ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

Pattern Recognition of Speech Signals using Curvelet transform and Artificial Intelligence

[1] Shaik Subhani
Associate professor, Dept. of IT, SNIST, Hyderabad

Abstract: Pattern recognition of speech signals is ability to translate a spoken word to text format. This paper presents an authority speech recognition system based on curvelet transform and artificial neural network techniques to enhance the recognition rate. This research comprised in two distinct phases, a feature extractor and recognizer is presented. In feature extraction phase, curvelet transform extract the features from the input speech signal and detail components of these signals which assist in achieving higher recognition rate. For feature matching, artificial neural networks is used as classifiers. The performance evaluation has been described in terms of accurate recognition rate, interfering sounds, hit rates, false alarm and miss rates. The rate of accurate classification was about 95.3 % for the sample speech signals.

Index Terms— Pattern Recognition, Speech recognition, Curvelet transforms, Feature extraction, Artificial intelligence

I. INTRODUCTION

Speech is the most efficient mode of communication between peoples. This, being the best way of communication, could also be a useful interface to communicate with machines. Past few decades, developers have analysis speech for a extensive of applications ranging from mobile transportation to automatic evaluation machines. Speech recognition reduces the transparency caused by exchange of communication techniques. Speech signal communication not used a large amount in the field of computers and electronics because of the complexity. However, with current methods and algorithms can process the speech signals simply and identify the text. There are dissimilar ways to speech recognition in Artificial Neural Networks, Hidden Markov Model (HMM), Vector Quantization (VQ), support vector machine etc. This paper presents an authority speech recognition system based on discrete curvelet Transform (DCT) and ANN methods to enhance the identification rate. Speech recognition is a charming appliance of digital signal processing in the real-world applications. It is still a rising area and carries tough potential in the near future as computing power enhance. Speech recognition process needs deep processing due to huge samples per window. The growth of techniques to signal processing in the lack of models lead to queries for the processes of making signals using Artificial neural networks.These techniques recognize stationary signals within a given time and lack of capability to process localized proceedings correctly. Curvelet analysis has been confirmed as efficient signal computing techniques for a different signal processing issues. Many different methods, algorithms, and mathematical models developed to help speech analysis and speech recognition. This section points out advances and techniques that have been and are being applied to the speech recognition process. Thiang, et al. offered speech identification using Linear Predictive Coding and Artificial Neural Network for domineering movement of mobile robot [6]. Akkas Ali et al. described automatic speech recognition method for Bangla words. He can be feature extraction was done by LPC and Gaussian Mixture Model. Hindered words recorded in thousand times which gave 84% accuracy [8]. Ms.Vimala et al. proposed speaker independent isolated speech detection system for Tamil language. Multiple designs like Feature extraction, acoustic model and pronunciation dictionary were applied using HMM which produced 88% of accuracy in 2500 words [9]. Cini Kurian et al. developed diverse acoustic models for Malayalam continuous speech recognition. HMM is used to compare and assess the Context Dependent, Context Independent models and Context Dependent models from this Context Independent model gain 21% [7]. Hemdal & Hughes et al. took the basis of finding speech sounds and providing labels their exist a flat number of distinctive phonetic units in spoken language which are generally characterized by a set of acoustics characteristics differ with respect to time in speech signal [10]. Suma Swamy et al. proposed an efficient speech recognition mechanism experimented with Mel Frequency Cepstrum Coefficients (MFCC), Vector Quantization (VQ) and HMM, she recognize 98% recognize accuracy [11]. The Research article is organized as follow: Section I provides speech signals with background work. Section II states the Mathematical Formulation of speech recognition. Section III presents our research with classification network. In Section IV, compares the performance of our optimized systems against many other systems. Finally, Section V describes

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

the conclusions of this paper and suggestions for future work.

II. MATHEMATICAL FORMULATION OF SPEECH RECOGNITION

Speech recognition is a multi level pattern recognition process; acoustical signals are tested and prepared into an order of sub word phrases, units, sentences and words. The continuous speech waveform is initially separated into frames with stable span [2]. Each frame gives features represented by discrete parameter vectors; assume that duration of a single vector, i.e. one frame of speech curve form can be considered as stationary. This is not severely accurate but it is approximation of speech recognition. For each spoken word, let O = o1o2……oτ be a sequence of parameters vectors, where ot is at time and s∈ {1, … … . . , τ}. Given a dictionary C with words yi ∈ C, the recognition problem is summarized by

$\tilde{W} = \arg^{max}_{w} P(Y|J)$ (1)

Where $\tilde{Y}$ is the recognized word, P is the probability measure and Y = y1 ……yk a word sequence. Bayes' Rule permits to transform (1) in a suitable calculable form:

$P(Y|J) = \frac{P(J|Y)P(Y)}{P(O)}$ (2)

Where P(O|W) represents the acoustic model and P(Y) the language model; P(J) can be unheeded. The set of prior probabilities P(Y), the majority of spoken word probably based on the likelihood P(J|yi). The combination of the acoustic probability and language probability model is weighted. So the language model is scaled by an empirically represented constant s, called language model scale factor (LMSF). LMSF is represented empirically to get best performance on recognition. This weighting has a side effect as a penalty for inserting new words. We add a scaling factor p word insertions called word insertion penalty (WIP) also calculated empirically. Thus equation one becomes

$\tilde{Y} = \arg^{max}_{w} P(Y|J)\ P(Y)t|Y|^{p}$ (3)

In the log domain, the total likelihood is calculated as $\log_{|y|}\ \tilde{Y} = \log_{|y|}\ P\ (J|Y) + t\ \log_{|Y|}\ P(Y) + p$ (4)

Where |Y|the length of the word sequence Y, t is the language model scale factor (LMSF) and p, word insertion penalty. Global posterior probability is P(N|J, θ) such that N is the model given the acoustic data J and the parameters θ. Express the possibility of global posterior probability in terms of local posteriors $P(p_l^n|q_k^{n-1}, j_n, \theta)$ (where q_k^n denotes the specific state qk of N at time n) and language model priors. We have

$P(N|O) = \sum_{l_1}^{L} \sum_{l_N}^{L} P(q_{l_1}^1, \dots \dots, q_{l_N}^N, N|J)$ (5)

Where the posterior probability of the condition in sequence and modal can be fragmented into the multiplication of an acoustic model over language models:

$P(q_{l_1}^1, \dots \dots \dots, q_{l_N}^N, N|O) = P(q_{l_1}^1, \dots \dots \dots, q_{l_N}^N|J)\ P(N|J, q_{l_1}^1, \dots \dots \dots, q_{l_N}^N)$ (6)

$\simeq P(q_{l_1}^1, \dots \dots, q_{l_N}^N|J)$

$P(N|\ q_{l_1}^1, \dots \dots, q_{l_N}^N)$ (7)

After Rewriting the above equation (7) becomes

$P(N|J) \simeq \sum_{l_1,\dots,l_N} [\prod_{n=1}^{N} P(q_{l_n}^n|\ J_{n-c}^{n+d}) \frac{P(q_{l_1}^1|N)}{q_{l_n}^n}]\ P(N)$ (8)

The posterior probability, where O_{n-c}^{n+d} is limited to local Context. With the Bayes rule, we can show that:

Pre-processing of signal

B. Pre-processing of signal .

Speech recognition systems in this phase are used as subsequent feature extraction with increase efficiency and classification. Finally enhance the performance of speech

$\frac{P(J_{n-c}^{n+d}|q_{l_n}^n)}{P(\ J_{n-c}^{n+d})} = \frac{P(q_{l_n}^n|J_{n-c}^{n+d})}{P(q_{l_n}^n)}$ (9)

The distinction between the hybrid and the likelihood approaches lies at the local level. The hybrid system estimates local posteriors and is then discriminate at the frame level. The likelihood system estimates local probability density functions. Both systems can give us an estimate of the global posterior.

III. FRAMEWORK FOR SIGNAL PROCESSING

Methodology

The curvelet transform is a multi scale directional transform that accept a best possible non-adaptive sparse demonstration of object, edges and curves. The time-frequency and multi-resolution property of curvelet transform for input speech signal, it is decomposed into different channels in frequency [3]. The breakdown procedure can be iterated with successive manner, so that single signal is decomposed into a lot of lower resolution components. The choice of the best decomposition phase hierarchy based on the nature way of signal analysis like low-pass filter range [4]. It is efficient technique for extracting non-stationary signals data. The extracted curvelet coef ficients offer a flat version of energy

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

distribution in time and frequency of the signal. The following curvelet based speech recognition diagram represented in figure:

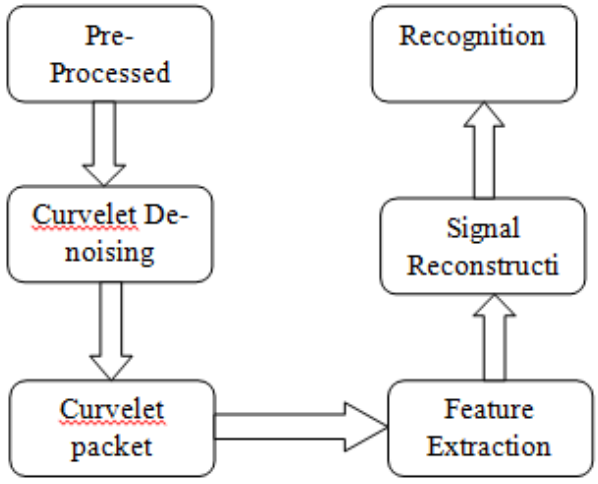

Figure 1. Curvelet Feature Extraction phases

signal recognition. At the end of this process, filtered and compressed frames of speech are forwarded to the feature extraction phase [5]. The following figure 2 represents the pre processing pipeline of speech signal.

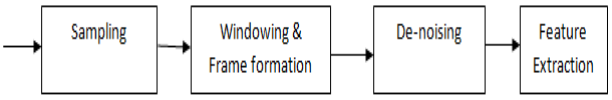

Figure 2. preprocessing pipeline of speech signal

C. Curvelet Function performance

Construction of curvelet functions; initially require defining special window functions that fulfil certain acceptability conditions. Let us regard as the scaled Meyer windows:

$$V(t)=\begin{cases}1 & |t|\le 1/3\\ \cos[\frac{\pi}{2}v(3|t|-1)] & 1/3 \le|t|\le 2/3\\ 0 & \text{else,}\end{cases}$$

$$X(r)=\begin{cases}\cos[\frac{\pi}{2}v(5-6r)] & 2/3 \le r\le 5/6,\\ 1 & 5/6 \le r\le 4/3,\\ \cos[\frac{\pi}{2}v(3r-4)] & 4/3 \le r\le 5/3,\\ 0 & \text{else}\end{cases}$$

Where v is a smooth function

$$\begin{cases}0 & x\le 0,\\ 1 & x\ge 1,\end{cases}$$

satisfying

$V(x) =$
$u(x) + u(1-x) = 1,\ x \in R$

An arbitrarily smooth window v is given by

$$V(x)=\begin{cases}0 & x\le 0\\ \frac{s(x-1)}{s(x-1)+s(x)} & 0< x <\\ 1 & \\ 1 & x \ge 1\end{cases}$$

With $s(y) = e^{-\frac{1}{(1+x)^2}+\frac{1}{(1+x)^2}}$

The above two functions U (t) and W(r) satisfy the conditions

$$\sum_{l=-\infty}^{\infty} V^2(t-l) = 1,\quad t \in R \qquad (10)$$

$$\sum_{j=-\infty}^{\infty} X^2(2^j r) = 1,\quad r\in 0 \qquad (11)$$

D. Feature Classifier of speech signal

Artificial Neural Networks are processing models stimulated by the brain. Machine learning and pattern recognition are capable of these models. It was working to recognize and classify vowel signals into their respective class. An artificial neuron with k given inputs converting a set $X \subset \mathbb{R}k$ of input signals (a k-neuron on X) is a function.

$E: Rk \times X \ni (\vec{y}, \vec{x}) \to F(\vec{y}, \vec{x}) = f(\vec{y}, \vec{x}) \in R,$

Where $\vec{w}$ is a weights vector, $\langle .,. \rangle$ is a actual scalar result, and $f: \mathbb{R} \to \mathbb{R}$ is called an activation function of the neuron. If f is a linear operator, then the neuron is called linear. A function

$E^*:= E(\vec{y}, .) : X \ni \vec{x} \to E^* (\vec{x}) \in \mathbb{R}$, is said to be a trained k-neuron on X.

The least-squares method is used for analysis of the learning process of a linear artificial neural network. The properties of Gram matrices will be used to analyse linear

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

training processes of artificial neural network analysis. Consider an Artificial neural network consisting of a single linear M-neuron. It is sufficient to accept a neuron and identity the activation function. In this case, the square deviation function is given by

$F(Y_1, \ldots\ldots, 1_N) = \sum_{n=1}^{n}[\, y(y_{1,\ldots\ldots}\, y_N)](n) - Z^n)2$

Where $y(w_1, \ldots\ldots\ldots, w_M)(n) = \sum_{m=1}^{M} x_m^{(n)} w_m$. Right side of the equation value calculates for the describing the learning process:

$$\frac{\partial F(w_1,\ldots\ldots 1_M)}{\partial w_{m'}} = \frac{\partial}{\partial w_{m'}} \sum_{n=1}^{N}[\sum_{m=1}^{N} x_m^{(n)} y_m - z^{(n)}]^2$$

Setting $H^{(n)} := \sum_{m=1}^{N} x_m^{(n)} y_m - z^{(n)}$,We obtain:

$$\frac{\partial E(w_1,\ldots\ldots 1_M)}{\partial w_{m'}} = \sum_{n=1}^{N} 2.H^{(n)} \cdot \frac{\partial (y^{(n)} - z^{(n)})}{\partial w_{m'}} = 2.\sum_{n=1}^{N} H^{(n)}.\frac{\partial y^{(n)}}{\partial w_{m'}} = 2.\sum_{n=1}^{N} H^{(n)}.\frac{\partial (\sum_{m=1}^{M} x_m^{(n)} m}{\partial w_{m'}} = 2\sum_{n=1}^{N} H^{(n)}.x_{m'}^{n} = 2.\sum_{n=1}^{N} x_{m'}^{n} \cdot [(\sum_{m=1}^{M} x_m^{(n)} w_m) - Z^{(n)}]$$

A linear one-layer ANN learning process is described by the system of T differential equations which are independent of each other. Each of them models the learning process of a single neuron. Indeed: Since the first component does not depend on $w_{t',m'}$, it is equal to zero. Thus:

$$\frac{\partial F}{\partial w_{t'm'}} = \frac{\partial}{\partial w_{t'm'}} \sum_{n=1}^{N}[(\sum_{m=1}^{N} x_m^{(n)} w_{t,m}) - z_t^{(n)}]2 = 2.\sum_{n=1}^{M} x_{m'}^{n}[(\sum_{m=1}^{N} x_m^{(n)} w_{t',m}) - z_{t'}^{(n)}]$$

The unique signal can be synthesized using the inverse discrete curvelet transform (IDCT).

IV. RESULTS AND DISCUSSION

In speech processing, feature extraction is one of the major level to enhance speech processing applications in real world. A bulky set of feature extraction techniques to apply on speech processing paths; however the division through curvelets is one of the popular techniques in these days for robustness. This research paper presents the progress and execution of curvelet transform technique using samples speech signals. The training and testing data set levels are recorded by using a PC-based audio input device. For resolution purposes the recoded samples are stored in matrices whose rows or columns represent as sample. The speech signal processing and analysis is done according to Section II, and the outcome weighted cepstral coefficients are the input to the pattern classifier. The following Table one shows the results of training and testing datasets using standard performance measures and defined as follows:

Table 1. Evaluated standard performance measures

S. No	Parameter	Training	Testing
1	Sensitivity	100%	100%
2	Specificity	100%	99 %
3	Accuracy	100%	97.8%

Sensitivity = $\frac{Tp}{Tp+Fn}$, and Accuracy = $\frac{Tp+Tn}{Tp+Tn+Fn+Fp}$

Specificity = $\frac{Tn}{Tn+Fp}$, and Sensitivity = $\frac{Tp}{Tp+Fp}$

The following Figure 3 Performance comparison of training and test datasets are evaluated using standard performance measures:

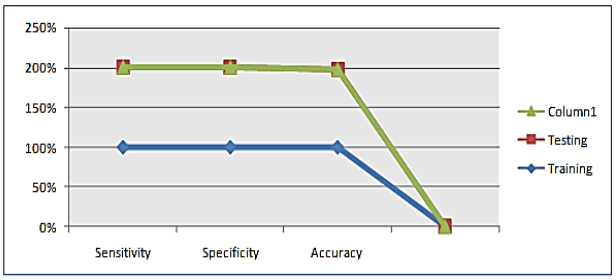

Figure3. Performance comparison of training and Test Datasets

The results show a classification above 97.8%, which demonstrates the suitability of the method for recognition.

V. CONCLUSION

A speech recognition scheme requires solutions to the problems to use modern algorithms and methods can process the speech signals simply and identify the text. This research concludes an expert speech recognition mechanism for isolated words based DCT and ANN methods was proposed. Artificial neural network play important role for classification of speech signal. They control similarly like human brain than conventional computer logic. In Feature Extraction phase, Curvelet transform extract the features from the given input speech signal and elements of these signals which help to achieving higher recognition rate. Artificial neural networks are used as feature matching classifiers. The throughput evaluation has been confirmed in terms of

accurate recognition rate, utmost noise power of interfering sounds, miss rates, hit rates, and false alarm rate. The rate of correct classification was about 97.8 % for the sample speech signals. The graphical results demonstrate that the proposed method can construct an accurate and robust classifier.

REFERENCES

[1] Bandhit Suksiri and Masahiro Fukumoto," Implementation of Artificial Neural Network and Multilevel of Discrete Wavelet Transform for Voice Recognition", Graduate School of Engineering, Kochi University of Technology (KUT), Kami City, Kochi 782-8502, Japan. https://www.ijecs.in /index.php/ijecs /article/download/387/338/citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.186.7405&rep=rep1matlab.izmiran.ru/help/toolbox/wavelet/ch01_i21.html. http://recognize-speech.com/preprocessing.

[2] Thiang and Suryo Wijoyo," Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot", 2011 International Conference on Information and Electronics Engineering IPCSIT vol.6 (2011) © (2011) IACSIT Press, Singapore.

[3] Cini Kuriana, Kannan Balakrishnan, "Development & evaluation of different acoustic models for Malayalam continuous speech recognition", in Proceedings of International Conference on Communication Technology and System Design 2011 Published by Elsevier Ltd, December 2011, pp.1081-1088.

[4] Md. Akkas Ali , Manwar Hossain and Mohammad Nuruzzaman Bhuiyan," Automatic Speech Recognition Technique for Bangla Words", International Journal of Advanced Science and Technology Vol. 50, January, 2013.

[5] Ms.Vimala.Ca and Dr.V.Radhab," Speaker Independent Isolated Speech Recognition System for Tamil Language using HMM", International Conference on Communication Technology and System Design 2011. Procedia Engineering 30 (2012) 1097 – 1102. file:///C:/Users/Admin/Downloads/AD0624555.pdf.

[6] P. Mishra and P. K. Mishra," A Study of various speech features and classifiers used in speaker identification", IJERT, ISSN: 2278- 0181.