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

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

An Approach to Extract Features from Speech Signal for Efficient Recognition of Speech

Narendra Kumar Saini[1], Vipra Bohara[2], Laxmi Narayan Balai[3],

P.G. Scholar, Yagyavalkya Institute of Technology, Jaipur, Rajasthan, India[1]

Assistant professor, Yagyavalkya Institute of Technology, Jaipur, Rajasthan, India[2]

H.O.D.(Electronics & Comm.), Yagyavalkya Institute of Technology, Jaipur, Rajasthan, India[3]

ABSTRACT:-The automatic recognition of speech, enabling a herbal and easy to apply method of communique among human and machine, is an active area of research. Speech processing has considerable software in voice dialing, cellphone verbal exchange, call routing, home equipment manage, speech to textual content conversion, text to speech conversion, lip synchronization, automation systems and so forth. Nowadays, speech processing has been developed as novel approach of protection. Characteristic vectors of authorized customers are saved in database. Speech capabilities are extracted from recorded speech of a male or girl speaker and in comparison with templates to be had in database. Speech features may be extracted by various techniques like LPC, PLP, MFCC and PLP- Relative Spectra and many others. A few parameters like PLP (Perceptual Linear Prediction) and MFCC(Mel-Frequency Cepstral Coefficient) considers the nature of speech whilst it extracts the capabilities, whilst LPC predicts the future features primarily based on preceding functions. Education models like neural network are trained for feature vector to expect the unknown sample. Techniques like Vector Quantization (VQ), Dynamic Time Warping (DTW), Assist Vector Device (SVM), and Hidden Markov Model (HMM) can be used for classification and recognition. We have defined neural community in our paper with LPC, PLP and MFCC parameters.

KEYWORDS*: LPC, PLP, MFCC and Neural Network.*

I. INTRODUCTION

Speech signals are natural happening signals and hence, are random signals. The information conveying speech signals are function of an independent fluctuating variable called time. In speech recognition the voice recognize by the information convey by a speech signal like pitch, stress, power spectral density, vowel duration [2]. It conveys records approximately phrases, expression, style of speech, accent, emotion, speaker identity, gender, age, the nation of fitness of the speaker and so on. There was loads of advancement in speech recognition technology, but still it has massive scope. Speech based totally gadgets locate their programs in our each day lives and have big advantages mainly for the ones people who are suffering from a few form of disabilities[11][2]. We are able to say that such people are constrained to expose their hidden skills and creativity. We also can use those speech based gadgets for security measures to reduce cases of fraud and theft [10].Speech is acoustic signal which contains records of concept this is created in speaker's thoughts. Speech is bimodal in nature, automated speech recognition (ASR) is considers acoustic statistics contained in speech signal. In noisy environment, it's miles less accurate. Audio visual speech recognition (AVSR) out weights ASR because it makes use of acoustic and visible data contained in speech. Speech processing can be executed at distinct 3 levels.

1. Signal level processing considers the anatomy of human auditory gadget and process the speech signal in the form of small chunks known as frames.

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

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

2. Phoneme level processing, speech phonemes are acquired and processed. Phoneme is the basic unit of speech.
3. Word level processing.

This version concentrates on linguistic entity of speech. The Hidden Markov Model (hmm) can be used to symbolize the acoustic state transition within the word.

The paper is prepared as follows: phase 2 describes acoustic characteristic extraction. In phase three, details of the characteristic extraction techniques like LPC, PLP and MFCC are discussed.

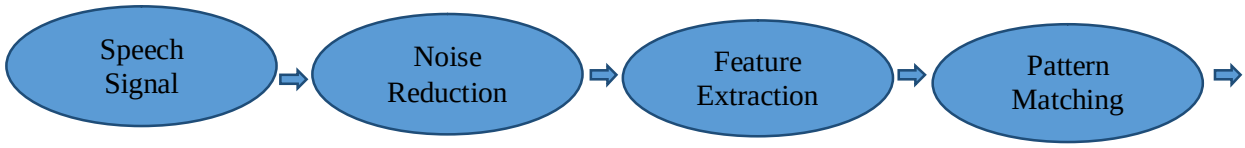


Fig.1: A speech recognition system overview

II. RELATED WORK

As I have mentioned in reference paper [1] which is related to linear predictive coding for speech data extraction. I also used this extraction technique but I also used MFCC technique in this paper which is quite efficient than LPC coding .In this paper I have gained 91.20% efficiency in speech recognition for the 40 isolated word by using MFCC technique so PLP and MFCC are derived at the concept of logarithmically spaced bank, clubbed with the concept of human auditory device and hence had the higher reaction compare to LPC parameters.Hidden Markov Model and Neural network are taken into considerationbecause the most dominant pattern popularity strategiesused within the subject of speech reputation.as human voice is nonlinear in nature, linear predictivecodes are not a very good choice for speech estimation.

III. BASIC IDEA OF ACOUSTIC FEATURE EXTRACTION

The task of the acoustic the front-quit is to extract feature functions out of the spoken utterance. Usually it takes in a body of the speech sign every 16-32ms and up to date every 8-16ms and plays certain spectral evaluation. The regular front-end includes among others, the subsequent algorithmic blocks: Fast Fourier transformation (FFT), calculation of logarithm (log), the discrete cosine transformation (DCT) and sometimes linear discriminate analysis (LDA). Considerably used speech abilities for auditory modeling are cepstral coefficients obtained via linear predictive coding (LPC). Every different well-known speech extraction is primarily based on Mel-frequency cepstral coefficients (MFCC).Methods based on perceptual prediction which is good below noisy conditions arePLP and Rasta-PLP (relative spectra filtering of logArea coefficients). There are a few one of a kind techniques like RFCC, LSP and so forth. To extract capabilities from speech. MFCC, PLP and LPC are the most widely used parameters in vicinity of speech processing.

IV. FEATURE EXTRACTION METHODS

Feature extraction in automatic speech recognition (ASR) is the computation of a series of characteristic vectors which offers a compact representation of the given speech signal. It is usually accomplished in 3 fundamental stages. The primary level is calledthe speech analysis or the acoustic front-cease, which performs spectra-temporal evaluation of the speech signaland generates raw features describing the envelope ofthe strength spectrum of short speech durations. The second level compiles a prolonged characteristic vectorComposed of static and dynamic functions. Ultimately, theLast

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

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

degree transforms these prolonged feature vectors into more compact and strong vectors which can be then provided to the recognizer.

(i). Mel Frequency Cepstrum Coefficients (MFCC)

The maximum conventional and dominant technique used toextract spectral features is calculating Mel-frequencycepstral coefficients (MFCC). MFCCs are one of themost famous characteristic extraction techniques utilized inspeech recognition based on frequency domain usingthe Mel scale that is based at the human ear scale [7].MFCCs being taken into consideration as frequency area functions are a whole lot greater accurate than time domain features.

Mel-frequency cepstral coefficients (MFCC) is arepresentation of the actual cepstral of a windowed shorttime sign derived from the short Fourier remodel(FFT) of that sign. The difference from the real cepstral is that a nonlinear frequency scale is used, which approximates the conduct of the auditory device. Moreover, the ones coefficients are strong and dependable to versions steady with audio device and recording conditions. MFCC is an audio characteristicextraction method which extracts parameters from thespeech just like ones which are utilized by human beings forlistening to speech, at the same time as at the equal time, deemphasizesall other information. The speech signal is first dividedinto time frames which include an arbitrary variety ofsamples. In maximum systems overlapping of the frames isused to easy transition from frame to frame. Everytime frame is then windowed with hamming window toget rid of discontinuities at the edges.

The Hamming window filter coefficient w(n) of length n are calculated according to the given formula-

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right)$$

$$, 0 \le n \le N-1$$

$$w(n) = 0 \qquad \text{, otherwise}$$

Wherein N is total wide variety of sample and n is modern-daypattern. After the windowing, speedy Fouriertransformation (FFT) is calculated for every frame toextract frequency components of a sign within the time domain.FFT is used to speed up the processing. The logarithmic Mel-scaled clear out financial institution is applied to the Fourier transformed frame [9]. This scale is approximatelylinear up to at least one kHz, and logarithmic at extrafrequencies. The relation among frequency of speech and Mel scale may be hooked up as:

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

MFCCs use Mel-scale filter bank where the higher frequency filters have greater bandwidth than the lower frequency filters, however their temporal resolutions are theidentical.

Fig.2. MFCC feature extraction technique

The last step is to calculate discrete cosine transformation (DCT) of the outputs from the filter-bank. DCT level coefficients in keeping withimportance, wherein the 0th coefficient is excludedin view that it is unreliable. The general system of MFCCextraction is shown on figure 1.

For each speech signal frame, a coefficient set of MFCC is computed. This set of coefficients is referred to as an acoustic vectorwhich represents the phonetically vitaltraits of speech and could be very useful for furtheranalysis and processing

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

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

in voice recognition. Wecan take audio of 2 sec which offers approximate128 frames each contain 128 samples (window size = 16 ms). We are able to use first 20 to 40 frames that provide accurateestimation of speech. Total of 40 two MFCC parameters encompass 12 unique, 12 delta (first order derivative), 12 delta-delta (second orderspinoff), three log strength and 30th parameter.

(ii). Linear Predictive Codes (LPC)

It is proper to compress signal for efficient transmission and storage speech signal. Digital signal is compressed earlier than transmission for efficient usage of channels on Wi-Fi media [8]. For medium or low bit rate coder, PLCis maximum extensively used. The LPC calculates an energy spectrum of the signal. It's far used for formant evaluation. LPC is one of the most powerful speech analysis strategies and it has won recognition as a formant estimation approach.At the same time as we pass the speech sign from speech evaluation filter to take away the redundancy in signal, residual errors is generated as an output. It may be quantized with the aid of smaller number of bits evaluate to unique signal. So now, as opposed to shifting entire signal we are able to switch this residual mistakes and speech parameters to generate the unique signal. A parametric model is computed based on least mean square error theory, this method being known as linear prediction (LP). By means of this technique, the speech sign is approximated as a linear combination of its p previous samples. in this approach, the acquired LPC coefficients describe the formants. The frequencies at which the resonant peaks arise are referred to as the formant frequencies [6]. Consequently, with this technique, the locations of the formants in a speech sign are envisioned via computing the linear predictive coefficients over a sliding window and locating the peaks inside the spectrum of the ensuing LP filter out. We've excluded 0th coefficient and used next 10 LPC coefficients.

In speech generation, for the duration of vowel sound vocal cords vibrate harmonically and so quasi periodic alerts are produced. While in case of consonant, excitation

Source may be taken into consideration as random noise [5]. VocalTract works as a clear out, that's accountable for speech response. Biological phenomenon of speech era may be without problems transformed in to equivalent mechanical version. Periodic impulse educate and random noise may be considered as excitation source and virtual clear out as vocal tract.

(iii). Perceptual Linear prediction (PLP)

The perceptual linear prediction PLP version developed byHermansky. PLP fashions the human speech based onthe concept of psychophysics of hearing [3] [4]. PLPdiscards irrelevant information of the speech and for this reasonimproves speech popularity price. PLP is equal toLPC except that its spectral characteristics have beentransformed to healthy characteristics of human auditory device.

$$\Omega(\omega) \rightarrow E(\omega) \rightarrow S(\omega) = (E(\omega)^{0.33}$$

Figure 3 shows steps of PLP computation. PLP approximates three major perceptual elements namely: the critical-band resolution curves$\Omega(\omega)$, the equal-loudness curve $E(\omega)$, and the intensity-loudness power $S(\omega)$,which are called the cubic-root.

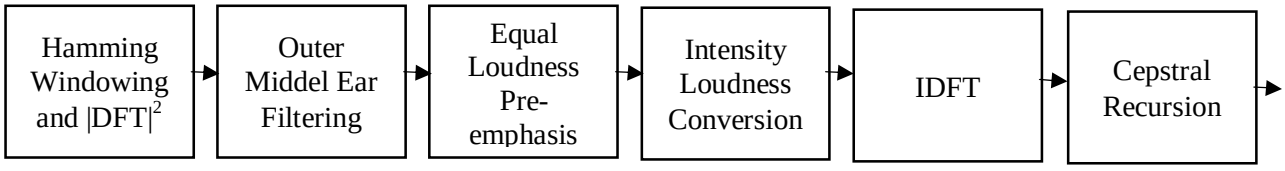


Figure 3. PLP parameter computation

Figure 3 shows the PLP parameter extraction process.

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

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

The windowed signal power spectrum can be calculated by the formula-

$$P(\omega) = Re(S(\omega))^2 + Im(S(\omega))^2$$

Frequency warping into the bark scale is applied. Thefirst step is a conversion from frequency to bark, whichis a better illustration of the human hearing resolution in frequency. The bark frequencycorresponding to anspeech signal frequency is-

$$\Omega(\omega) = 6\, ln\left[\frac{\omega}{1200\pi} + \sqrt{\left[\left(\left(\frac{\omega}{1200\pi}\right)^2\right) + 1\right]}\right]$$

The auditory warped spectrum is convoluted with theenergy spectrum of the simulated critical-band masking curve to simulate the critical-band integration of humanlistening to. The smoothed spectrum is down-sampled atperiods of ≈1 bark. The 3 steps frequencywarping, smoothing and sampling are integrated right into anunmarried filter-bank known as bark filter out financial institution. An equalloudnesspre-emphasis weight the clear out-bank outputsto simulate the sensitivity of hearing. The equalizedvalues are converted in keeping with the strength regulation ofStevens by elevating each to the strength of 0.33. Theresulting auditory warped line spectrum is similarlyprocessed with the aid of linear prediction (LP). Making use of LP to theauditory warped line spectrum manner that we computethe predictor coefficients of a (hypothetical) signal thathas this warped spectrum as an energy spectrum. Sooner or later,cepstral coefficients are acquired from the predictorcoefficients by a recursion this is equivalent to thelogarithm of the model spectrum accompanied via an inverseFourier remodel.

The PLP speech analysis technique is more adapted tohuman listening to, in evaluation to the traditional linear prediction coding (LPC). The primary difference amongPLP and LPC evaluation techniques is that the LP versionassumes the all-pole transfer characteristic of the vocal tractwith a specified range of resonances within the evaluation band. The LP all-pole version approximatespower distribution equally properly at all frequencies of theanalysis band. This assumption is inconsistent withhuman listening to, because beyond 800 Hz, the spectralresolution of listening to decreases with frequency andlistening to is likewise extra touchy within the center frequencyvariety of the audible spectrum [3].

V. NEURAL NETWORK

The generalization is the beauty of synthetic neuralnetwork. It affords high-quality simulation of factsprocessing analogues to human nervous system.Multilayer feed ahead network with back propagationset of rules is the commonplace preference in class andsample recognition. Hidden Markov model, Gaussian aggregate version, vector quantization are the number of thestrategies for acoustic capabilities to visible speechmotion. Neural community is one of the excellent selectionsamongst all. Genetic set of rules can be used with neuralcommunity for performance development by using optimizingparameter combination.

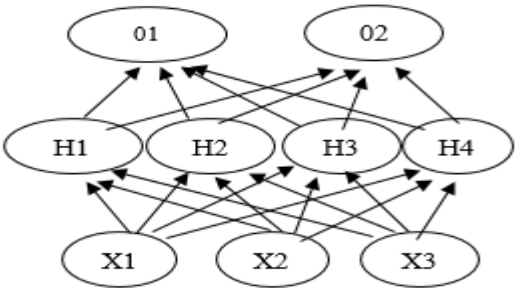
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018



01,02: Output Layer
H1,H2,H3,H4: Hidden Layer
X1,X2,X3: Input Layer

we will use multi-layer feed forward returned propagation neural community as proven in parent 4 with general varietyof features as range of input neurons in enter layer for LPC, PLP and MFCC parameters respectively. Asshownin discern four neural community consists of inputlayer, hidden layer and output layer. Variable quantity ofhidden layer neurons may be tested for first-rate consequences. Wecan educate community for exceptional combinations of epochswith purpose as minimal blunders price.

V. CONCLUSIONS AND DISCUSSION

We've got mentioned a few function extraction techniques and their problems and cons. LPC parameter isn't always sodesirable due to its linear computation nature. Itbecame visible that LPC, PLP and MFCC are the mostoften used functions extraction techniques within the fields of speech recognition and speaker verification packages. Hidden Markov Model and Neural network are taken into considerationbecause the most dominant pattern popularity strategiesused within the subject of speech reputation.as human voice is nonlinear in nature, linear predictivecodes are not a very good choice for speech estimation. PLPand MFCC are derived at the concept oflogarithmically spaced bank, clubbed with theconcept of human auditory device and hence had the higher reaction compare to LPC parameters.

REFERENCES

(1)Amrutha R's "Feature Extraction of Speech Signal using LPC "IJARCCEVol. 5, Issue 12, December 2016.
(2) C. Ittichaichareon, S. Suksri and T. Yingthawornsuk, speech Recognition using MFCC, *IJCSET*, 2012.
(3) Lei Xie, Zhi-Qiang Liu, "A Comparative Study of Audio Features For Audio to Visual Cobversion in MPEG-4 Compliant Facial Animation," Proc. of ICMLC, Dalian, 13-16 Aug-2006.
(4)H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," Acoustical Society of America Journal, vol. 87, pp.1738–1752, Apr. 1990.
(5)ChengliangLi,Richard M Dansereau and Rafik A Goubran , "Acoustic speech to lip feature mapping for multimedia applications", proceedings of the third international symposium on image and signal processing and analysis, vol. 2, pp. 829-832, 18-20 Sept. 2003.
(6)Honig, Florian Stemmer, Georg Hacker, Christian Brugnara, Fabio, "Revising Perceptual Linear Prediction ", In INTERSPEECH-2005, pp. 2997-3000. 2005.
(7) S. Dhingra, G. Nijhawan and P. Pandit, Isolated Speech Recognition using MFCC and DTW, *International journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,8(2),* 2013.
(8)B. P. Yuhas, M. H. Goldstein Jr., T. J. Sejnowski, and R. E. Jenkins, "Neural network models of sensory integration for improved vowel recognition," Proc. IEEE, vol. 78, Issue 10, pp. 1658–1668, Oct. 1990.
(9) L. Muda, M. Begam and I. Elamvazuthi, Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping(DTW) Techniques, *Journal of Computing, 3(2),*2010.
(10) "Voice Command Recognition System based on MFCC and DTW" by Anjali, A. Kumar and N. Birla, *International Journal of Engineering Science and Technology, 2(12),*2010.
(11) V. Sharma and P. Sharma, Discrete and continuous Mouse Motion using Vocal and Non-Vocal Characteristics of Human Voice, *IJCSET-4,* 2013.