

AUTOMATED SPEECH RECOGNITION SYSTEM WITH NOISE CONSIDERATION

Mrs. S. Premalatha¹

Dr. T. Kesavamurthy²

¹Department of Electronics and Communication Engineering, KSR Institute for Engineering and Technology, Tiruchengode -637215, India, pramalathaksr2@yahoo.com

²Department of Electronics and Communication Engineering, PSG College of technology, Coimbatore-641004, India, tkm71in@yahoo.co.in

Abstract: In the real world environment, Speech recognition is one of the most progressing research area. The quality as well as performance of automated speech recognition is disturbed by noises exist in the speech signals. Noises are inevitable in the speech that is transferred via external medium containing noise. In the previous research, it is resolved with the help of Multivariate Autoregressive Spectrogram Modeling (MARS-M). On the other hand, in the previous method, noise reduction in the speech recognition is carried out by means of taking the greater energy coefficients as well as the signal with greater correlation. By concentrating on these, it is presumed that the noise existence is evaded significantly. In the research system, it is solved by means of presenting the new technique known as Background Noise concerned Automated Speech Recognition System (BN-ASR). In the presented system, Noise reduction is carried out with the help of the Normalized data nonlinearity (NDN)-LMS adaptation technique. This technique could adaptively remove the noises exist in the signals. Subsequent to noise reduction feature extraction is carried out with the help of the technique called Synchrony-Based Feature Extraction that will foresee the averaged localized synchrony response of the noise filter. At last, for the precise recognition of speech signals, Hybrid Particle Swarm Optimization-Artificial Neural Network (HPSO-ANN) is introduced. In the matlab simulation environment, the complete implementation of the research method is performed and it is clear that the research technique results in providing the best possible outcome compared to the previous research techniques.

Keywords: Speech recognition, noisy features, normalized data, synchrony response, Noise filter

1. Introduction

Automatic speech recognition (ASR) is the self-governing, computer-driven transcription of spoken language into readable text in real time[1]. ASR lets a computer to find out the words, which a person speaks into a microphone or telephone and transform it into written text[2]. The objective of an ASR system is to exactly as well as proficiently transform a speech signal into a text message transcription of the spoken words self-reliant of the speaker, environment or the device utilized to record the speech (that is to say the microphone)[3].

An ASR system focuses on inferring those original words provided the observable signal [4]. The

most general as well as finest technique is the probabilistic method [5]. A speech signal relates to any word (or sequence of words) in the vocabulary containing a certain probability [6]. So, presuming a word x or word sequence X was spoken, we calculate a score for comparing these words with the speech signal. This score is computed from the acoustic characteristics of speech sub-units (phonemes in the acoustic model), linguistic knowledge regarding which words could follow which other words [7]. Comprising extra knowledge as the pronunciation score presented in this research has as well shown to be useful [8]. Lastly, we sort the probable word sequence hypotheses by score, and elect to choose the hypothesis with the greatest score as recognition outcome [9].

ASR is a central engineering research topic for epochs [10]. In spite of the constant technological advances, particularly the rise of HMM-based speech recognition systems as well as the countless amount of enhancements to the HMM-methodology, the modern speech recognition systems fall till now far behindhand human performance in recognition abilities[11]. This is particularly correct for noisy conditions, that is to say, situations with semi-stationary as well as transitory background sounds combined with the input speech being recognized [12]. Humans could carry out source separation with comparative easiness, and they could enhance their recognition performance by means of using multimodal- and contextual cues to fill gaps in input stream with most appropriate elucidation for missing data while the linguistic content of the input itself is vague [13].

HMM-based recognizers need a match amid the features taken out from the training data (state models) as well as the features acquired from the input signals in the recognition [14]. As this perfect match is hardly a sensible supposition in real life applications, diverse kinds of techniques were designed to handle noise in speech recognition processing. This proposed method reviews one family of such noise robust speech recognition techniques, called missing-feature techniques, which try to handle partly corrupted spectral input [15]. As the HMM

recognizers are the present advanced technique in speech recognition, the missing feature techniques in this review all depend upon HMM-based recognition systems if not stated then.

In the presented technique, Noise reduction is carried out by means of utilizing the Normalized data nonlinearity (NDN)-LMS adaptation technique that could adaptively remove the noises in the signals. Subsequent to noise reduction, feature extraction is carried out with the help of the technique called Synchrony-Based Feature Extraction that will find out the averaged localized synchrony response of the noise filter.

The complete organization of the presented technique is provided in this manner: In this part, thorough explanation regarding the part of automatic speech recognition is provided. In part 2, conversation of numerous previous research techniques, which try to carry out the noisy reduction as well as precise speech recognition is conversed. In part 3, thorough conversation regarding the presented technique with the appropriate samples as well as diagram is provided. In part 4, results as well as discussion of the presented technique got from the Matlab simulation is given. At last in part 5, complete summary of the presented technique based outcomes attained is conversed thoroughly.

2. Related Works

[16] Proposed speech recognition with the help of Linear Predictive Coding (LPC) as well as Artificial Neural Network (ANN) for directing movement of mobile robot. Input signals were sampled unswervingly from the microphone and after that the removal was carried out by means of LPC and ANN. [17] presented speaker independent secluded speech recognition system for Tamil language. Feature extraction, acoustic model, pronunciation dictionary as well as language model were developed with the help of HMM that generated 88% of accurateness in 2500 words.

[18] Identified development as well as assessment of diverse acoustic models for Malayalam continuous speech recognition. In this proposed method, HMM is utilized to match up as well as assess the Context Dependent (CD), Context Independent (CI) models and Context Dependent tied (CD tied) models from this CI model 21%. The database comprises 21 speakers with 10 males as well as 11 females. [19] presented an effectual speech recognition system that was tested with Mel Frequency Cepstrum Coefficients (MFCC), Vector Quantization (VQ), HMM that identify the speech by 98% accurateness. The database comprises five words spoken by 4 speakers at ten times.

[20] Introduced an automatic speech recognition system for secluded as well as connected words of Hindi language

with the help of Hidden Markov Model Toolkit (HTK). Hindi words are utilized for dataset taken out by MFCC as well as the recognition system attained 95% accurateness in secluded words and 90% in connected words. [21] presented Hindi automatic speech recognition by means of HTK. Secluded words are utilized to identify the speech with 10 states in HMM topology that generated 96.61%.

[22] Introduced automatic speech recognition method for Bangla words. Feature extraction was carried out by, Linear Predictive Coding (LPC) and Gaussian Mixture Model (GMM). Over-all, there are 100 words recorded in 1000 times that produced 84% accurateness. [23] Presented Malayalam word identification for speech recognition system. The research method was carried out with syllable based segmentation by means of utilizing HMM on MFCC for feature extraction.

[24] Presented Sanskrit speech recognition by means of utilizing HTK. MFCC and two state of HMM were utilized for extraction that generates 95.2% to 97.2% accurateness correspondingly. [25] Presented real time speaker recognition system for Hindi words. Feature extraction carried out with MFCC with the help of Quantization Linde, Buzo and Gray (VQLBG) algorithm. In order to eliminate the silence, Voice Activity Detector (VAC) was presented.

2. Related Works

In the research technique, Noise reduction is carried out with the help of the Normalized data nonlinearity (NDN)-LMS adaptation technique that could remove the noises in the signals. Subsequent to noise reduction feature extraction is carried out with the help of Synchrony-Based Feature Extraction that will identify the averaged localized synchrony response of the noise filter.

3.1. Noise Reduction Using Nsn-Lms

In this proposed method, we present a new least-mean-square (LMS) algorithm to filter/speech sounds in the adaptive noise cancellation (ANC) problem. It is dependent upon the minimization of the squared Euclidean norm of the difference weight vector under stability restraint well-defined over the a posteriori estimation error. So, the Lagrangian methodology is utilized with the intention of presenting a nonlinear adaptation rule stated in regard to the product of differential inputs as well as errors that signifies a generalization of the normalized (N) LMS algorithm, which is observed as the solution to a restrained optimization problem. The problem of interest is defined in this manner: provided the tap-input vector $x(n)$ as well as the desired response $d(n)$, identify the tap weight vector $w(n+1)$ with the purpose of reducing the squared Euclidean norm of the change $\|w(n+1) - w(n)\|$ in the tap-weight vector $w(n+1)$ regarding its old value $w(n)$,

relate to the restraint $w(n+1)^H x(n) = d(n)$, here H signifies the Hermitian transpose. This restraint signifies that the a posteriori error sequence vanishes $[e^{[k+1]}(n) = d(n) - w(k+1)^H x(n) = 0, \text{ for } k=1:n]$.

With the intention of resolving this optimization problem, the technique of Lagrange multipliers is utilized with the Lagrangian function as given in Eq. (1)

$$L(w(n+1)) = \|\xi w(n+1)\|^2 + \text{Re}[\lambda^* e^{[n+1]}(n)] \quad (1)$$

Here λ^* is known as the Lagrange multiplier, therefore getting the famous adaptation rule in (1) with the normalized step size provided by $\mu = \hat{\mu} / \|x(n)\|^2$. The latter restraint is excessively obstructive in real applications; therefore, when we relax it, additional interesting solution is derived. Take the restrained optimization problem, which gives the subsequent cost function as given in Eq. (2):

$$L(w(n+1)) = \|\xi w(n+1)\|^2 + \text{Re}[\lambda^* \xi e^{[n+1]}(n)] \quad (2)$$

Here $\xi e^{[n+1]}(n) = e^{[n+1]}(n) - e^{[n+1]}(n-1)$. This equilibrium constraint guarantees stability in the series of a posteriori errors, that is to say the best possible solution $w^{opt}(n+1)$ is the one, which renders the series of errors as smooth as likely. Considering the partial derivative of (2) regarding the vector $w^H(n+1)$ and setting it equivalent to zero results in Eq.(3)

$$\frac{\partial L(w)(n+1)}{\partial w^H(n+1)} = \frac{\partial \delta w^H(n+1) \delta w(n+1)}{\partial w^H(n+1)} \quad (3)$$

Since $e^{[n+1]}(k) = d(k) - w^H(n+1)x(k)$ for $k=n, n-1$ and $\text{Re}[z] = \frac{1}{2}(z+z^*)$, then

$$\frac{\partial L(w)(n+1)}{\partial w^H(n+1)} = \delta w(n+1) - \frac{1}{2} \lambda^* \delta x \quad (4)$$

Here $\delta x(n) = x(n) - x(n-1)$ is known as the difference amid two successive input vectors. Therefore, the step of the algorithm is given in Eq. (5)

$$\delta w(n+1) = \frac{1}{2} \lambda^* \delta x(n) \Rightarrow w(n+1) = w(n) + \frac{1}{2} \lambda^* \delta x \quad (5)$$

Lastly, subsequently multiplying both sides of (5) by $\delta x^H(n)$, the Lagrange multiplier is stated as in Eq. (6)

$$\lambda^* = \frac{2 \delta x^H(n) \delta w(n+1)}{\|\delta x(n)\|^2} = \frac{-2(\delta e^{[n+1]}(n) - \delta e^{[n]}(n))}{\|\delta x(n)\|^2} \quad (6)$$

Here $\delta e^{[n]}(n) = e^{[n]}(n) - e^{[n]}(n-1)$ is known as the difference in the a priori error sequence [denoted by $\delta e(n)$ for short], as the numerator on the left-hand side of (6) is equivalent to $x^H(n) w(n+1) - x^H(n-1) w(n+1) - x^H(n) w(n) + x^H(n-1) w(n)$. So, employing the equilibrium restraint on the right-hand side of (7) $\delta e^{[n]}(n) = 0$ results in Eq. (7)

$$\lambda = \frac{2 \delta e^{[n]}(n)}{\|\delta x(n)\|^2} \quad (7)$$

Lastly, the minimum of the Lagrangian function fulfills the subsequent restrained stability update condition (CS-LMS) which is given in Eq. (8)

$$w(n+1) = w(n) + \frac{\delta x(n) \delta e^{[n]}(n)}{\|\delta x(n)\|^2} \quad (8)$$

The weight adaptation rule is made more vigorous by means of presenting a small positive ϵ constant into the denominator to thwart numerical uncertainties in case of a vanishingly small squared norm $\|\delta x(n)\|^2$ and by means of multiplying the weight increment by a constant step size with the aim of controlling the speed of the adaptation. The equilibrium imposes the convergence of the algorithm when $\|\delta x(n)\|^2 \neq 0$. Numerous learning techniques, in which the learning depends upon the simultaneous modification of processing variables, were presented in the past for decorrelation, blind source separation, or deconvolution applications. Stochastic information gradient (SIG) algorithms increase (or decrease) the Shannon's entropy of the series of errors by utilizing an estimator dependent upon an instantaneous value of the probability density function (Pdf) as well as Parzen windowing. Thus, the CS-LMS algorithm is taken as a generalization of the single sample-based SIG algorithm with the help of variable kernel density estimators.

Actually, it is general to utilize ensemble-average learning curves with the aim of studying the statistical performance of adaptive filters. The derivation of these curves is somewhat dissimilar for the ANC problem because of the existence of the anticipated clean signal $s(n)$. By means of utilizing the definition of the weight-error vector $\epsilon(n) = w_0 - w(n)$ and Eq. (9) with the step size described as μ , we might state the evolution of $\epsilon(n)$ as

$$\epsilon(n+1) = \epsilon(n) - \mu \delta x(n) \times (\delta s(n) + \delta v(n) - (w_0 - \epsilon(n))) \quad (9)$$

Here $\delta[\cdot](n) = [\cdot](n) - [\cdot](n-1)$ and $v(n)$ represents the noise in the primary signal $d(n)$. When $v(n)$ is presumed to be produced by the multiple regression model: $v(n) = w_0^H x(n) + e_0(n)$, the weight-error vector is stated as in Eq. (10)

$$\epsilon(n+1) = (I - \mu \delta x(n) \delta x(n)^H) \epsilon(n) \quad (10)$$

By means of invoking the direct-averaging technique [11], the equation above results in Eq. (11)

$$\epsilon_0(n+1) = (I - \mu R_{\delta x}) \epsilon_0(n) - \mu \delta x(n) \delta \hat{e}_0^{[n]}(n) \quad (11)$$

Here $\delta \hat{e}_0^{[n]}(n) = \delta e_0(n) + \delta s(n)$, and the mean-squared error created by the filter is provided by eq. (12)

$$J(n) = J_0 + E[|s(n)|^2] + E[\varepsilon_0^H(n)x(n)x(n)^H\varepsilon_0(n)] \quad (12)$$

Here $J_0 = E[|e_0(n)|^2]$ and $J_{\min} = J_0 + E[|s(n)|^2]$

The stochastic evolution on the natural modes is analyzed by transmuting Eq. (12) into Eq. (13)

$$v(n+1) = (I - \mu \Lambda)v(n) - \phi(n) \quad (13)$$

by means of using the unitary similarity transformation to the correlation matrix $R_{\delta x}$, here $\Lambda = Q^H R_{\delta x} Q$ is a diagonal matrix encompasses the eigenvalues λ_k of $R_{\delta x}$, is a unitary matrix whose columns create an orthogonal set of eigenvectors and the stochastic force vector is stated as $\Phi(n) = \mu Q^H \delta x(n) \delta \hat{e}_0^i(n)$. This vector contains the subsequent properties.

- The mean of the stochastic force vector is zero: $E[\Phi(n)] = 0$
- The correlation matrix of the stochastic force vector is a diagonal matrix: $E[\Phi(n)\Phi^H(n)] = \mu^2 \hat{J} \Lambda$, here $\hat{J} = 2 \left(E[|e_0(n)|^2] + E[|s(n)|^2] - \Re\{r_s(1)\} \right)$, and $r_s(1) = E[s^*(n+1)s(n)]$.

The primary two moments of the natural modes $v(n)$ is acquired by means of utilizing these properties that let one to show the evolution of $J(n)$ with time step n . The third term of Eq. (13), in light of the direct-averaging technique, is equivalent to Eq. (14)

$$J_{ex}(n) = \frac{1}{2} \sum_{k=1}^L \lambda_k E[|v_k(n)|^2] + E[tr\{v^H Q^H \Re\{R(1) Q v\}\}] \quad (14)$$

Here $R(1) = E[x(n+1)x^H(n)]$. Presuming that the input signal is weakly correlated ($R(1) \sim 0$), the second term is bounded in the last equality of Eq. (14) with the primary term (natural evolution), that is to say

$E[tr\{v^H Q^H \Re\{R(1) Q v\}\}] \leq (1/2) \sum_{k=1}^L \lambda_k E[|v_k(n)|^2]$, and after that it is derived as in Eq. (15)

$$J_{ex}(n) \leq \sum_{k=1}^L \lambda_k E[|v_k(n)|^2] = \sum_{k=1}^L \lambda_k \left(\frac{\mu \hat{J}}{2 - \mu \lambda_k} + (1 - \mu \lambda_k) \right) \quad (15)$$

Here $v_k(n)$ represents the k th component of natural mode $v(n)$. When the exponential factor is ignored with raising n which is depicted in Eq. (16)

$$J_{ex}(\infty) \leq \sum_{k=1}^L \lambda_k \left(\frac{\mu \hat{J}}{2 - \mu \lambda_k} \right) \approx \frac{1}{2} \mu \hat{J} tr\{R_{\delta x}\} \quad (16)$$

The reduction in $J_{ex}(\infty)$ is attained every time is shown in Eq. (17)

$$J_{ex}(\infty) \approx \frac{1}{2} \mu \hat{J} tr\{R_{\delta x}\} \approx \mu \hat{J} tr\{R\} \geq \frac{3}{4} J_{\min} \quad (17)$$

That is to say the anticipated signal is strongly correlated.

It as well follows from classical analysis that 1) the greater value of μ poises the trade-off amid $J_{ex}(\infty)$ and the average time constant as given in Eq. (18)

$$\tau \approx \frac{L}{\mu tr\{R_{\delta x}\}} \quad (18)$$

Here L is known as the filter length, and 2) a necessary condition for stability is that $0 < \mu < 2/\lambda_k$, for all k .

3.2. Synchrony Based Feature Extraction

We have discovered numerous means of taking out synchrony info for speech recognition. The processing in this proposed method tries to take out synchrony with the intension that reproduces the frequency content of the real signal, more willingly than just the center frequency of every analysis channel. Primarily, we pass the outcome of every channel of the auditory model via a second bandpass filter with the similar frequency response as the auditory filter for that channel with the aim of decreasing the harmonic distortion presented by nonlinearities in the peripheral auditory processing.

The short-time Fourier transform of the outcomes of the bandpass filters is calculated, and these frequency responses are averaged crosswise channels. This generates a high resolution spectral representation at low frequencies for which the auditory nerve is synchronized to the input equal to approximately 2.2 kHz, and that comprises the consequences of all of the nonlinearities of the peripheral processing. We eliminate the horizontal striations characteristically seen in narrowband spectrograms (that reproduce the pitch of the incoming signal) by means of employing a discrete-cosine transform (DCT) to the frequency response, using a short-time "lifter" to the inverse transform, and after that returning to the frequency domain by means of utilizing an inverse DCT.

The features utilized for speech recognition are designed by means of uniting the synchrony outcomes at low frequencies with the mean rate outputs at greater frequencies. Initially, the synchrony outcomes that arise primarily as a linear function of frequency are changed alongside the frequency axis with the intension that they when the mean rate outcomes are warped alongside the frequency axis with the intension of relating to the nonlinear dependence on frequency of the center frequencies of the analysis channels utilized to design the mean rate outcomes.

Components from the frequency-warped synchrony outcomes amid around 0 and 2.5 kHz and components from the mean rate outcomes amid around 2.2 to 8 kHz are

conserved for additional processing. These outcomes are related to a final DCT that creates a group of coefficients, which are alike to a certain degree to cepstral coefficients. The current implementation utilizes the first 8 DCT coefficients derived from the synchrony outcomes and the primary 5 DCT coefficients derived from the mean rate outcomes. These are concatenated into a vector of 13 coefficients that function as the basic input to the speech recognition system. Delta as well as delta-delta coefficients are got in the similar style as is usually accomplished for the CMU SPHINX system.

3.3. Speech Recognition Using Improved Artificial Neural Network

A Hybrid Particle Swarm Optimization-Artificial Neural Network (HPSO-ANN) is presented in the research method with the intention of enhancing the speech recognition outcomes of the entire system. HPSO-ANN classifier encompasses two main steps learning as well as testing. Figure 1 elucidates the HPSO-ANN architecture with PSO algorithm.

In learning stage, the best possible size for the speech dataset such as the entire amount of layers, the amount of hidden units in the middle layers, and amount of units in the input as well as output layers in regard to accurateness on a test set, and the training algorithm were imposed, it would run the extraction process from a noisy training set, with the aim of attaining the better speech improvement. The objective of learning is to reduce the cost function dependent upon the error signal $e_i(t)$, in regard to parameters (weights), as a result, the real response of all the output neuron in the network techniques the target response. The info of the HPSO-ANN classifier was elucidated in the current work.

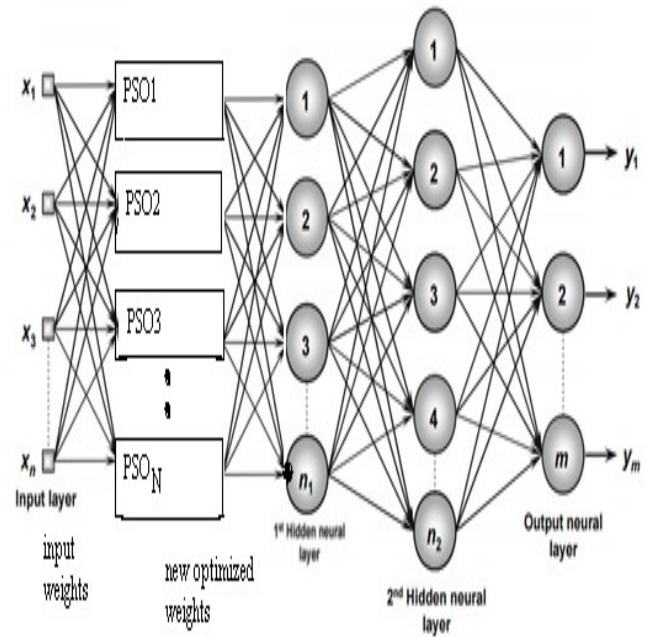


Figure 6. Improved Artificial Neural Network (IANN)

4. Results And Discussion:

The datasets were collected from regional Tamil news. The audio samples range is about range of 100-150 samples and from diverse news channels, it is presumed. The samples are in the formats of '.wav' format as well as the time duration is 2 minutes for each sample. The presented Background Noise concerned Automated Speech Recognition System (BN-ASR) algorithm, correctly get the speech recognition. The performance metrics were taken as sensitivity, specificity, precision, accuracy, recall and f-measure metrics were calculated by previous techniques for instance multivariate AR modeling (MAR) and presented BN-ASR algorithm.

Performance metrics	MA R	BN-ASR
Precision	71.43	93.3
Recall	100	100
F-measure	85.7	96.6
Accuracy	92	96

Table .1. Comparison of classifier algorithms

Precision: It is the proportion of the true positives contrary to true positives as well as false positives result for imposition and real features. It is given below

$$\text{Precision}(P) = \frac{T_p}{T_p + F_p} \quad (19)$$

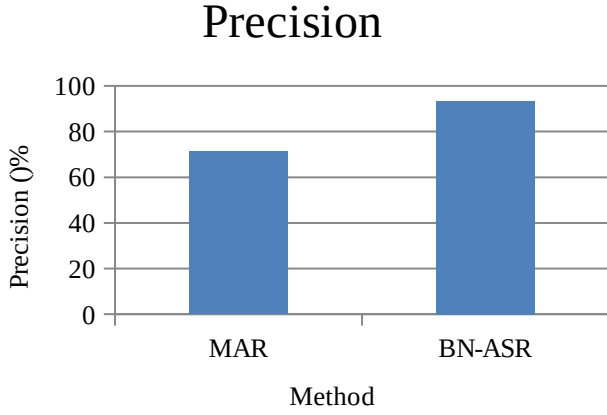


Figure. 2. Precision comparison

According to the Figure 2, it is clear that the presented BN-ASR system contains greater precision while compared with other techniques.

Recall: It is calculated on the root of the data retrieval at true positive forecast, false negative. Commonly, it is identified in this manner,

$$\text{Recall}(R) = \frac{T_p}{T_p + F_n} \quad (20)$$

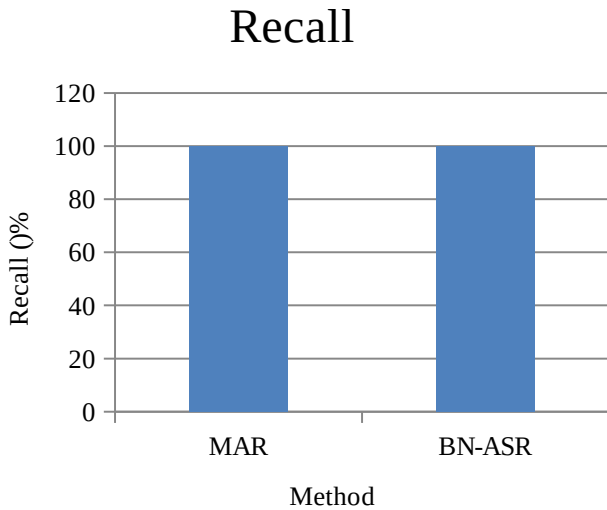


Figure. 3. Recall comparison

According to Figure 3, it is clear that the presented BN-ASR system contains greater recall. So, the

presented FW-VCVQ-NOE algorithm is superior to the existing MAR algorithms.

F-measure: It is a degree of an accurateness of the test. It presumes the precision p as well as the recall r of the test to quantify the score.

$$\text{F-measure} = 2 \cdot \frac{P \cdot R}{P + R} \quad (21)$$

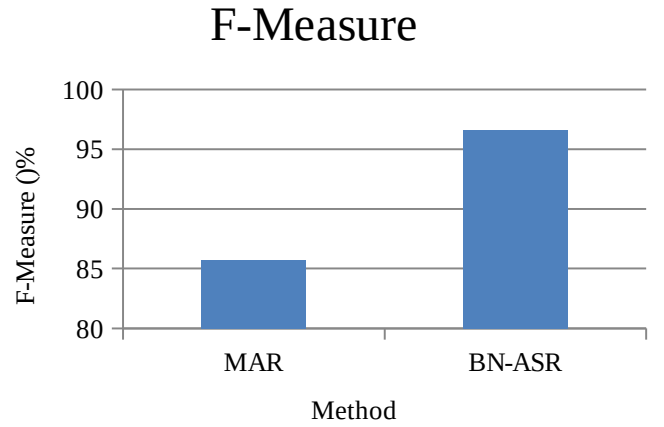


Figure.4. F-measure comparison

According to Figure 4, it is clear that the presented BN-ASR system contains greater f-measure. So, the presented BN-ASR algorithm is superior to the previous MAR algorithms.

Accuracy: It is stated as the complete correctness of the model and is calculated as the over-all actual classification parameters ($T_p + T_n$) that is categorized by the summation of the classification parameters ($T_p + T_n + F_p + F_n$). The accuracy is computed in this manner:

$$\text{Accuracy} = \frac{T_p + T_n}{(T_p + T_n + F_p + F_n)} \quad (22)$$

Here T_p is the number of accurate predictions that an instance is negative, T_n is the number of inappropriate predictions that an instance is positive, F_p is the number of inappropriate of predictions that an instance negative, and F_n is called the number of accurate predictions that an instance is positive.

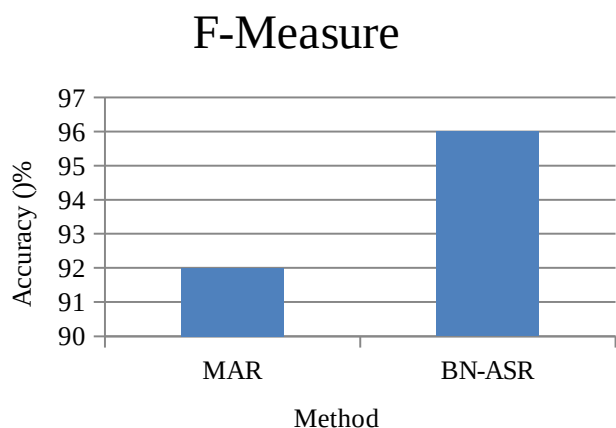


Figure.5. Accuracy comparison

According to Figure 5, it is clear that the presented BN-ASR system contains greater accuracy

5. Conclusion

Automated speech recognition acts as a significant role in the real world environment that will be affected by numerous aspects. One among the foremost aspect that has emotional impact on the speech recognition performance is existence of noises in the speech signals. So, the noise related speech recognition turn out to be the most widespread research method in the real world that must be concentrated more. In the presented research technique, Noise reduction is carried out by means of utilizing the Normalized data nonlinearity (NDN)-LMS adaptation technique. This technique could adaptively remove the noises exist in the signals. Subsequent to noise reduction, feature extraction is carried out by means of utilizing the technique known as Synchrony-Based Feature Extraction that would foresee the averaged localized synchrony response of the noise filter. The complete implementation of the presented technique is carried out in the matlab simulation environment and it is predicted that the presented research technique results in providing the best possible outcome compared to the previous research techniques.

References

1. Yu, D., & Deng, L. (2016). Automatic Speech Recognition. SPRINGER LONDON Limited.
2. Ranjeet, P., Prakash, T., Amruta, S., & Monali, S. (2016). Automatic Speech Recognition System. Imperial Journal of Interdisciplinary Research, 2(3), 165-169.
3. Erdogan, H., Hershey, J., Watanabe, S., & Le Roux, J. (2016). U.S. Patent Application No. 14/620,514.
4. Raut, P. C., & Deoghare, S. U. (2016). Automatic Speech Recognition and its Applications.
5. Hofmann, T. (2017, August). Probabilistic latent semantic indexing. In ACM SIGIR Forum (Vol. 51, No. 2, pp. 211-218). ACM.
6. Deligne, S., Gopinath, R. A., Kanevsky, D., & Viswanathan, M. (2017). U.S. Patent No. 9,754,586. Washington, DC: U.S. Patent and Trademark Office.
7. Kamper, H., Jansen, A., & Goldwater, S. (2016). Unsupervised word segmentation and lexicon discovery using acoustic word embeddings. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 24(4), 669-679.
8. Komissarchik, J., & Komissarchik, E. (2018). U.S. Patent Application No. 15/641,298.
9. Hofer, J., & Stemmer, G. (2016). U.S. Patent No. 9,530,404. Washington, DC: U.S. Patent and Trademark Office.
10. Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., ... & Truong, K. P. (2016). The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE Transactions on Affective Computing, 7(2), 190-202.
11. Kumar, P., Gauba, H., Roy, P. P., & Dogra, D. P. (2017). Coupled HMM-based multi-sensor data fusion for sign language recognition. Pattern Recognition Letters, 86, 1-8.
12. Chen, Z. (2017). Single Channel auditory source separation with neural network. Columbia University.
13. Barker, J., Marxer, R., Vincent, E., & Watanabe, S. (2017). The third 'CHiME' speech separation and recognition challenge: Analysis and outcomes. Computer Speech & Language, 46, 605-626.
14. Najafian, M., & Hansen, J. H. (2017, March). Environment aware speaker diarization for moving targets using parallel DNN-based recognizers.

- In Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on (pp. 5450-5454). IEEE.
15. Williams, C. E. (2017). Toward a Decolonial Homiletic: In Search of a Local Preaching Voice for Trinidad and Tobago (Doctoral dissertation, Princeton Theological Seminary).
 16. Wijoyo, S., & Wijoyo, S. (2011). Speech recognition using linear predictive coding and artificial neural network for controlling movement of mobile robot. In Proceedings of 2011 International Conference on Information and Electronics Engineering (ICIEE 2011) (pp. 28-29).
 17. Radha, V. (2012). Speaker independent isolated speech recognition system for Tamil language using HMM. *Procedia Engineering*, 30, 1097-1102.
 18. Kurian, C., & Balakrishnan, K. (2012). Development & evaluation of different acoustic models for Malayalam continuous speech recognition. *Procedia Engineering*, 30, 1081-1088.
 19. Swamy, S., & Ramakrishnan, K. V. (2013). An efficient speech recognition system. *Computer Science & Engineering*, 3(4), 21.
 20. Choudhary, A., Chauhan, M. R., & Gupta, M. G. (2013). Automatic Speech Recognition System for Isolated and Connected Words of Hindi Language By Using Hidden Markov Model Toolkit (HTK). In Proceedings of International Conference on Emerging Trends in Engineering and Technology (pp. 244-252).
 21. Saini, P., Kaur, P., & Dua, M. (2013). Hindi automatic speech recognition using htk. *International Journal of Engineering Trends and Technology (IJETT)*, 4(6), 2223-2229.
 22. Sampath Kumar, A & Vivekanandan, P 2018, 'Gene Selection Using Multiple Queen Colonies In Large Scale Machine Learning', *Journal of Electrical Engineering*, vol. 18, no. 4, pp. 97-111. ISSN: 1582-4594
 23. Ali, M. A., Hossain, M., & Bhuiyan, M. N. (2013). Automatic speech recognition technique for Bangla words. *International Journal of Advanced Science and Technology*, 50.
 24. Moneykumar, M., Sherly, E., & Varghese, W. S. (2014). Malayalam Word Identification For Speech Recognition System. *An International Journal of Engineering Sciences*, 22-26.
 25. Pokhariya, J. S., & Mathur, S. (2014). Sanskrit Speech Recognition using Hidden Markov Model Toolkit. *International Journal of Engineering Research & technology*, 3, 93-98.
 26. Nijhawan, G., & Soni, M. K. (2014). Real Time Speaker Recognition System for Hindi Words. *International Journal of Information Engineering and Electronic Business*, 6(2), 35.