I J C T A, 9(2) 2016, pp. 575-590

Environmental Audio Source Separation Based on Improved K-means Clustering with Adaptive Genetic Algorithm

T. Sivaprakasam[1] and P. Dhanalakshmi[2]

ABSTRACT

In daily life environmental sounds that are present around us include intricately mixed sounds emitted from different sources. This work aim is to contribute a method for sound source separation by means of the enhanced *k*-means clustering with adaptive genetic algorithm. At first removes the features from the input audio signal by means of Mel Frequency Cepstral Coefficients (MFCC) and spectral features. Based on the features we gather the input audio signal to background sound and event sound. The suggested method employs the enhanced *k*-means with adaptive genetic algorithm for collecting the audio signal. Now the traditional *k*-means clustering algorithm is enhanced using the centroids selection by the adaptive genetic algorithm. At last we categorize the event sound by means of the Neuro fuzzy classifier. The Neuro fuzzy classifier is employed to categorize the event sound based on the features. So, that the suggested method attains the better classification with high accuracy.

Keywords: Mel-frequency cepstral coefficients (MFCCs), Spectral Centroid, Spectral Roll-off, Spectral Decrease, Spectral Flux, Spectral Slope, Improved k-means clustering, Adaptive genetic algorithm, Neuro fuzzy classifier.

1. INTRODUCTION

Humans live in a complex audio environment, and have very good skills of following a specific sound source while ignoring or simply acknowledging the others. For example we can follow a conversation in a busy background consisting of other people talking or music. The performance of automatic methods in computational auditory scene analysis (CASA) is much more limited in this task. Acoustic mixture signals contain multiple simultaneously occurring sound events, and machine listening systems are still far from the level of human performance in recognizing them [1]. Source separation of music audio signals is a fundamental task for music information retrieval (MIR). High-quality source separation could help users find their favorite songs according to the content (such as vocals or instruments). It would also let them enjoy active music listening based on the remixing of existing instrumental parts. Nonnegative matrix factorization (NMF) has recently played a key role in the source separation of single-channel audio signals. It can approximate a nonnegative matrix (the amplitude or power spectrogram of a given mixture signal) as the product of two nonnegative matrices— a set of basis spectra and a set of the corresponding activations [3]. In the blind source separation problem the objective is to separate multiple sources, mixed through an unknown mixing system (channel), using only the system output data (observed signals) and in particular without using any (or least amount of) information about the sources or the system. The blind source separation problem arises in many fields of studies, including speech processing, data communication, biomedical signal processing, etc [2].

Multidimensional data analysis plays a major role in solving BSS problems. Proper representation of multivariate data commonly encountered in signal processing, pattern recognition, neural networks and

[1] Assistant Professor, Department of Computer science and Engineering, Annamalai University, India
[2] Associate Professor, Department of Computer science and Engineering, Annamalai University, India
[1]*E-Mail: tsivaprakasam@gmail.com*; [2]*E-Mail: abidhan01@gmail.com*

statistical analysis is essential for visualization of the underlying geometry of the data structure [4, 5]. Linear transformations are used to exploit possible dependencies and reduce dimensionality of the multivariate data sets. Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are the two prominent methods for linear transformation. Particularly ICA has been very effective in separating independent sources from the mixed signals. Hence, the principles of ICA are essential to deal with the issues of BSS [6].

There are multiple potential applications of convolutive blind source separation. In acoustics different sound sources are recorded simultaneously with possibly multiple microphones. These sources may be speech or music, or underwater signals recorded in passive sonar. In radio communications, antenna arrays receive mixtures of different communication signals. Source separation has also been applied to astronomical data or satellite images. Finally, convolutive models have been used to interpret functional brain imaging data and bio-potentials [7].

2. RELATED WORK

Various environmental sounds exist around us in daily life. Recently, environmental sound recognition has drawn great attention for understanding the environment. However, because environmental sounds derive from multiple sound sources, it was difficult to recognize them accurately [9]. If they were able to separate sound sources before sound recognition as a pre-process, then recognition would be easier and more accurate. They assumed that monaural microphones are widely installed in mobile devices used as recording devices. Two-phase clustering using non-negative matrix factorization (NMF) was proposed to separate monaural sound sources. In this proposal, the time-variant gain feature was used as an attribute of an environmental sound for more efficient sound separation [10, 11].

Multichannel signal processing using a microphone array provides fundamental functions for coping with multi-source situations, such as sound source localization and separation that are needed to extract the auditory information for each source [12]. Auditory uncertainties about the degree of reverberation and the number of sources are known to degrade performance or limit the practical application of microphone array processing. Such uncertainties must therefore be overcome to realize general and robust microphone array processing. These uncertainty issues have been partly addressed-existing methods focus on either source number uncertainty or the reverberation issue, where joint separation and dereverberation has been achieved only for the over determined conditions [13, 14]. All-round method that achieves source separation and dereverberation for an arbitrary number of sources including underdetermined conditions. Our method uses Bayesian nonparametric that realize an infinitely extensible modeling flexibility so as to bypass the model selection in the separation and dereverberation problem, which was caused by the source number uncertainty. Evaluation using a dereverberation and separation task with various numbers of sources including underdetermined conditions demonstrates that (1) our method is applicable to the separation and dereverberation of underdetermined mixtures, and that (2) the source extraction performance was comparable to that of a state-of-the-art method suitable only for over determined conditions [15].

Sound source separation from a multichannel microphone array capture via estimation of source spatial covariance matrix (SCM) of a short-time Fourier transformed mixture signal. In many conventional audio separation algorithms the source mixing parameter estimation was done separately for each frequency thus making them prone to errors and leading to suboptimal source estimates [16].

3. OUTLINE OF THE WORK

In real-world audio signals several sound sources are usually mixed. The process in which individual sources are estimated from the mixture signal is called sound source separation. The separation of percussive and harmonic sounds remains a challenging problem in sound source separation. The common problem in existing software estimation method is given below,

- In real time sound source separation, isolating speech from environmental noise remains a challenging problem, especially in the presence of highly non-stationary noise.
- One of the major problems in the existing method is accuracy. The accuracy of the sound source will be affected through the separation of background sound from the event sound.
- The problem of real time sound source separation is overlapping of sound by employing non negative matrix factorization techniques.

The important purpose of the suggested method is to sound source separation based on enhanced k means clustering algorithm with neuro fuzzy classification. Initially we coach the input sample audio signal and next extract the features from the sample signal by means of the Mel frequency Cepstral coefficients (MFCC) and spectral features. After that we collect the features based on the enhanced k means clustering algorithm with Adaptive genetic algorithm. Based on this we cluster the audio signal into background and multiple event sound. At last the suggested method classifies the multiple event sound by means of the neuro fuzzy classifier. The detail process of the suggested method is revealed in the block diagram Fig. 1.

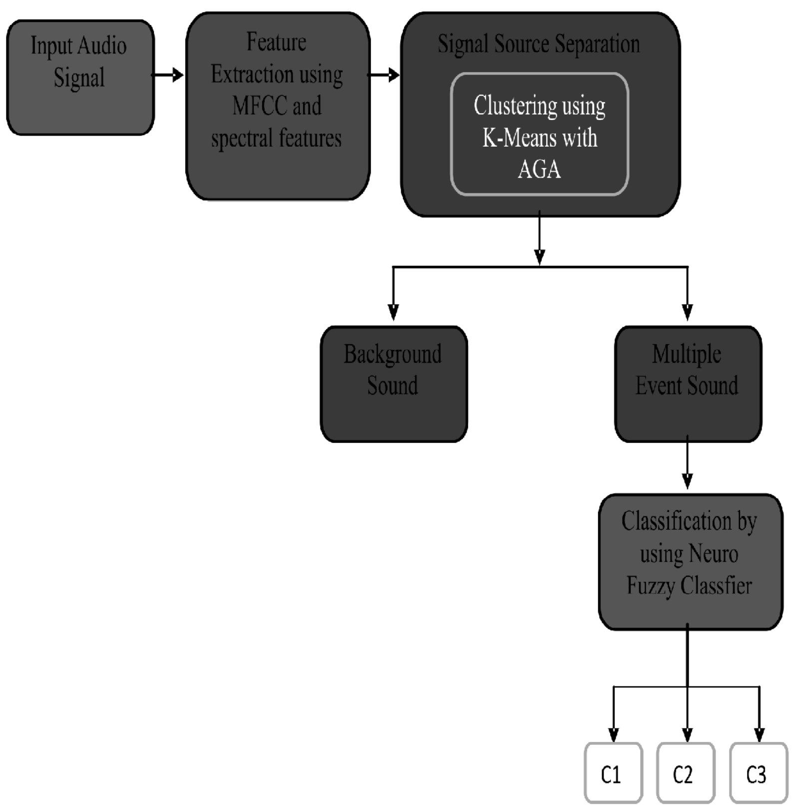

Figure 1: The block diagram of Sound Source Separation

3.1. Acoustic features

The purpose of feature extraction is to extract useful discriminative information from the waveform which will result in a compact set of feature vectors [11]. We have experimented with many different types of feature for classification of auditory events. Environmental sounds in general are unstructured data comprising of contributions from a variety of sources, and unlike music or speech, no assumptions can be made about predictable repetitions nor harmonic structure in the signal. Due to the inherent diverse nature, there are many features that can be used, or are needed, to describe audio signals.

Mel Frequency Cepstral Coefficients (MFCCs)

Mel Frequency Cepstral Coefficients (MFCCs) are obtained through a frame based analysis of a signal where the waveform is divided into a sequence of frames, the purpose is to smooth the frequency spectra and reduce the effects of acoustic variation. A sinusoidal transform(DFT) is performed using a hamming window overlapping each frame to obtain an amplitude spectrum, which is then converted to a Mel-scale spectrum using triangular filters, emphasising frequencies according to their perceptual importance on this scale.

$$M(f) = 2595 \log\left(1 + \frac{f}{700}\right) \tag{1}$$

where *M(f)* is the logarithmic scale of *f* normal frequency scale.

Spectral Centroid

The spectral centroid is a measure used in digital signal processing to characterize a spectrum. It indicates where the “center of mass” of the spectrum is. Perceptually, it has a robust connection with the impression of “brightness” of a sound. It is calculated as the weighted mean of the frequencies present in the signal, which is determined using a Fourier transform, with their magnitudes as the weights

$$Centroid = \frac{\Sigma_{n=0}^{N-1} f(n)x(n)}{\Sigma_{n=0}^{N-1} x(n)} \tag{2}$$

where $x(n)$ represents the weighted frequency value, or magnitude, of bin number n, and $f(n)$ represents the center frequency of that bin.

Spectral Roll-Off

Spectral Roll-off is the frequency R_s so that α^{th} percentile of the power spectral distribution is below this frequency. It is correlated to harmonic/noise cutting frequency.

$$\sum_{k=1}^{R_n} x_n[k] = \propto \sum_{k=1}^{K} x_n[k] \tag{3}$$

where α is the percentage (i.e. $\alpha = 0.95$) and $X_n[k]$ denotes the k^{th} component of the power spectrum vector for the n^{th} signal frame.

Spectral Decrease

Spectral decrease represents the amount of decreasing of the spectral amplitude. This formulation comes from perceptual studies and it is supposed to be more correlated to human perception. The formula is:

$$Decrease = \frac{1}{\Sigma_{n=1}^{N-1} x(n)} \cdot \sum_{n=1}^{N-1} \frac{x(n) - x(0)}{N-1} \tag{4}$$

where $x(n)$ represents the weighted frequency value or magnitude of bin number n.

Spectral Flux

Spectral flux is calculated by first calculating the difference between the current values of each magnitude spectrum bin in the current window from the corresponding value of the magnitude spectrum of the previous window. Each of these differences is then squared, and the result is the sum of the squares.

$$SF = \frac{1}{(N-1)(k-1)} \sum_{n=1}^{N-1} \sum_{k=0}^{k-1} [\log A(n,k) - \log A(n-1,k)]^2 \tag{5}$$

where $A(n, k)$ is the discrete Fourier transform of the nth frame of input signal.

$$A(n,k) = \left| \sum_{m=-\infty}^{\infty} x(m) w(nL-m) e^{j\frac{2\Pi}{L}km} \right| \tag{6}$$

$x(m)$ is the original audio data, $w(m)$ is the window function, L is the window length, K is the order of discrete fourier transform (DFT), and N is the total number of frames.

Spectral Slope

The spectral slope represents the amount of spectral energy decrease as a function of frequency. It assumes that the amplitude spectrum follows a linear model:

$$A(k) = mk + b \tag{7}$$

The slope m is computed by linear regression.

$$m = \frac{\frac{K}{2} \Sigma_{k=0}^{\frac{K}{2}-1} kA(k) - \Sigma_{k=0}^{\frac{K}{2}-1} k \, \Sigma_{k=0}^{\frac{K}{2}-1} A(k)}{\frac{K}{2} \Sigma_{k=0}^{\frac{K}{2}-1} k^2 - \left(\Sigma_{k=0}^{\frac{K}{2}-1} \right)} \tag{8}$$

In this equation (7), (8) where K is the total number of frequency values, $A(k)$ is the spectral amplitude with frequency index k.

3.2. Improved k-means clustering algorithm

In our suggested method, we employ the enhanced k means clustering algorithm is employed to divide the background sound and multiple event sound. Now the traditional k means clustering algorithm is enhanced using the centroids value selection. To choose the centroids value, our suggested method employs the adaptive genetic algorithm. Initially we find out the number of cluster K we choose the centroids value arbitrarily. Next the k-means clustering algorithm has three steps, it is specified beneath.

- Find out the centroids coordinate by means of AGA.
- Find out the distance of each object
- Set the object

Step 1: Initialization

Initialize the solution with the characteristics or features (X, Y) then select the cluster K. Our objective is to set these objects into k groups. The proposed method K values are selected based on the gap statistic method.

Gap statistics

The Gap statistic is a standard method for determining the number of clusters in a set of data. At first assumes that a set of sample (xi) then the resultant clusters are obtained. C_1, C_2...C_K. then calculate the pair wise distance $d^2(x_i,x_{i'})$ of each cluster m. And then find the sum of within the cluster Sk. it is given in the following equation.

$$S_k = \sum_{m=1}^{k} \frac{1}{2n_m} \sum_{i,i' \in C_m} d^2(x_i, x_{i'}) \tag{9}$$

In gap statistic S_k value is compared to the $\log(S_k)$.

$$Gap_n(k) = E_n^* \{\log(S_k)\} - \log(S_k) \tag{10}$$

Where, E_n^* is the expectation of sample size n of the reference distribution. Based on these gap statistics we assign the k value of the proposed method.

Step 2: Find distance

Compute the distance between cluster centroids to each object. Now we employ the Euclidean distance to locate the object centroids distance.

$$Euclidean\, distance = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \tag{11}$$

Step 3: Objects clustering

The squared Euclidean distance from each object to each cluster is calculated, and each object is allocated to the closest cluster.

Step 4: Iteration 1: Find centroids

Making out the members of each group, now we calculate the novel centroids of each group based on these novel memberships. For computing the centroids of the suggested method employ the adaptive genetic algorithm.

Adaptive genetic algorithm

Adaptive genetic algorithm is employed to find the centroids value in our suggested method. Initially we initialize the centroids value and next locate the fitness by means of the symmetric index.

$$fitness = max.symmetric\ index \tag{12}$$

The best fitness value is employed as the centroids value of the suggested method.

Step 5: Iteration-1: Find distance

The next step is to calculate the distance of all objects to the novel centroids.

Step 6: Iteration-1: Objects clustering

Similar to step 3, we allocate each object based on the minimum distance.

Step 7: Stopping criteria

Steps 4, 5 and 6 are replicated until no object moves clusters.

3.3. Classification using Neuro fuzzy classifier

By means of neuro fuzzy classifier the multiple event sounds are classified. The removed features are specified as the input to the neuro fuzzy classifier for categorizing all the multiple event sound to their class. The neuro fuzzy system has a three layered architectural plan. Fig. 2 illustrates the fundamental structure of the neuro fuzzy classifier. Neuro-Fuzzy classifier is a fuzzy based system that is coached by a learning algorithm obtained from Neural Networks. The learning algorithm only executes on the local information and offers the local modifications in the fuzzy system. In common, a neuro-fuzzy system produces very dominant solutions instead of employing the system components individually. In Fig. 2 the specified process is revealed.

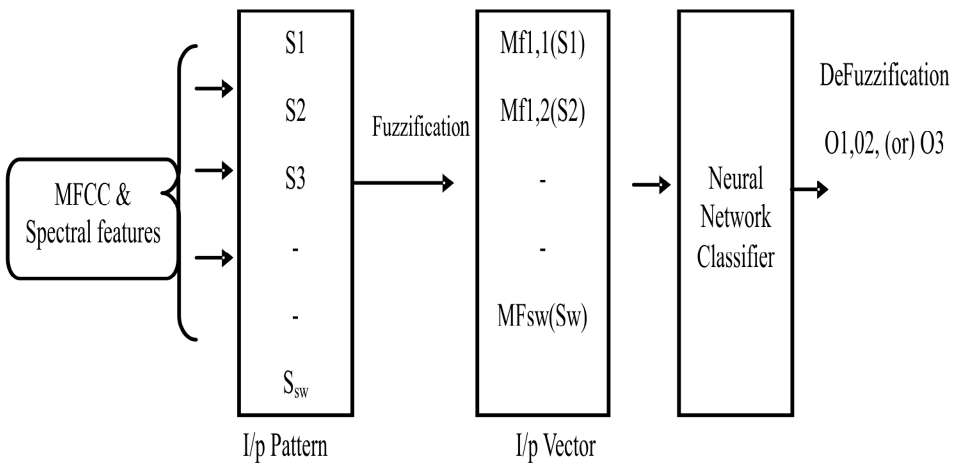

Figure 2: The detailed process of neuro fuzzy classifier

Fuzzification

The input values are the removed features from the multiple event sound. Now Mel Frequency Cepstral Coefficients (MFCC) and spectral features is employed to extract the features. These input features values are fuzzified by means of membership functions that make easy the membership of each feature to dissimilar classes. The unseen and inter-related information are removed from the features to the classes through the MF, which leads to get more accuracy of the classification phase by means of Neuro-fuzzy classifier. The membership matrix includes with sum of the window size and three columns, in which the number of rows is the same to the number of features and the number of columns is the same to the number of classes.

The membership matrix $mf_{a,c}(s_m)$ explains the degree of belonging of different features (a) to different classes (c).

Where, S^m – m^{th} feature value of pattern S.

m - 1, 2, …, m, here number of features is the sum of the window size.

c - 1, 2, …, c here number of classes is 2.

The depiction of pattern is as follows,

$$S = [\textit{sum of the window size}]^T \tag{13}$$

The formula employed to calculate the membership values is represented as below,

$$mf(s) = \begin{cases} 0 & if\ s \le a \\ \dfrac{s-a}{b-a} & if\ a \le s \le b \\ \dfrac{c-s}{c-b} & if\ b \le s \le c \\ 0 & if\ s \ge c \end{cases} \tag{14}$$

Fig. 3 illustrates a triangular membership function for a single fuzzy set. Now, we can see that, a and c value is zero and it attains progressively to a maximum of value one at the centre point b between a and c. Figure 5 illustrates the plot considering all the three membership functions containing overlapping values. Now, the curves for, low, medium and high are revealed for a specific one attribute. The membership functions of the medium needed three parameters and membership function of Low and high are needed two parameters.

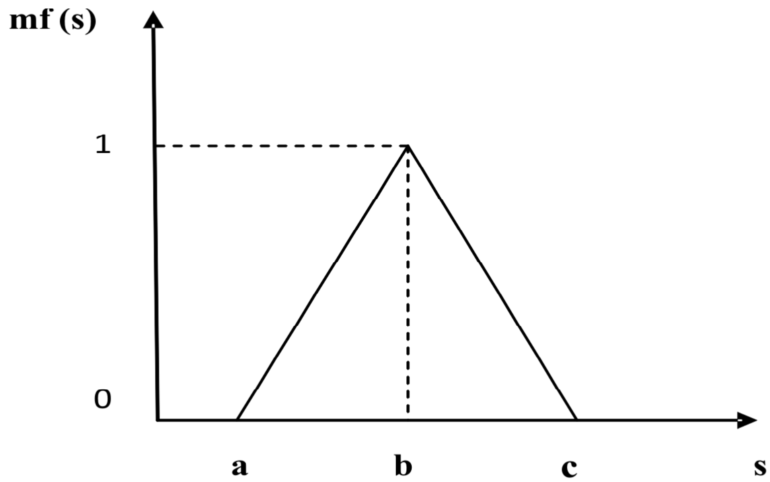

Figure 3: Triangular membership function

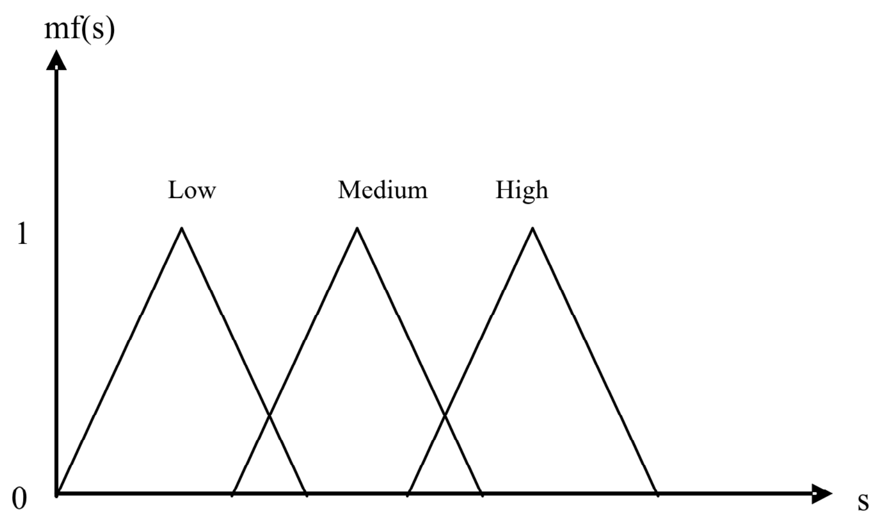

Figure 4: Triangular membership function with defined parameters and their values

The membership function after the fuzzification process is uttered for a pattern *S* as follows,

$$mf(S)=\begin{bmatrix} mf_{1,1}(x_1) & mf_{1,2}(x_1) & mf_{1,3}(s_1) \\ mf_{2,1}(x_2) & mf_{2,2}(x_2) & mf_{2,3}(s_2) \\ mf_{3,1}(x_3) & mf_{3,2}(x_3) & mf_{3,3}(s_3) \\ . & & \\ . & & \\ mf_{sw}(x_5) & mf_{sw}(x_5) & mf_{sw}(s_{sw}) \end{bmatrix} \tag{15}$$

All rows and columns in the membership matrix are tumbled and changed into a vector by this cascading. This produced vector is specified as the input to the Neural Network (NN).

Neural Network

In this, Feed Forward Multi-layer Perceptron classifier is employed which has three layers such as input layer, hidden layer and output layer. The entire number of input nodes of the NN is the same to the product of the number of features and classes.

In this work the total number of output nodes from the NN is similar as that of the number of classes, and now 3 output nodes are produced from the NN. The total number of unseen nodes is equal to the square root of the product, of the number of input nodes and output nodes. In Fig. 5 the configuration of neural network is revealed.

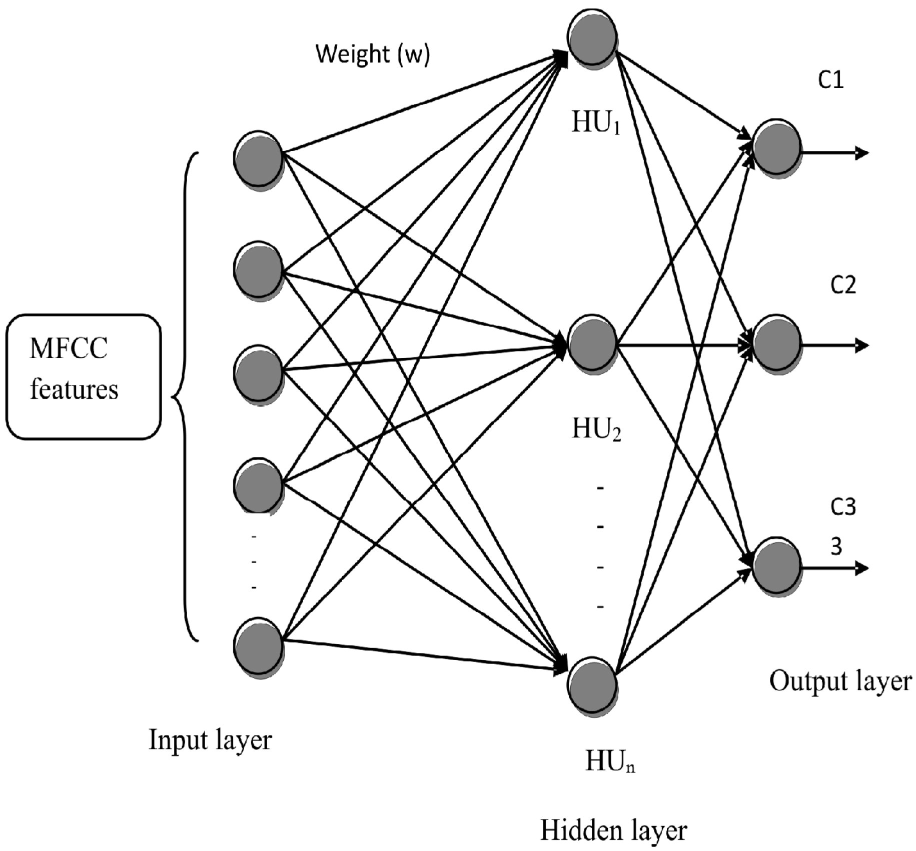

Figure 5: The structure of Neural Networks

Defuzzification

Next the defuzzification process is performed on the output nodes of NN, by executing a MAX (maximum) operation. The output is a single value, from this value; we can competent to categorize the multiple event sound.

4. EXPERIMENTAL ANALSYSIS AND RESULTS

Various input audio signals are taken as the input for the proposed signal separation work. The real world audio signals are bird, animals and river etc. these input audio signals are separate the background sound from the event sound. These input audio signal samples of the proposed work are shown in fig. 6.

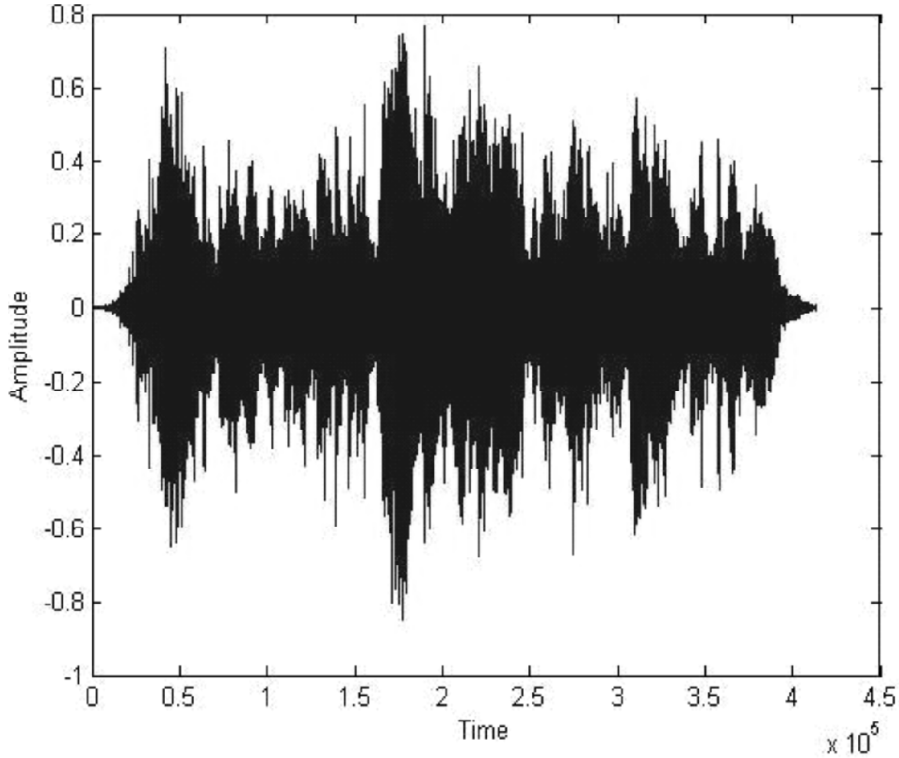

Figure 6 a): Sample jungle bird audio signal

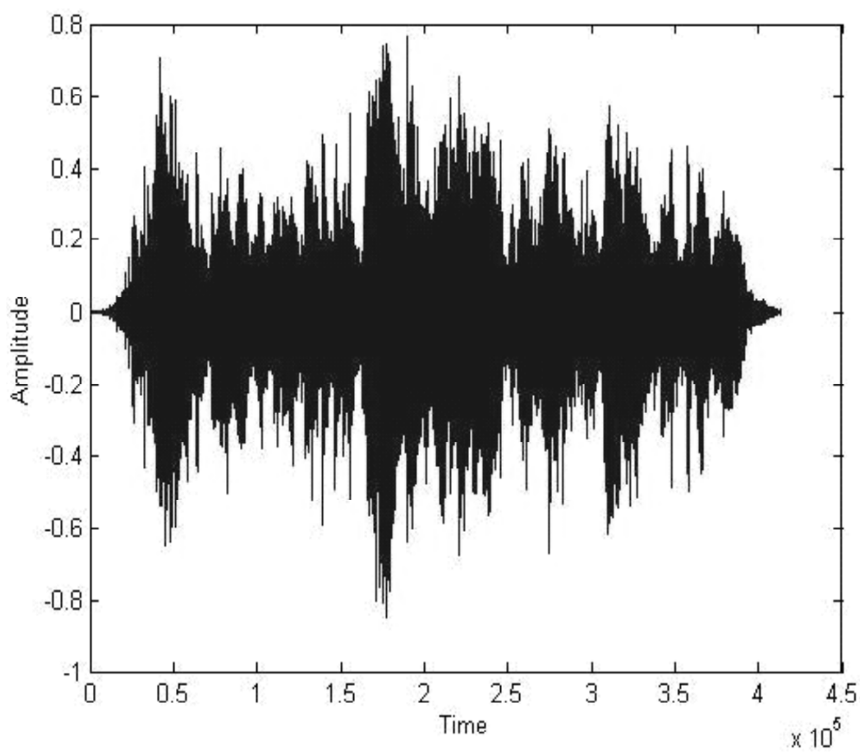

Figure 6 b): Sample tree frog audio signal

The sample input audio signal in Fig.7 are given as the input and that are cluster into background and multiple event sound using the improved k means clustering algorithm with AGA. The results obtained from the improved k means clustering algorithm with AGA is given in Fig.7.

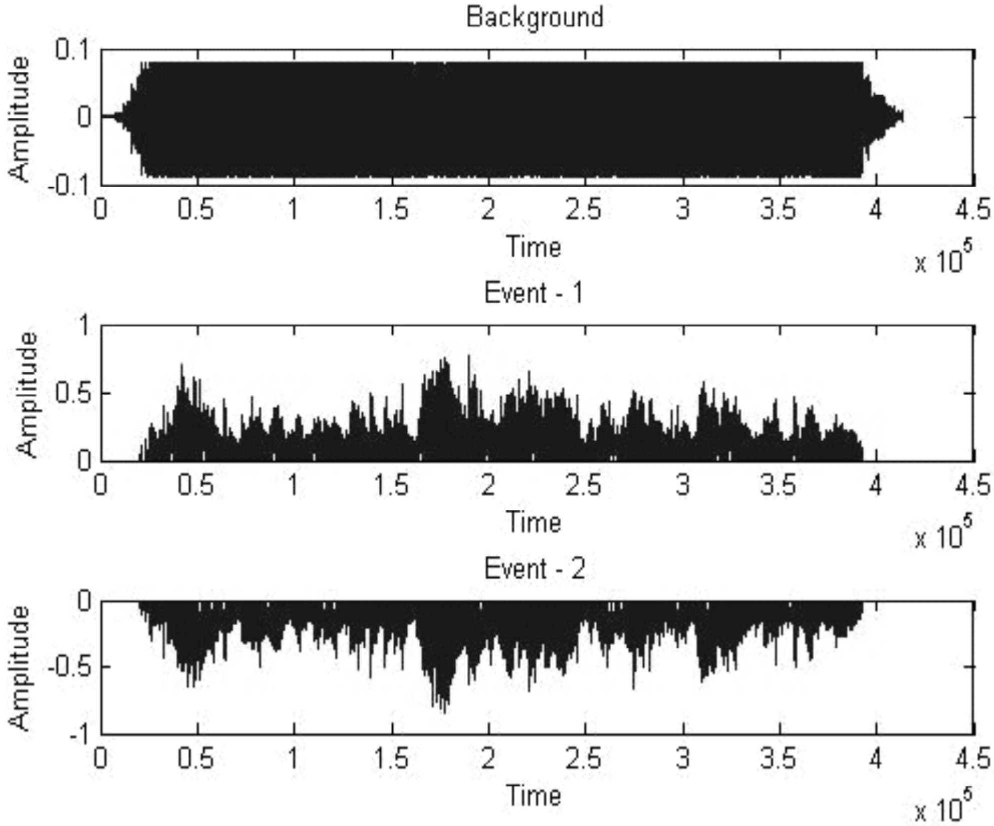

Figure 7 a): Jungle birds sound source separation

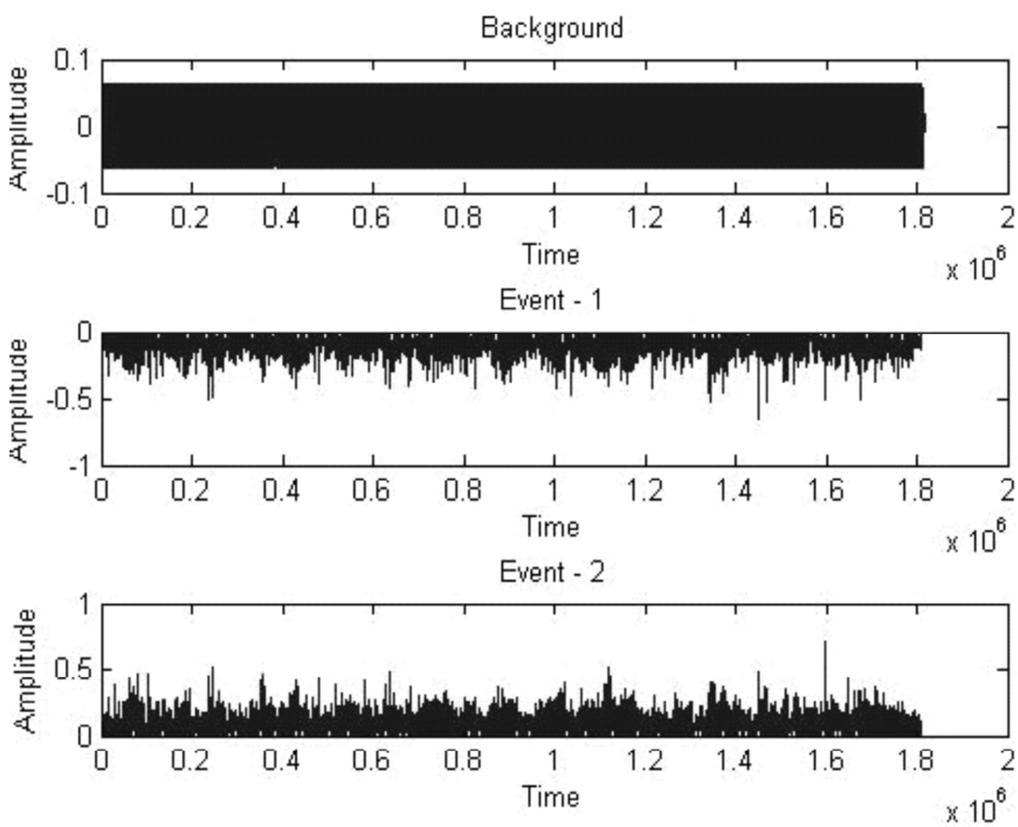

Figure 7 b): Tree frogs sound source separation

Finally we get the background sound and multiple event sound separately. To classify the multiple event sound the proposed method use the neuro fuzzy classifier, this gives the accurate proposed signal results.

4.1. Evaluation Metrics

We require assessing some measures in order to assess the suggested sound source separation. The assessment measures employed in our work are:

- Sensitivity
- Specificity
- Accuracy

To evaluate the detectors, several performance measures can be computed. Most of them use the following values: number of positive instances that are classified as positive (True Positives-TP), number of negative instances that are classified as positive (False Positives-FP), number of positive instances that are classified as negative (False Negatives-FN) and number of negative instances that are classified as negative (True Negative-TN). Given the values of true and false positive and negatives, it is possible to drive the following performance measures.

Sensitivity

The ratio of actual positives which are properly recognized is the measure of the sensitivity. It relates to the capacity of test to recognize positive results.

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \tag{16}$$

Specificity

The ratio of negatives which are properly recognized is the measure of the specificity. It relates to the capacity of test to recognize negative results.

$$Specificity = \frac{TN}{TN + FP} \times 100 \tag{17}$$

Accuracy

We can calculate the measure of accuracy from the measures of sensitivity and specificity as given beneath.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{18}$$

Effectiveness of Classification Results

Effectiveness of classification results by means of neuro fuzzy classifier for the event sound with different metrics value. It is revealed in table 1.

We can launch the competence for the event sound source using neuro fuzzy classifier from the table.1 in our suggested method. False Positive Rate and False Negative Rate values are 0.3333 and 0.3333, correspondingly, which makes clear that our suggested method has low error rate. The suggested neuro fuzzy classifier has the sensitivity value is 75% to the event sound source. Additionally, specificity is another metric that details the percentage of the event sound source; here the classifier provides 87.5%

Table 1
Effectiveness of classification results using Neuro fuzzy classifier for the event sound

Metrics	*Neuro fuzzy classifier*
TP	1
TN	2.3333
FN	0.3333
FP	0.3333
Sensitivity	75%
Specificity	87.5%
Accuracy	83.3%

specificity value. The accuracy value of the suggested method is 83.3% so that the suggested method illustrate that the high classification accuracy of the event sound source classification. The suggested effectiveness of classification in Fig. 8.

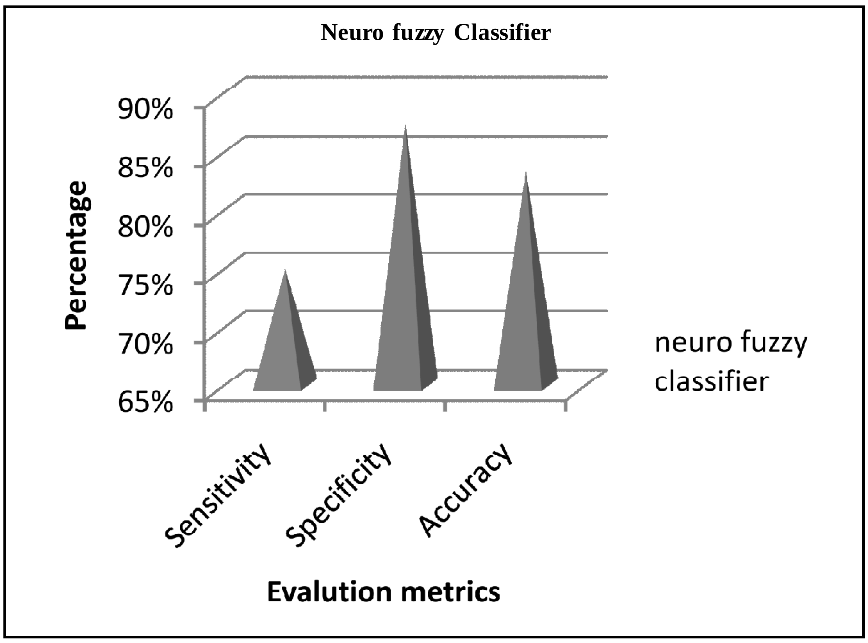

Figure 8: Effectiveness of the proposed classification

5. PERFORMANCE ANALYSIS

The presentation of our suggested sound source separation work is examined based on the evaluation measures Sensitivity, Specificity and Accuracy with different work.

5.1. Comparative Analysis for our proposed work with the existing works

Sound source separation using improved *k*-means clustering algorithm is compared with traditional *k*-means clustering algorithm. The result is tabulated in table.2 and the comparison result is plotted in Fig. 9.

Table 2
Comparison value with and without Optimization

Input Sample	*Improved k-means algorithm with optimization*	*k-means algorithm without optimization*
Jung bird	1.92E-11	2.20E-11
Tree frog	1.07E-12	1.14E-12

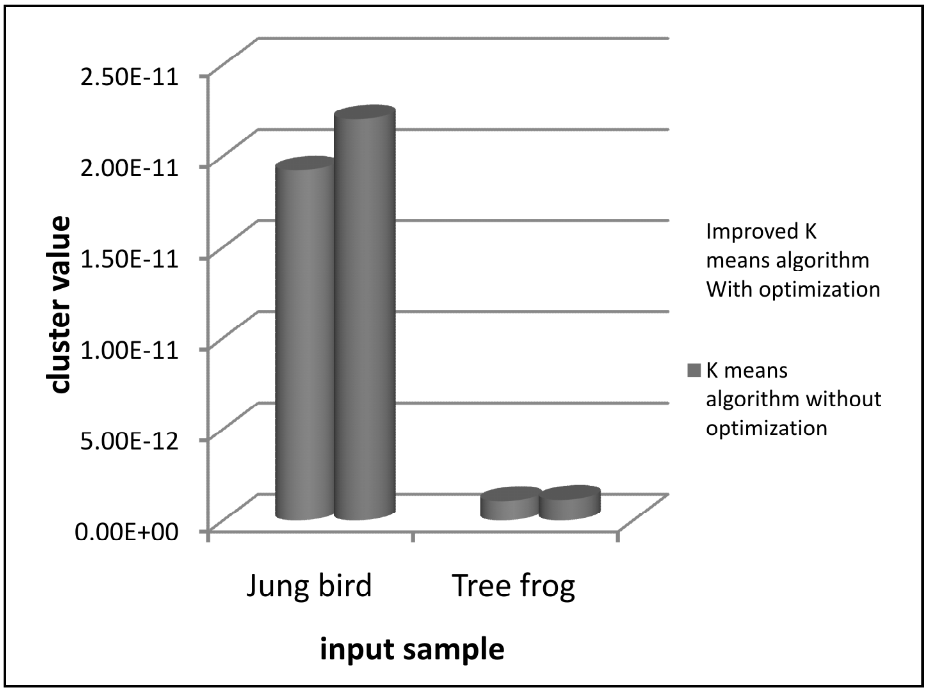

Figure 9: Comparison value of with and without optimization

Table 3
Comparison result of event sound source with other classifier

Metrics	*Artificial neural network*	*Neuro fuzzy classifier*
TP	0.6667	1
TN	2	2.3333
FN	0.6667	0.3333
FP	0.6667	0.3333
Sensitivity	50%	75%
Specificity	75%	87.5%
Accuracy	66.7%	83.3%

Sound source of our suggested work makes use of neuro fuzzy classifier for the classification of event. We can detect that our suggested method has very good accuracy for the categorization of event sound source using neuro fuzzy classifier by comparing the other classifier. Now we contrast the artificial neural network for our similarity. The comparison effect revealed in the Table 3.

The enhanced accuracy of the classification of event sound source is offered by our suggested method. The artificial neural network presents low accuracy compared to the suggested neuro fuzzy classifier. The sensitivity value of the artificial neural network is 50% it is extremely low contrasted to our suggested classifier. The sensitivity value of our classifier is 75%. The specificity value of the suggested method attains the 87.5% when compared to the artificial neural network it is high, because the specificity value of the artificial neural network is 75%. On the other hand the accuracy value is 83.3% for our neuro fuzzy classifier and the artificial neural networks encloses low accuracy result of 66.67%. From these results, it is recognized using neuro fuzzy classifier in our suggested method offers very good classification of event sound with high accuracy, sensitivity and specificity vales. Hence our work illustrates the improved classification of event sound source.

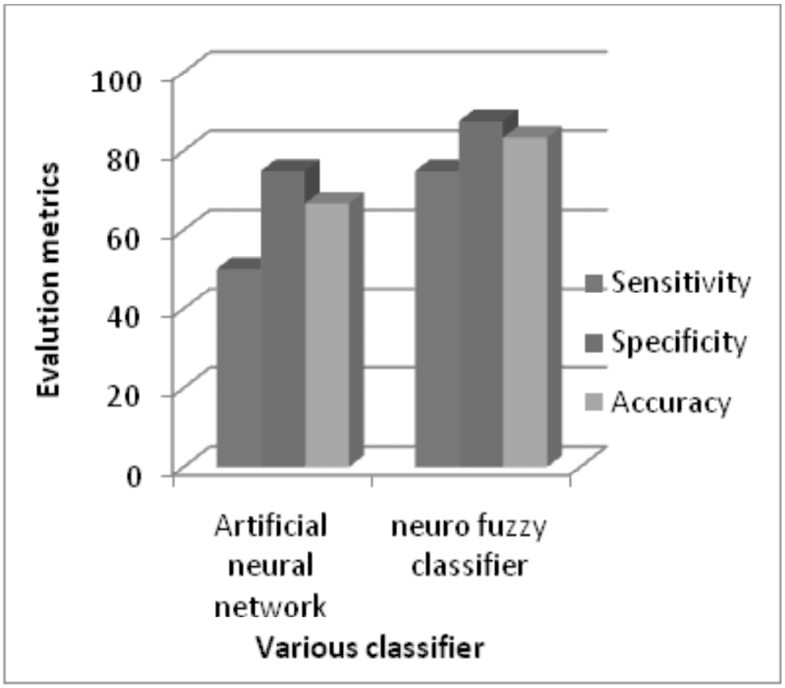

Figure 10: Comparison graph for the event sound source with other classifier

The accuracy of the suggested method is compared to presented work. The suggested method attains the improved accuracy value.

Table 4
Comparison Result of our proposed method

Metrics	*Proposed method (Neuro fuzzy classifier)*	*Existing method (Artificial Neural network)*
Accuracy	83.3%	66.7%

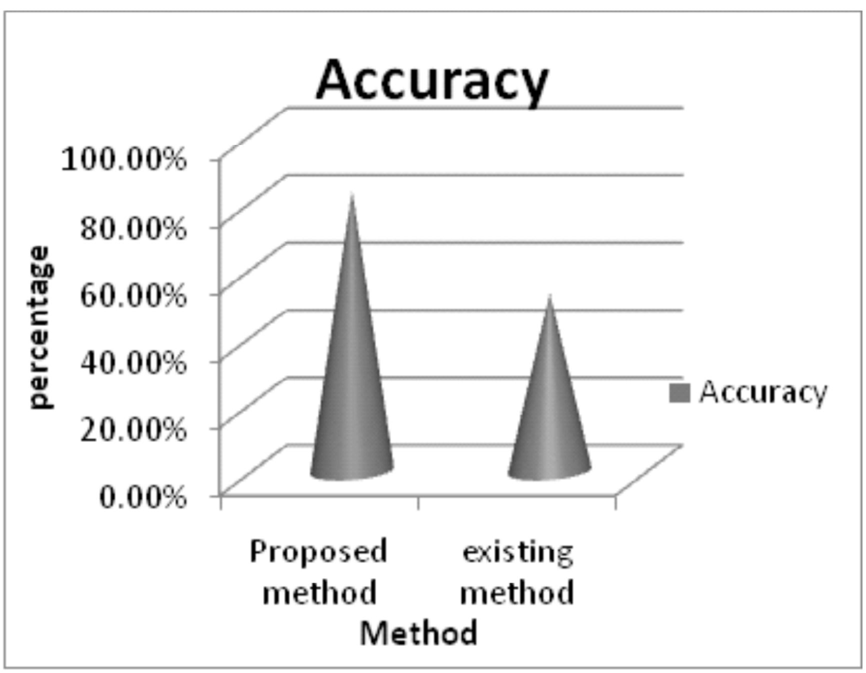

Figure 11: Comparison of existing method

6. CONCLUSION AND FUTURE WORK

Sound source separation by means of enhanced *k*- means clustering with adaptive genetic algorithm. The input audio signal first detaches the background sound and event sound by means of the enhanced *k*-means clustering algorithm. Now the centroids values are chosen by the adaptive genetic algorithm. Next we categorize the event sound by means of the neuro fuzzy classifier. The effect of the suggested method accomplishes the better classification with high accuracy, sensitivity and specificity value of the event sound source. The suggested classifier has high accuracy values when compared to the other classifier.

REFERENCES

[1] Toni Heittola, Annamaria Mesaros, Tuomas Virtanen and Antti Eronen, "Sound Event Detection in Multisource Environmental using Source separation", IEEE International Conference, 2008.

[2] Kamaran Rahbar and James P. Reilly, "A New Frequency Domain Method for Blind Source Separation of Convolutive Audio Mixtures", IEEE Transactions on Speech and Audio Processing, Vol.13, No.5, 2005.

[3] Kazuyoshi Yoshii, Ryota Tomioka, Daichi Mochihashi and Masataka Goto, "Beyond NMF: Time-Domain Audio Source Separation without Phase Reconstruction", ISMIR, pp.369-374, 2013.

[4] Emmanuel Vincent, Remi Gribonval and Cedric Fevotte, "Performance Measurement in Blind Audio Source Separation", IEEE Transaction on Audio, Speech And Language Processing, Vol.14, No.4, 2006.

[5] Shoji Makino, Shoko Araki, Ryo Mukai and Hiroshi Sawada, "Audio Source Separation Based On Independent Component Analysis", In. proc. Of International Symposium on Circuits and Systems, 2004.

[6] Niva Das, Aurobinda Routray and Pradipta Kishore Dash, "ICA Methods for Blind Source Separation of Instantaneous Mixtures: A Case Study", Neural Information Processing – Letters and Reviews, Vol. 11, No.11, 2007.

[7] Michael Syskind Pedersen, Jan Larsen, Ulrik Kjems and Lucas C. Parra, "A Survey of Convolutive Blind Source Separation Methods", Speech Processing and Speech Communication, Springer, 2007.

[8] S. Mirzaei, H. Van hamme and Y. Norouzi, "Blind Audio Source Separation of Stereo Mixtures using Bayesian Non-Negative Matrix Factorization", Signal Processing, Vol. 115, 2011.

[9] Joan Bruna, Pablo Sprechmann and Yann LeCun, "Source Separation With Scattering Non-Negative Matrix Factorization", Machine Learning for Speech and Audio Processing, 2012.

[10] Tom Barker and Tuomas Virtanen, "Non-negative Tensor Factorisation of Modulation Spectrograms for Monaural Sound Source Separation", ISCA, pp.827-831, 2013.

[11] Satoshi Innami and Hiroyuki Kasai, "NMF-based environmental sound source separation using time-variant gain features", Computers and Mathematics with Applications, Vol. 64, pp.1333-1342, 2012.

[12] Alexey Ozerov and Cedric Fevotte, "Multichannel Nonnegative Matrix Factorization in Convolutive Mixtures for Audio Source Separation", IEEE Transaction On Audio, Speech, and Language Processing, Vol.18, No.3, 2010.

[13] Jaya Kulchandani and Kruti J. Dangarwala, "Blind Source Separation via Independent Compontent Analysis: Algorithms and Applications", International Journal of Computer Science and Information Technologies, Vol.5, No.5, 2014.

[14] Virtanen. T, "Monaaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria", IEEE Transactions on Audio, Speech, and Language Processing, Vol.15, No.3, 2007.

[15] Otsuka. T, Ishiguro. K, Yoshioka. T and Sawada. H, "Multichannel Sound Source Dereverberation and Separation for Arbitrary Number of Sources Based on Bayesian Nonparametrics," IEEE ACM Transactions on Audio, Speech and Language Processing, Vol.22, No.12, 2014.

[16] Nikunen. J and Virtanen. T, "Direction of Arrival Based Spatial Covariance Model for Blind Sound Source Separation", IEEE ACM Transactions on Audio, Speech and Language Processing, Vol.22, No.3, 2014.

[17] Toni Heittola, Annamaria Mesaros, Tuomas Virtanen, Antti Eronen, "Sound Event Detection in Multisource Environments Using Source Separation", In proceedings of international conference on the environmental science, pp.1-5, 2011.