Kalamani[1], Dr.S.Valarmathy[2], S.Anitha [3]

Assistant Professor (Sr.G), Dept of ECE, Bannari Amman Institute of Technology, Sathyamangalam, India [1]

Professor and Head, Dept of ECE, Bannari Amman Institute of Technology, Sathyamangalam, India [2]

PG Student, Dept of ECE, Bannari Amman Institute of Technology, Sathyamangalam, India[3]

ABSTRACT: Automatic speech recognition system consist of two stages: One is Pre-processing stage and another one is classification stage. In pre processing stage continuous speech signal is recorded and segmented. The classification stage is used to classify the extracted features. The segmentation algorithm is hybrid of short time energy and spectral centroid. It has high segmentation accuracy. The Hit Rate rate is 95.33% and False Alarm rate is 4.67%. In this paper MFCC is used for feature extraction and ELM, KNN classifiers are used for speech classification. Compare to KNN classifier ELM classifier has high classification accuracy.

KEYWORDS: Speech segmentation, Spectral Centroid, Speech Classification, KNN, ELM

I. INTRODUCTION

Automatic speech recognition is used to convert a speech signal into text signal accurately and efficiently. A speaker-independent system does not use training data. The speaker-dependent systems use training data. Segmentation is used to identify the boundaries of words, syllables, or phonemes. The advantages of speech segmentation is to reduce the computational load and power consumption of the system [1]

Automatic speech recognition can be divided into three different components such as signal preprocessing, feature extraction and signal classification. In pre processing stage noise can be eliminated. In feature extraction most discriminative features can be extracted that is used to characterize a speech signals. In this paper Mel frequency cepstral coefficient method is used. Classification is used to classify the extracted features and relates the input sound to the best fitting sound in a known vocabulary set [2].

In all classification methods, the data is separated into training and test sets. Each instance in the training set contains a target value which represents the corresponding class and a set of attributes. The test data do not contain a target value. The objective of the classifier is to produce a model from the training data which predicts the target values of the test data [3].

II. RELATED WORK

The time domain features such as short time energy (STE) and zero crossing rate (ZCR). The frequency domain features such as spectral centroid (SC) and spectral flux (SF). The segmentation methods are described as follows.

Md. Mijanur Rahman and Md. Al-Amin Bhuiyan (2012) proposed the Speech Segmentation method using Short-term Speech Features Extraction. Continuous Bangla speech sentences segmented using time domain features and frequency domain features. The time-domain features, such as short-time signal energy, short-time average zero crossing rate and the frequency-domain features, such as spectral centroid and spectral flux. A simple dynamic thresholding criterion is applied in order to detect the word boundaries. J.Sangeetha and S.Jothilakshmi (2012) proposed the Continuous Speech Segmentation for Indian Language. Convert speech into corresponding text, it is necessary to identify the boundaries and phrases present in the continuous speech signal. Automatic continuous speech segmentation for Indian languages using short time energy and zero crossing rate. The beginning and ending for each utterance can be detected. Hemakumar G and Punitha P (2014) proposed the Segmentation of Kannada Speech Signal.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

Automatically segments the continuous Kannada speech signal into syllables and sub-words using the dynamic threshold computation by the combination of short time energy and magnitude of signal. In pre processing hamming window is used. Md. Mijanur Rahman et al. (2010) proposed the Segmentation and Clustering of Continuous Bangla Speech. The segmentation approach was used to segment the continuous speech into uniquely identifiable and meaningful units. After segmentation, the segmented words were clustered into different clusters according to the number of syllables and the sizes of the segmented words. Nipa Chowdhury et al. (2010) proposed the Separating Words from Continuous Bangla Speech Continuous Bangla speeches are fed into the system and the word separation algorithm separate speech into isolate words**.** The algorithm is developed by considering prosodic feature with energy.

This paper is organized as follows: Section 2 describes techniques for segmentation of the speech signal. Section 3 describes segments detection of speech signal. Section 4 describes the hybrid speech segmentation. In section 5 describes the speech classification. In section 6 describes the performance measures. Section 7 and 8 describes the results and conclusion.

III. HYBRD SPEECH SEGMENTATION ALGORITHMS

The hybrid speech segmentation algorithms are spectral centroid and short time energy.

A. Short Time Energy [5]

The energy signal is time varying signal. It is a measure of how much signal there is at any one time. By the nature of production, the speech signal consist of voiced, unvoiced and silence regions [4].The hamming window is used to calculate the short time energy [4].

The equation of the short time energy is [5]

$$En = \frac{1}{N}\sum_{m=1}^{N}[x(m)w(n-m)]^2 \tag{1}$$

where, x (m) is a discrete-time audio signal and w (m) is a rectangle window

The equation of hamming window is [5]

$$w(n) = \alpha - \beta\cos(\frac{2\Pi n}{N-1}) \tag{2}$$

where ,
α=0.54, β=1-α=0.46.

B. Spectral Centroid[5]

Spectral centroid indicates where the "center of gravity" of the spectrum is [4]. This feature is a measure of the spectral position, with high values corresponding to “brighter” sounds [5].The equation of Spectral centroid is defined as [5]

$$SC_i = \frac{\sum_{m=0}^{N-1} f(m)X_i(m)}{\sum_{m=0}^{N-1} X_i(m)} \tag{3}$$

f (m) is a Center frequency, X_i (m) is a amplitude of the signal.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

The DFT is given by [5]

$$X_k = \sum_{n=0}^{N-1} x(n) e^{-j2\Pi k \frac{n}{N}}, \mathrm{k} = 0 \ldots \mathrm{N}-1 \quad (4)$$

IV. SPEECH SEGMENTS DETECTION

A simple dynamic based threshold method is used to detect the speech segments. The following steps are present in these thresholding methods [5].

1. Get the feature sequence from the previous feature extraction module.
2. Apply median filtering to smooth the feature sequences.
3. Compute the Mean or average values of these sequences.
4. Find the threshold value.[5]

Threshold [5]

$$T = \frac{Mean}{2} \quad (5)$$

Here, the both short time energy and spectral centroid the above steps are applied to find the threshold value [5].

The two threshold values are T1 and T2. T1 is threshold value for energy and T2 is threshold value for spectral centroid. Based on these two threshold values speech segment is detected [5].

V. HYBRID SPEECH SEGMENTATION

The hybrid speech segmentation is the combination of short time energy and spectral centroid. The hybrid speech segmentation method has five major steps [6, 5].

1. Speech Acquisition
2. Signal Preprocessing
3. Speech Segmentation
4. Dynamic Thresholding.
5. Speech Segments Detection [5]

1) Speech Acquisition

Speech acquisition is acquiring of continuous speech sentence s through the microphone [5].

2) Signal Preprocessing

Preprocessing is elimination of back ground noise, framing and windowing. Back ground noise is removed from the data. Continuous speech has been separated into frames. That method is known as framing. Windowing is used to determine the portion of the speech signal [5].

3) Speech Segmentation

In this section we have been computed the hybrid of short time energy and spectral centroid of each frame of the speech signal [5].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

4) Dynamic Thresholding

This method is used to find the threshold values. The two threshold values are T1 and T2. After computing two thresholds, the speech word segments are formed by successive frames for which the respective feature values are larger than the computed threshold values [5].

5) Speech Segments Detection

A simple dynamic based threshold method is used to detect the speech segments [5].

BLOCK DIAGRAM OF AUTOMATIC SPEECH RECOGNITION

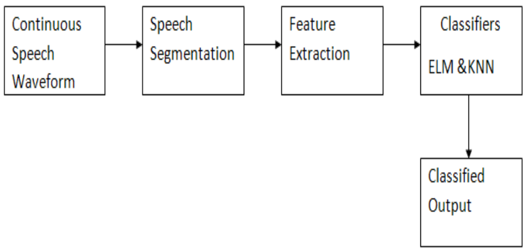

Figure 1. Block Diagram of Automatic speech Recognition

In figure 1 shows the block diagram of automatic speech recognition. In speech segmentation block hybrid speech segmentation algorithm is used to segment the continuous speech waveform. MFCC method is used for feature extraction. In classification block ELM and KNN classifiers are used.

VI. SPEECH CLASSIFICATION

Speech Recognition is a special case of pattern recognition. There are two types of phases: Training and Testing. Classification is common in both phases [7]. The test pattern is declared to belong to that whose model matches the test pattern best [7]. In training phase, the parameters of the classification model are estimated using the training data. In testing phase, test speech data is matched with the trained model of each and every class.

A. K-Nearest neighbor (KNN):

KNN classifier is a type of instance based learning technique and predicts the class of a new test data based on the closest training examples in the feature space [8]. Euclidean distance was used as distance measurement [9]. The KNN algorithm is among the simplest of all machine learning algorithms. Both for classification and regression, it can be useful to weight the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. KNN is a variable-bandwidth, kernel density estimator with a uniform kernel. Using an appropriate nearest neighbor search algorithm makes KNN computationally tractable even for large data sets.

B. Extreme learning machine (ELM):

Extreme Learning Machine (ELM) is used to study the automatic speech recognition and it’s also used for speech emotion recognition[10].The weights between the input neurons and the hidden neurons in ELM were randomly assigned based on some continuous probability density function while the weights between the hidden layer and the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

output of the probability density function while the weights between the hidden layer and the output of the single layer feed forward network was determined analytically in [11,12].

VII. PERFORMANCE MEASURES

The performance measures of the speech signal is defined as follows: [5]
Hit rate is defined as number of correctly recognized words. The equation of Hit rate is given by [5].

$$\text{Hit Rate} = \frac{\text{No. of correctly identified word}}{\text{Total no. of words}}$$

False Alarm rate is defined as number of words incorrectly recognized. The equation of false alarm rate is given by [5].

$$\text{False Alarm Rate} = \frac{\text{No. of erroneous word identified}}{\text{Total no. of words}}$$

VIII. RESULTS AND DISCUSSION

The hybrid speech segmentation algorithm has been implemented in Mat lab [5].Various human speech sentences in Tamil language have been recorded and segmented. Hybrid speech segmentation algorithm has been implemented and analyzed. The performance of speech recognition system is often described in terms of accuracy [5].

Table 1. Results for Hit Rate and False Alarm Rate

Si.no	Sentences	Total no. of words	SC	STE	SF	STE&SC	Hit Rate of STE&SC (%)	False Alarm Rate of STE&SC (%)
1.	Tamil 1	3	3	2	1	3	100	0
2.	Tamil 2	5	4	5	1	4	80	20
3.	Tamil 3	5	5	5	1	5	100	0
4.	Tamil 4	6	5	5	1	5	83.33	16.67
5.	Tamil 5	3	2	3	1	3	100	0
6.	Tamil 6	5	5	1	1	5	100	0
7.	Tamil 7	5	5	5	1	5	100	0
8.	Tamil 8	3	2	3	3	3	100	0
9.	Tamil 9	5	5	5	1	5	100	0
10.	Tamil 10	4	3	3	1	4	100	0
11.	Tamil 11	3	3	1	1	2	66.67	33.33
12.	Tamil 12	3	2	1	1	3	100	0
13.	Tamil 13	4	4	2	1	4	100	0
14.	Tamil 14	3	2	2	1	3	100	0
15.	Tamil 15	4	3	3	1	4	100	0
	Average of Hit Rate and False Alarm Rate						**95.33**	**4.67**

In table.1 shows the details Hit Rate and False Alarm Rate results for hybrid speech segmentation. The existing method is SF, SC and STE. The hybrid method is combination of STE and SC [5].

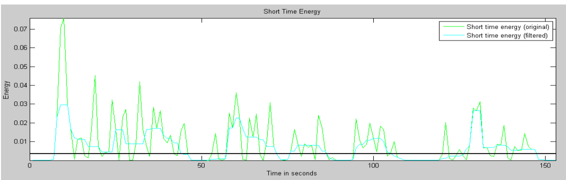

Figure2. Original and filtered signal of short time energy.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

In figures 2 shows the details of the original speech signal of and short time energy and how it will be after the pre processed signal. This pre processed stage will makes the signal in standard format which leads at increasing the segmentation accuracy rate.

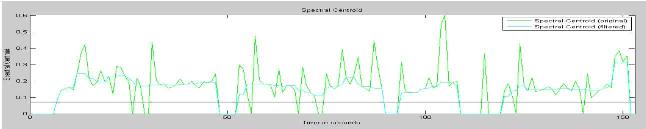

Figure3. Original and filtered signal of spectral Centroid

In figures 3 shows the details of the original speech signal of and spectral centroid and how it will be after the preprocessed signal [5]. This preprocessed stage will makes the signal in standard format which leads at increasing the segmentation accuracy rate. In filtered output DC component is removed and it gives standardized signal [5]

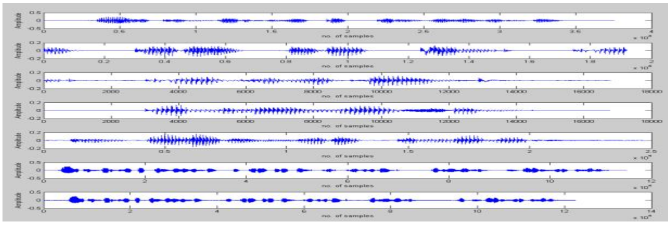

Figure4. Time Domain results for short time energy and spectral centroid.

In this figure4. Indicates the time domain results of short time energy and spectral centroid. It shows the segmented output of the input signal [5].

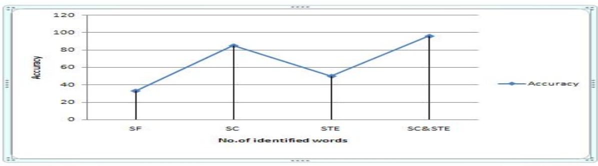

Figure5. Comparisons of Segmentation Algorithms

The line chart gives the comparison of various segmentation methods. Accuracy of four segmentation methods is compared. It shows that the hybrid of short time energy and spectral centroid has high segmentation accuracy [5].

Table 2. Results for ELM classifier

Class	Accuracy	Computation time for training	Computation time for testing
2	100	0.6984	1.0609
3	99.98	0.2375	0.5343
4	99.96	0.1726	0.1921
5	97.65	0.1362	0.1875
6	95.45	0.0940	0.1583

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

7	90.04	0.0712	0.1183
8	85.32	0.0707	0.1121
9	82.36	0.0371	0.0534
10	75.25	0.0498	0.0981

In table 2 gives the results of Accuracy and Average time of training and testing for ELM classifier. Number of classes increases the classification accuracy will be decreased.

Table 3 Results for KNN classifier

Class	Classification Accuracy	Missed classification Accuracy	Computation time for testing
2	100	0	4.0057
3	93.33	6.67	5.0159
4	75	25	1.7219
5	72	28	1.0406
6	63.33	36.67	0.9835
7	62.85	37.14	0.8021
8	62.50	42.5	1.3866
9	62.22	37.78	0.7934
10	60	40	1.0781

In table 3 gives the results of Accuracy and Average time of testing for KNN classifier. Number of classes increases the classification accuracy will be decreased.

IX. CONCLUSION

In this paper, hybrid speech segmentation algorithms and ELM, KNN classifiers are discussed and comparisons are made between various segmentation algorithms. The Hit Rate and False Alarm Rates of hybrid speech segmentation are calculated. The hybrid method gives the good accuracy in speech segmentation. This method increases the accuracy rate and decreases the error rate. The Hit Rate rate is 95.33% and False Alarm rate is 4.67%. Compare to KNN classifier ELM classifier has high classification accuracy.

ACKNOWLEDGEMENT

The authors would like to thank friends, reviewers and Editorial staff for their help during preparation of this paper.

REFERENCES

1. J. Sangeetha and S. Jothilakshmi " Robust Automatic Continuous Speech Segmentation for Indian Languages to Improve Speech to Speech Translation" International Journal of Computer Applications (0975 – 8887) Volume 53– No.15, September 2012.
2. Georgi T. Tsenov, and Valeri M. Mladenov, "Speech Recognition Using Neural Networks", 10th symposium on neural network applications in electrical engineering, September 2010.
3. Sonia Suuny, David Peter S, K. Poulose Jacob, "performance of different classifiers in speech recognition" IJRET, Volume: 2 Issue: 4, APR 2013.
4. M.Kalamani, Dr.S.Valarmathy, S.Anitha, R.Mohan, "Review of Speech Segmentation Algorithms for Speech Recognition", International Journal of Advanced Research in Electronics and Communication Engineering, Volume 3, Issue 11, November 2014.

5. M.Kalamani, Dr.S.Valarmathy, S.Anitha, "Modified Speech Segmentation Algorithm for Continuous Speech Recognition" international journal of advanced research trends in engineering and technology (ijartet) vol. ii, special issue viii, February 2015
6. Bello J P, Daudet L, Abdallah S, Duxbury C, Davies M, and Sandler MB, "A Tutorial on Onset Detection in Music Signals", IEEE Transactions on Speech and Audio Processing 13(5), pp 1035–1047,2005.
7. Santosh K.Gaikwad, Bharti W.Gawali, Pravin Yannawar, "A Review on Speech Recognition Technique," International Journal of Computer Applications (0975 – 8887) Volume 10– No.3, November 2010
8. R. O. Duda, P. E. Hart and D. G. 2012. Stork, Pattern classification. John Wiley and Sons
9. M. Hariharan, Sazali Yaacob, M. N. Hasrul and Oung Qi Wei "speech emotion recognition using stationary wavelet transform and timbral texture features", ARPN Journal of Engineering and Applied Sciences vol. 9, no. 8, august 2014
10. Nidhi Desai, Prof.Kinnal Dhameliya, Prof.Vijayendra Desai, "Feature Extraction and Classification Techniques for Speech Recognition: A Review", international journal of Emerging Technology and Advanced Engineering Volume 3, Issue 12, December 2013
11. G.-B. Huang, H. Zhou, X. Ding and R. Zhang. 2012. Extreme learning machine for regression and multiclass classification. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on. 42: 513-529.
12. G.-B. Huang, Q.-Y. Zhu and C.-K. Siew. 2006. Extreme learning machine: theory and applications. Neurocomputing. 70: 489-501