An Integrated Approach for Deaf Speech Enhancement Using Modified SOM Clustering for Deaf Speech Enhancement

NirmalaDevi Jaganathan and BommannaRaja Kanagaraj

1Department of Information Technology, Excel Engineering College, Namakkal, India
2Department of Electronics and Communication Engineering KPR Institute of Engineering and Technology, Coimbatore, India.

Received 25 April 2016; Accepted 28 May 2016; Available 2 June 2016

ABSTRACT
S Development in deaf speech recognition techniques shows the significant progression that supports deaf speakers in communicating with associated inhabitants. The use of efficient clustering techniques paves the way for better recognition of the unclear deaf speech so that it could be understandable to normal speakers. The SOM clustering has been found to be the efficient technique but it has a drawback with its fixed architecture as it is not supported by the larger speech data. Also, the deaf speech can be enhanced by reducing the distance between the deaf and normal speech upon recognition, which can be measured by introducing a new distance measure. Hence, in this paper to improve the clustering performance, the normal and deaf speech signals are clustered using two new clustering algorithms: Non-negative Matrix Tri-Factorizations Clustering (NMTF) and Modified Self Organizing Map (M-SOM) clustering. The M-SOM algorithm is developed by combining the adaptive features of SOM clustering and the Ward clustering methods. Using Ward clustering, each node of the SOM is considered as clusters and at each step, two clusters with minimum distance are merged together. It estimates the distance measure for each cluster by the use of nodal character of the map and a topological location of clusters. Thus, the speech signals can be clustered effectively. The comparative study of three clustering methods, SOM, NMTF and M-SOM show that the M-SOM clustering provides better association of similar group. Hence effective recognition of the deaf and normal speech signal is possible which in turn help to estimate the differential corrective measure for possible deaf speech enhancement.

KEYWORDS: Clustering, Non-negative Matrix Tri-Factorizations clustering, Modified SOM clustering, Ward clustering, Bayesian Information Criterion (BIC), Deaf Speech Enhancement

INTRODUCTION
Speech processing has become an interesting field of study in the recent years; however it includes acquisition, manipulation, analysis, recognition, enhancement, storage and transfer of speech signals [1]. Speech processing techniques such as spatial filtering, spectral and model-based approaches are developed for improving the quality the speech signals [2]. Modern techniques for the speech enhancement process play a vital role in the deaf speech analysis which increases the accuracy of speech recognition, speech separation, and deaf speech enhancement. Such deaf speech processing has paved the way for implementing numerous applications in the medical field, in addition to developing gadgets for assisting deaf speakers. The fundamental problem associated with deaf speakers is unclear production of audible speech. This demands enhancement of deaf speech signals to understand the properties of deaf speech compared to the normal speakers [8].

To Cite This Article: NirmalaDevi Jaganathan and BommannaRaja Kanagaraj., An Integrated Approach for Deaf Speech Enhancement Using Modified SOM Clustering for Deaf Speech Enhancement. Advances in Natural and Applied Sciences, 10(8); Pages: 56-64
The advancement in deaf speech signal enhancement techniques has been considered to be a significant accomplishment in the arena of deaf speech analysis. Even though, the deaf speech processing follows the similar techniques of normal speech processing, the complexity involved in understanding the deaf speech patterns for a specific word uttered is much higher. Further speech descriptive parameters vary indiscriminately for deaf speakers which limit assessment. Therefore a suitable method need to be developed that could estimate speech parameters for possible quantification, discrimination and recognition. The study on the quantifying parameters of deaf speech signal reveals that the variation is little obvious with respect to normal speakers [12]. Hence certain decision making algorithm may be applied to demarcate the speech signals. Such grouping helps to recognize the speaker category, thereby assisting to evaluate the distance metrics to understand their degree and depth of dislocation. This measurement can further be used to synthesis normal speech from deaf speech signal. In this paper, proper clustering methods are employed to constellate the speech signals into clusters or groups, which further, as mentioned, assist enhancement using correction factors.

Non-negative Matrix Factorization (NMF) and Self-Organizing Maps (SOM) clustering methods have been employed in previous works for effective clustering of speech signals [13]. Though SOM clustering provided improved clustering efficiency than NMF clustering technique, it has fixed architecture which is a major drawback. The fixed architecture of SOM cannot support large speech data sequence, as in the present case, where comparison between intra- and inter- category is carried. Furthermore, algorithm need to supervise different words uttered. Two new clustering techniques called Non-negative Matrix Tri-Factorization (NMTF) clustering and Modified SOM clustering (M-SOM) have been proposed. The proposed NMTF and M-SOM methods assign the input speech signals into a matrix and then segment the row and columns of the matrix to form the speech clusters. The NMTF clustering segments the matrix into three factors and clusters the speech signals. The M-SOM utilizes the adaptive features of SOM along with the Ward clustering approach for effective clustering.

The paper is organized as follows; section II, explains the related research works briefly. Section III, presents the methodologies adopted to derive suggested clustering methods. Section IV, provides the experimental results and their discussions. Section V, concludes the work with the findings.

Related Works:

Guang Li et al [1] proposed Mandarin digital signal recognition technique that uses the chaotic neural networks and the fuzzy c-means clustering for speech clustering. The Mandarin technique extracts the features of the Freeman’s KIII model and Fuzzy C-means methods and employs in the neural network as the Mandarin classifier. The approach enhances the possibility of detecting the style and features of the speech signal thus improving the signal recognition. But this approach suffers from problems such as optimization of parameter, pattern recognition problems and more iteration of the neural networks.

Tatsuya Hirahara et al [3] suggested the enhancement of non-audible murmur (NAM) speech signals using the body-conducted vocal-tract resonance signals. The method proposed three speech enhancement techniques namely direct amplification of weak vocal-tract resonance signals using a wired urethane-elastomer NAM microphone, a wireless urethane-elastomer-duplex NAM microphone, a soft-silicone NAM microphone for enhancing whispered speech. The highly sensitive sensors improve the collection of the murmur speech and thus the speech could be enhanced. Still the sensor system demands certain improvement by reducing the noise.

Nafiseh Esfandian et al [4] presented a feature selection method for the effective speech recognition. The approach utilized the Gaussian mixture models (GMM) and weighted K-means (WKM) clustering techniques for reducing the dimensions of the features space in the spectral-temporal domain. The clustering techniques employ the elements of centroid vectors and covariance matrices of clusters as attributes of the secondary feature vector of each frame. The phonemes features are segmented using spatial GMM. Spatial GMM input vectors include the position attributes in addition to the representation attributes at that point to reduce the inaccuracies in speech recognition. The approach has a minor problem by the formation of noise sequences in the clustered segments.

Christian D. Sigg et al [5,6] presented generative dictionary learning for the effective speech enhancement. The objective of speech signals enhancement is to separate a single mixture speech combined with noise into its underlying clean speech and interferer components. In this approach, the clean speech is recovered from the degraded speech by sparse coding of the mixture by LARC in a composite dictionary consisting of the concatenation of a speech and an interferer dictionary. The presence of signal source confusion risk problems reduces the performance.

Nasser Mohammadiha et al [7] presented a speech enhancement method based on the Bayesian formulation of Non-negative Matrix Factorization (BNMF) for the efficient enhancement of the speech signals. In this speech enhancement method, the temporal dependencies of speech and noise signals are used to generate information prior distributions. These information prior distributions are then applied in a Bayesian framework to perform NMF. An environmental noise classification technique called Hidden Markov Model (HMM) is utilized for effective separation of noise signals from the original speech signals. Additionally an online
dictionary learning scheme is proposed to learn the noise basis matrix in order to improve the speech signal enhancement. The fundamental problem with the approach is the computational complexity.

Jian Zhou et al [9] proposed an approach focusing on the unsupervised learning of phonemes of whispered speech in a noisy environment with the use of NMF for the effective clustering and extraction of phonemes of whispered speech. The approach uses the noise bases from the prior learning’s which are trained by the NMF method. The divergence function with a sparseness constraint term is selected for the development of the algorithm to obtain multiplicative update rules of the phoneme base matrix and the corresponding weight matrix. The approach does perform differently when the results of prior learning are not perfect.

Xixin Wu et al [10] proposed automatic speech clustering method based on an extended version of the k-means clustering called as x-means clustering. The automatic clustering approach presented in this paper, utilizes the human perception based weighted distance for the classification of the human speech with different styles and emotions, capture the contribution of different acoustic features on human perception. The approach has been effectively used for the automatic classification of internet speech data in order to improve the clustering of the original speech signals from the noise signals. The method performs poorly when there is no prior knowledge on human perception.

Herman Kamper et al [11] suggested a novel lexical clustering approach using the acoustic embeddings. The clustering approach is meant for the unsupervised lexical clustering of speech segments and the approach utilized Infinite Gaussian mixture model (IGMM), a probabilistic method for automatic selection of number of clusters. IGMM clustering improves the overall clustering of the lexical variable-length word segments which are embedded in a fixed-dimensional acoustic space. The English corpus models are utilized to evaluate the performance of the clustering technique. The segmentation of the transcribed lexical words is a major drawback.

Methods:

Deaf speech enhancement is a tedious process in the speech processing domain. The deaf speech or muted speech can be enhanced by comparing the deaf speech signals with the normal speech signals. Even though such comparison provides the degree of deviation, parametric differences and validation of category, the enhancement of deaf speech signal pertaining to correction is not possible. Generally, the existing methodologies help for characterization of deaf speech signal under various environments and using different techniques. Therefore, the unique scheme that uses the speech parameters for enhancement by identifying deviation and exploring this to derive corrective measure is in high demand. In such case, followed by speech parameters evaluation, to isolate the groups, an appropriate clustering technique is needed. In this paper, an effort has been taken to identify the suitable clustering technique to group the normal and deaf speech effectively. With such clustering, it is possible to evaluate the deviation across the group by determining the corrective measure [14].

In our previous work, the features such as Pitch, Formants, Intensity, Jitter, Shimmer, Spectral wave, Spectrogram, LPC Spectra for Voiced and Unvoiced region, PLP (Perceptual Linear Prediction), MFCC, Cepstrum, Vocal tract extraction, Signal to Noise Ratio, MELP are extracted for both the normal and deaf speech signals. The NMF and SOM clustering has been presented to cluster the speech signals using these features. Though the approach provides efficient clustering, the fixed architecture of SOM makes it difficult to cluster large speech data sequences effectively. Similarly, the difference between the normal and deaf speech is more and arbitrary. Hence it lacks definite distance metric measurement. Consequently, the NMF and M-SOM clustering methods are proposed for the effective clustering of the speech signals. Based on the clustering results a distance formulae based on the energy level of the signals may be used to reduce the difference level of the signals that improves the deaf speech signals.

3.1 Non-negative matrix factorization (NMF) clustering:

A NMF is a group of algorithms that can be used to classify the speech signals based on the speech parameters. Normally, this clustering technique assigns the speech data into a matrix which is factored into two smaller matrices. The significance is that the product of the smaller matrices forms the parent matrix and all the three matrices have no negative elements. This non-negativity is the main advantage of using this approach for speech processing as it reduces the complexity of processing.

Let the speech data of both normal and deaf speakers can be represented in the form of a matrix X. The matrix X is the input matrix which can be factored into two smaller matrices U and V. X has the property that the product of the two smaller matrices will form the input matrix. The matrix multiplication must follow a unique property that the dimensions of smaller matrices should be lower than the input matrix X. This is the basis of the NMF clustering.

NMF has an inherent clustering property in which the columns of the input matrix can be clustered automatically. Using this property, the features of speech can be clustered. Let the input matrix X is with \( m \times n \) dimensions. Now the factorization is performed by generating a feature matrix U with m rows and k columns where k is a factor value. Similarly, another matrix called coefficients matrix V with k rows and n columns. In
this case, when the matrices are multiplied, it forms a matrix of $m \times n$ with the same size of $X$. Applying the Kuhn-Tucker condition, the non-negative matrices $U$ and $V$ can be obtained and normalized. Taking transpose of $V$ forms the clusters.

The rows of the transpose matrix can be clustered into $n$ clusters by the minimization process to classify the difference range between the different speakers. The normal speaker and deaf speaker speeches are clustered into groups based on the parameters extracted. Thus, the clustering of speech features in the difference range can be computed so that the speech can be enhanced in the testing process of the research where the deaf speech can be corrected.

Algorithm:

Input: Speech dataset $S$ of normal and deaf speakers; list of features $\theta = \{ F_1, F_2, ..., F_i, C \}$
Output: Clusters based on features
Initialize
Clustering based on features
For $i=1$ to $m$ do
List $= SU(F_i, C)$
If $F \geq \theta$ then
$S=S \cup \{F_i\}$;
Generate nonnegative $m \times n$ matrix $X$ using the speech data $S$
Factorizes $X$ into the non-negative $m \times k$ matrix $U$ and the non-negative $k \times n$ matrix $V^T$ that minimizes the objective function,
Applying Kuhn-Tucker condition $a_{ij}u_{ij} = 0$ and $\beta_{ij}v_{ij} = 0$
Obtaining the two non-negative matrices $U$ and $V$
Normalizing $U$ and $V$
Taking transpose of obtaining a matrix of $V$ and get $V^T$
$V_{k \times n}^T = V_{n \times k}$
(1)
Rows are speech data & columns are featured in $V^T$ matrix
Rows in the $V^T$ matrix formed the clusters thus $n$ number of clusters are created
From each cluster, choose a representative feature and include in $S$
Return $S$

3.2 Self-organizing Map (SOM) approach:

The self-organizing map approach is an unsupervised learning approach that can perform feature based clustering. As the approach is unsupervised, the clustering is driven by speech data and does not have fixed target results. This approach can classify the features using the competitive learning networks and uses a neighborhood function to preserve the properties of the input speech data. The principal goal of an SOM in our approach is to transform an incoming speech signal pattern of arbitrary dimension into a one or two-dimensional discrete map and to perform this transformation adaptively in a topologically ordered fashion.

The speech data from the normal and deaf speakers which are needed to be clustered are assigned to the nodes of the network. The input map units are taken as $n$ which are the training node vectors of the speech data and the output units are taken as $m$ which are the number of clusters. Initialize the neighbourhood distance and the weights from input nodes to the output nodes. Select an input data and check for the computational bounds while setting an initial leaning time $t$. Then compute the Euclidean distance of input nodes to the output nodes. Then select the least weight output node as the first category. Update the weights using weight update rule and increment value of $t$ to continue clustering of next data. Likewise, all data are utilized and the category to which each speech data belong is assigned. Thus, the clustering of the speech data can be performed effectively for both the speakers. This clustered speech data can be enhanced to correct the deaf speech.

Algorithm:

Input: Speech data as $N$ node vectors; list of features
Output: $m$ clusters based on features
Select output layer network topology
Assign speech data as input nodes
Initialize output nodes
Initialize current neighborhood distance, $D(0)$, to a positive value
Assign weights to nodes
Initialize weights from inputs to outputs to small random values
Let $t = 1$ // $t \equiv$ time for learning
While computational bounds are not exceeded
Do
Select a feature $F_i$ \  \  // $i = 1, 2, 3… n$
Compute the square of the Euclidean distance of $i$ from weight vectors ($w_j$) associated with each output node
\[
\text{Weight} = \sum_{k=1}^{n} (\mathbf{l}_{ik} - w_{ik}(t))^2
\]  
(2)
Select output node $j$ that has weight vector with minimum values
Update weights to all nodes within a topological distance given by $D(t)$ from $j$, using the weight update rule
Store the results under $F_i$
Increment $t$
Continue until $F_n$
End while

3.3 Non-negative Matrix Tri-Factorization for clustering (NMTF):

The NMTF clustering approach clusters the data into 3 factors is similar to the normal 2-factor NMF clustering which segments the speech data into 2 factors. But the orthogonal constraint feature is the one that varies the NMTF clustering. The orthogonally unconstraint NMTF clustering is similar to the unconstrained 2-factor NMF but the orthogonally constraint NMTF is different from the 2-factor NMF with new features that enhances the clustering by overcoming the architecture problems in SOM. The orthogonally constraint NMTF approach improves the clustering by using class aggregate distribution and multi-peak distribution.

The general 2-factor NMF factorizes the input non-negative speech data matrix into 2 non-negative matrices. $X$ is the non-negative input matrix while the two factored matrices are $F$ and $G$.

\[
X \approx FG^T
\]  
(3)

Where $X$ is in the order of $m \times n$; $F$ is the order of $m \times k$ and $G$ is in the order of $n \times k$. $X$, $F$, $G$ are all derived from the set of all non-negative matrices. The genuine condition is that the rank of the non-negative matrices $F$ and $G$ is always lower than the rank of $X$.

The orthogonality of the matrices is needed to be imposed in order to improve the indicative level of the NMF. The single orthogonal NMF can be achieved by employing orthogonality to one of the two factors $F$ or $G$. But the bi-orthogonal NMF can provide effective clustering with better accuracy. Hence the orthogonality can be imposed on both $F$ and $G$.

\[
\min_{F \geq 0, G \geq 0} \|X - FG^T\|^2
\]  
(4)

This function is only possible if the two conditions are satisfied such that $FF^T = I$, $GG^T = I$. Thus the bi-orthogonality can use these features to cluster the columns and rows similar to the k-means clustering. The problem with the double orthogonality is that it is very restrictive and gives poor matrix low-rank approximation.

The NMTF can overcome the poor matrix low-rank approximation problem by introducing $S$ of the order of $k \times \ell$. The NMTF also follows the rank condition. The NMTF can be represented by including an additional factor $S$ along with $F$ and $G$. $S$ absorbs the different scales of $X$, $F$ and $G$. $S$ provides the additional degrees of freedom so that the low-rank matrix representation remains accurate while the other factors $F$ and $G$ gives the row clusters and the column clusters respectively. NMTF can be represented by

\[
X \approx FSG^T
\]  
(5)

Where $X$ is in the order of $m \times n$; $F$ is the order of $m \times k$, $S$ is in order of $k \times \ell$ and $G$ is in the order of $n \times \ell$. $X$, $F$, $S$, $G$ are all derived from the set of all non-negative matrices. For orthogonality function, the 3-factors can be optimized

\[
\min_{F \geq 0, G \geq 0, S \geq 0} \|X - FSG^T\|^2
\]  
(6)

This approach allows the number of row clusters $(k)$ to slightly differ from the number of column clusters $(\ell)$ in order to improve the clustering application. The rank conditions are needed to be satisfied for the $F$ and $G$ factors which means that at most of the cases the number of row clusters is equal to the number of column clusters such that $k = \ell$. This provides an efficient approach of clustering $X$.

The main advantage of NMTF is the presence of factor $S$ in the middle as it provides the additional approximation required for effective clustering of the matrix $X$. The orthogonality can be imposed in the three factors $F$, $S$, $G$ of which the orthogonality of $F$ and $G$ are approached similar to that of $F$ and $G$ in the 2-factor NMF.
Consider the unconstrained NMTF given in eqn (6). In order to compute the 1-orthogonality, the NMTF can be reduced to unconstrained 2-factor NMF by using the mapping concept. This makes the degree of freedom of $FSG^T$ equal to that of $FG^T$. The mapping concept is applied as

$$F \leftarrow FS$$

(7)

But the mapping approach makes the 1-orthogonal NMTF similar to that of the 2-factor NMF. Hence the bi-orthogonality can be imposed in the unconstrained NMTF to differ it from that of 2-factor NMF. The bi-orthogonal factorization can be computed using the update rules.

$$G_{jk} \leftarrow G_{jk} \left(\frac{X^TFS}{XG^TSG^T}\right)_{jk}$$

(8)

$$F_{ik} \leftarrow F_{ik} \left(\frac{FXG^T}{FG^TSG^T}\right)_{ik}$$

(9)

$$S_{ik} \leftarrow S_{ik} \left(\frac{(FXG)^T}{F^TSFG^T}\right)_{ik}$$

(10)

Where $i$ is the row of the non-negative matrices and $j$ is the columns of the non-negative matrices. These update rules enable the efficient clustering of the matrix X orthogonally different from the 2-factor NMF. The issue arises at this point of the clustering as that the optimization problem in the factors. The optimization problem can be reduced by utilizing the correctness and convergence approaches.

$F$ and $G$ can be optimized similar to 2-factor NMF. So let $F, G$ be fixed matrices. When monotonically reducing under the update rule in eqn (10), the correctness is estimated as

$$J(s) = \|X - FSG^T\|^2$$

$$J(s) = Tr(XX^T - 2G^TXX^TFS + F^TSFG^TGS^T)$$

(11)

Where $J(s)$ is the correctness while the $Tr$ is the product of optimization. The complementary condition for the NMF is only if $S_{ik}$ gives

$$(-F^TXG + F^TSG^TG)_{ik}S_{ik} = 0$$

(12)

The convergence of the NMTF can be calculated by satisfying the auxiliary function approach.

$$Z(S, S') = \|X\|^2 - 2Tr(F^TXG) + \sum_{ik} \left(\frac{(FXG)_{ik}}{S_{ik}}\right)_{ik}$$

(13)

This condition is equal to the monotonically decreasing update rule in eqn (10).

$$S_{ik} = S_{ik} \left(\frac{(FXG)_{ik}}{F^TSFG^T}\right)_{ik}$$

(14)

Thus the convergence can be achieved and it makes the clustering very effective without the optimization problem.

### 3.4 Modified Self-Organizing Maps (M-SOM) clustering:

The M-SOM clustering approach has been proposed by combining the adaptive features of the SOM clustering and the Ward clusters to overcome the problem of the fixed architecture of the normal SOM clustering. The M-SOM utilizes the Ward clustering method to cluster the maps generated in the SOM. The SOM with the Ward consists of unsupervised training and mapping approaches. The training process consists of three phases: sampling the clusters, finding the maximum correspondence and adjustment phases. The training process provides the sorting of the input information in the form of user defined maps either in one-dimensional or the two dimensional maps. The multi-dimensional maps are represented in terms of vectors which have its own coordinate on the maps and neighbouring coordinates of two vectors on the maps closer to the input space. The topological map can be generated in the multi-dimensional input space providing a visual representation of the speech data. The proximity relation of the topological maps is kept locally closer regions in the closer input space. The visualization of the topological map is similar to the coloration of the geographical maps. The feature data extracted from each map generated its own coloring cells determined by the average value of each feature in the data present in the cell. When all the features are gathered together, it resembles a atlas of maps for effective clustering the multi-dimensional speech data.

The fixed architecture can be made flexible in order to adapt the larger speech sequence. The adaptive features are associated with the units of the resources and gather the statistical information. The column and row architecture of the SOM is extended to the larger speech data by splitting process. The parent columns and rows are split and the data are executed twice on the current architecture. The splitting process considers the cluster centers for the effective segmentation of the columns and rows. Now there will be a number of architectures and
the Bayesian Information Criterion (BIC) score is calculated for all the architecture. The map architecture with the larger BIC score is selected for the new speech data.

Ward clustering approach operates by considering the nodes in the maps as clusters in order to cluster the map points. Among the considered clusters, two clusters which have minimum distance between them are selected and merged together. This process of selection and merging two clusters continues at every step of the process. The distance measure used in the Ward approach is based on the principle that it should provide a low variance when inside a cluster while a high variance is provided between the clusters. The merging of the clusters results in the combination of small proportion of the variance. The Ward formula for the distance between the clusters can be given by

$$d_{pq} = \frac{n_p n_q}{n_p + n_q} \| \bar{x}_p - \bar{x}_q \|^2$$

(15)

Where \( p, q \) are two clusters and \( n_p, n_q \) are the number of data points in the clusters. The terms \( \bar{x}_p, \bar{x}_q \) are the cluster centroids given in the \( \| \| \) Euclidean norm. The cluster centroids and the number of data points are needed to be calculated to determine the distance.

Ward method has been modified to fit the SOM clustering approach as the distance measure uses the nodal character of the map and a topological location of clusters. Thus when the clusters are differently placed the distance between the two clusters becomes infinity. This proves that the Ward can be utilized for the clusters within range or present at the neighbouring regions. The adaptive features of the SOM can resolve this situation of infinity and provides effective clustering.

**Experimental Results:**

The results obtained are evaluated to study the performance of the SOM, NMTF, M-SOM clustering approaches. The performances of the three clustering algorithms are compared to find the better clustering approach. The clustering technique which has high inter-cluster distance and low intra-cluster distance is considered the most suitable algorithm to demarcate normal and deaf speech signal.

The Figure 1 depicts the implementation method for the deaf speech enhancement using the clustering algorithms. The input deaf speech signal is given to the system which extracts the speech parameters. The estimated values of speech are compared with the trained data set of normal and deaf speech samples. Further, using the SOM, NTMF and M-SOM clustering algorithms, similar data sets are grouped. The clustered mean waveform is then obtained with the normal and deaf speech samples. The differential correction measure may then be evaluated to get the enhanced waveform.

Figure 2 and 3 shows the speech waveform recorded for the normal and deaf speaker by uttering all the test words selected for work. The normal speech signal is represented in such a way it is definite in the signal space. From the figure it is clear that the normal speech signal is confined to the definite boundaries in which smaller changes are negotiable but the greater changes result in unclear speech signals. The deaf speech is a form of murmured speech which means it is not clear and spread along various regions of the signal space. The boundaries are not definite as the deaf speech signal varies with the change in frequencies.

The experiments are conducted to evaluate the clustering methods in terms of accuracy, precision, recall and f-measure. The SOM, NMTF and M-SOM clustering methods are compared to determine the efficient clustering approach.

**Accuracy:**

Accuracy is defined as the proportion of true positives and true negatives among the total number of features clustered. Accuracy is evaluated as,

$$\text{Accuracy} = \frac{(\text{True positive} + \text{True negative})}{(\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative})}$$

(18)

Figure 4 shows that when compared to the clustering algorithms SOM and NMTF the accuracy is improved for clustering using M-SOM. The accuracy of SOM clustering technique is 85.79%, NMTF offers accuracy of 80.72% whereas the clustering using M-SOM provides 88.84% accuracy.
Precision:

Precision value is evaluated according to the clustering level at true positive prediction, false positive.

\[
Precision = \frac{True\ Positive}{(True\ Positive + False\ Positive)}
\]  \hspace{1cm} (19)

The precision is improved using M-SOM clustering. When the methods are compared in terms of precision, the precision of clustering using SOM is 84.20, NMTF has 80.90 whereas the clustering using M-SOM has 88.56.

Recall:

Recall value is evaluated according to the clustering level at true positive prediction, false negative.

\[
Recall = \frac{True\ Positive}{(True\ positive + False\ negative)}
\]  \hspace{1cm} (20)

When the methods are compared in terms of recall, the recall of clustering using SOM is 84.79%, NMTF has 80.90% whereas the clustering using M-SOM has 89.16%.

F-measure:

The F-Measure computes average of clustering level precision and recall metrics

\[
F-Measure = \frac{2 \times precision \times recall}{precision + recall}
\]  \hspace{1cm} (21)

Fig. 1: Implementation diagram for enhancing the deaf speech signal using clustering technique

Fig. 2: Speech samples of Normal Speaker for the test words
When the methods are compared in terms of F-measure, the F-measure of clustering using SOM is 84.77%, NMTF has 80.77% whereas the clustering using Modified SOM has 88.68%. Table 1 shows the numerical comparison of the SOM, NMTF and M-SOM clustering approaches in terms of accuracy, precision, recall and F-measure.

From the above results it is clear that the M-SOM clustering approach has better clustering efficiency. The performance metrics given in Table 1 indicate that the M-SOM clustering algorithm provides accuracy of >87.33% for the words considered for study. The minimum is obtained for the word “Kathavu” (87.33%) and the maximum is attained for the word “Veedu”. In case of precision, M-SOM offers the least value of 87.12% for the word “Maram” and highest is noticed for the word “Kadhu” (89.99%). For recall, M-SOM clustering provides 90.56% for the word “Paal” and 87.33% for the word “Kathavu”, which are maximum and minimum respectively. When F-measure is considered, the word “Mani” takes the highest value of 90.16% and the word “Paal” claims the lowest of 87.14%. Upon considering all the words, the performance of the clustering algorithms can be rated sequentially in the following order, M-SOM, SOM and NMTF. This confirms that the M-SOM clustering algorithm has the possibility assisting speech enhancement progressively when compared to other methods.

The overall feature characteristics that vary among the normal and deaf speakers that are measured by using the clustering methods are applied to find the maximum distance variations among the two groups. When the deaf speaker word is given to the algorithm, it finds the best average distance. Based on the distance factor, the M-SOM clustering method is used to produce the clustered mean waveform. The correlation between the speech utterances of deaf and normal speakers for finding the distance factors may be explored in the future work.

**Conclusions:**

The deaf speech recognition is a much-needed solution to aid the deaf speakers and hard of hearing people. In order to improve the deaf speech enhancement process, the speech data’s are clustered based on the extracted features. For efficient clustering, three methods namely NMTF, SOM and M-SOM clustering approaches are applied. The suggested M-SOM clustering improves the clustering process of deaf speech data and provides improved performance compared to other clustering approaches. The effect of the technique is measured using the following parameters namely accuracy, precision, recall and F-measure. The overall performance is >86% for M-SOM where as the results are 81.86% and 84.56% for NMTF and SOM clustering. The inference made is that the M-SOM clustering approach can be used to find the best clustering with the features extracted. This efficient clustering of speech data provides suitable enhancement of the deaf speech which can further be used to derive different corrective measures. The obtained enhanced speech can then be applied as a voice aid which can be used by normal persons to understand the deaf speech.
Fig. 4: Accuracy comparison between clustering algorithms

Table 1: Analysis of performance metrics between the clustering algorithms

<table>
<thead>
<tr>
<th>Test words</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SOM</td>
<td>NMTF</td>
<td>Modified SOM</td>
<td>SOM</td>
</tr>
<tr>
<td>Amma</td>
<td>85.63</td>
<td>81.06</td>
<td>87.69</td>
<td>82.69</td>
</tr>
<tr>
<td>Appa</td>
<td>86.22</td>
<td>80.58</td>
<td>88.53</td>
<td>83.12</td>
</tr>
<tr>
<td>Bhandhu</td>
<td>84.56</td>
<td>80.21</td>
<td>89.71</td>
<td>85.11</td>
</tr>
<tr>
<td>Kathavu</td>
<td>86.75</td>
<td>80.66</td>
<td>88.66</td>
<td>84.70</td>
</tr>
<tr>
<td>Kathu</td>
<td>84.90</td>
<td>81.00</td>
<td>87.33</td>
<td>81.95</td>
</tr>
<tr>
<td>Mani</td>
<td>87.91</td>
<td>80.66</td>
<td>90.00</td>
<td>88.00</td>
</tr>
<tr>
<td>Maram</td>
<td>84.27</td>
<td>81.33</td>
<td>88.60</td>
<td>84.08</td>
</tr>
<tr>
<td>Veedu</td>
<td>83.49</td>
<td>81.52</td>
<td>88.54</td>
<td>86.32</td>
</tr>
</tbody>
</table>
REFERENCES

13. Hua Wang, FeipingNie, and Heng Huang, Chris Ding, 2011.‘Nonnegative Matrix Tri-Factorization Based High-Order Co-Clustering and Its Fast Implementation’, 11th IEEE International Conference on Data Mining.