International Journal of Advanced Scientific Technologies ,Engineering and Management Sciences (IJASTEMS-ISSN: 2454-356X) Volume.3,Special Issue.1,April.2017

Speaker Identification Using K-Nearest Neighbors (k-NN) Classifier Employing MFCC and Formants as Features

Sreelekshmi S Kumar,PG student and Syama R,Assistant Professor, Department of Elecronics and Communication Engineering,College of engineering kidangoor

Abstract— In communication ,speech is one of the major form .This work proposes a new method for speaker feature extraction based on Formants, Mel-frequency cepstral coefficients and K-NN denoted as FMFCNN . In this work extracted the formants and MFC coefficient features,this features are used as the input to K-NN and produce the corresponding output from training and the testing data. In early method we use sentences or words to extract features ,in this work partially recorded words, ie vowels are specifically used for speaker recognition. Every individual voice contain their own special features that clearly convey their emotions. This work which is useful for speech to text purpose, speaker recognition mainly in biometrics, google search, autonomous cars, navigation systems . FMFCNN results convey very good method for speaker recognition and identification. The MFC coefficients are very good results in speaker recognition as well as audio formants are effective method for speaker verification, combination of these two method result produce a better method for speaker identification than the former results The FMFCNN results are more superior.

Index Terms— Speaker identification and recognition,Formants,MFCC,K-nearest neighbouur

I.INTRODUCTION

From the begining of earth communication is very effective,and the speech which play a major role in communication environment. Every individual speech have their own special characteristics, and it convey different emotions.In day to day technologies speech processing mechanism lead a specific role in speaker identification as well as speaker recongnition methods. Speaker identification used in biometrics areas security system for more safety. Speech proceesing means study of speech signals and processing these signals. Which is regarded as a special case of digital signal processing in which applied to the speech signals, in this technique input is called speech recongnition and the output is called speech synthesis. In this work a new method FMFCNN is approached the combination of MFC coefficients and the formants as features extracted from the partially recorded audio mainly vowels.

In speech recongnition MFCC bring a major roll.The human speech contain discriminate features and the frequency 5KHz,speech signals are quasi-stationary ,means slowly time varying signals .Different choices of methods are available like LPC,DFT,DWTetc.The main purpose of the MFCC processor is to mimic the behavior of the human ears. MFCC which is susceptible to variation.Speech input is recorded as a sampling rate 1000Hz,this range of sampling frequency selected for minimize the effect of aliasing in the analog to digital conversion. Sampled signal can capture all frequency up to 5KHz,which cover most of energy of sounds that as generated by humans.

Formants are frequency peak,which have in the spectrum.they are especially for vowels,each formant corresponds to a resonance in the vocal tract .Formants considered as filters .During the speech signals the input signal filtered according to the physical structure of the oral tract.They named as F1,F2,F3etc.In vowels formant frequencies are very important,vowels are small recorded audio clip,which clearly lead the envelop of formants and we can get the peak correctly which is very effective feature in speaker recognition.

k-nearest neighbor is considered as lazy learning algorithm that classifies data sets based on their similarity with neighbors.Nearest neighbor have been used in statistical estimation and pattern recognition .KNN simple algorithm that stores all available data and classified with the new databases, classification is a technique which will examine the large pre-existing databases in order to generate the new information.Where K denote the number of data set items that are considered for the classification.The processing defers with respect to K value result is generated after analysis of stored data, it neglects any intermediate data.KNN is computationally simple algoritham and solve the complex problems.IT work with little information and learning process is simple.Training time was less for KNN than ANN

In FMFCNN method it's a combination of these MFCC Formants and the KNN system we used.

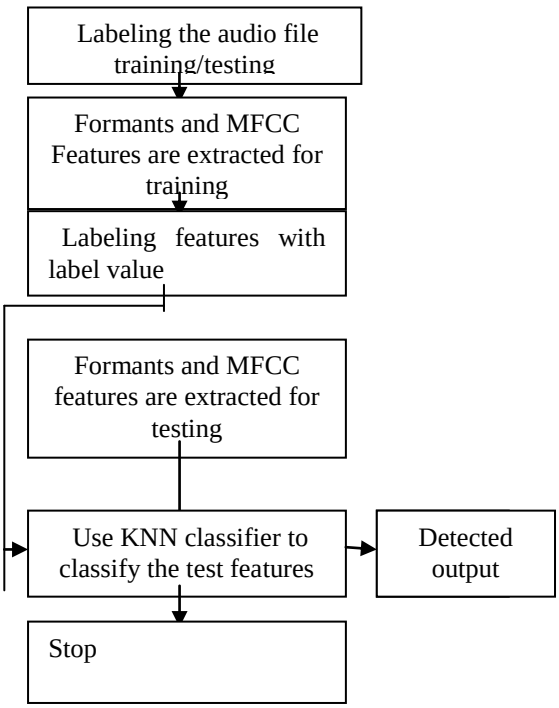

Fig 1.FMCNN method

II. LITERATURE SURVEY

The current work is the combined method of formants and the MFC coefficients within the KNN the previous study which convey the idea of formants and the wavelet entropy within the neural network. Disadvantage of wavelet packet entropy which leads the new idea of MFCC with Formants. There were many studies are carried by different researches to obtain the speaker identification. Also KNN which is much better than ANN because of training time is less KNN than ANN. Speaker identification using DWT,and the form DFT,wavelet which is better than Fourier transform.

Text independent speaker recognition using combined LPC and MFC coefficient within the artificial neural network.MFC coeffecients good in text-independent speaker recongnition,but alone it will failure in text independent. They we can use the linear predictive coding within the artificial neural network, but this cannot give the correct output. The wavelet packet they offer simultaneous localization in time frequency domain,and the better energy compaction, It will give better recognition than the LPC then the new method speaker recognition using wavelet and MFCC.But its combination work doesn't provide a better result.s

Researchers investigate Formants from the vowels they are partially recorded audio. Which will give the better recognition results, then researchers use this formants and the wavelet packet entropy within the ANN ,but the sample data decreses it become inefficient.MFC filters they are good in filtering of human sound.Also training time for KNN less than ANN,within limited amount of data,so this work convey the idea for use speaker identification and recognition using Formants and MFCC within the KNN.

III. PROPOSED METHOD

The Formant and MFCC are very useful in speaker identification and speaker recognition,and it have almost 100 percentage of correct classification in partially recorded audio,obtain the correct output from the minimum amount of data.

The steps for proposed method first of all take the vowel speech data base from different persons,and extract the formant and MFCcoefficients features from this speech database. Labelling these features with the label value These values are given to the KNN for classification purpose .In the testing process extract the formants and MFCC features give to the KNN from these two input KNN detect the output,mainly this work convey of feature extraction and k-nearst neighbor classifier

A .FEATURE EXTRACTION

In this paper two feature extraction method are used one is Formant and another one is Mel-scale feaquency cepstral coefficient.

Formant

Formants are frequency peak,which have in the spectrum,they are especially for vowels,each formant corresponds to a resonance in the vocal tract .Formants considered as filters .During the speech signals the input signal filtered according to the physical structure of the oral tract.They named as F1,F2,F3etc.In vowels formant frequencies are very important, vowels are small recorded audio clip,which clearly lead the envelop of formants and clearly lead the envelop of formants and we can get the peak correctly which is very effective feature in speaker recognition.

Formants are vowels pronounced high frequency peaks, we want 3 formants from feature extraction. Different pronunciation and different vowels are the input to the system.One person have similar formants otherwise we can use MFCC for better and accurate results.

Formants obtained by LPC ie Linear predictive coding which is a provened technology.Take the square root of the prediction polynomial. The speech signals which pass through the low pass filter because obtain the sampling frequency of the data.First of all the speech signals windowing by hamming window the use of hamming window which is prefer the middle portion of the recorded signal which will contain maximum amplitude values,it will give the better result.

This hamming window signal multiplied with the data then we can obtain the windowing function.Apply the pre-emphasis filter,which is a all pole regressive one mode filter.Input of the pre-emphasis filter is a transfer function

and it will filtered and get the filter coefficient.The transfer function is the system parameters.Then determine the LPC , find the root of the LPC coefficient. The LPC coefficientes are real valued functions and the roots are complex conjugate.Then we can find the angle,for this convert the angular frequency in radiance per sample represented by the angle to hertz represented as bandwidth,take the log then get the matrix. 3 Formants are obtained, ranges from 90-400Hz

MFCC

In speech recongnition MFCC bring a major roll.The human speech contain discriminate features and the frequency 5KHz,speech signals are quasi-stationary ,means slowly time varying signals .Different choices of methods are available like LPC,DFT,DWTetc.The main purpose of the MFCC processor is to mimic the behavior of the human ears.MFCC which is susceptible to variation. Speech input is recorded a sampling rate 1000Hz,this range of sampling frequency selected for minimize the effect of aliasing in the analog to digital conversion. Sampled signal can capture all frequency up to 5KHz,which cover most of energy of sounds that as generated by humans.

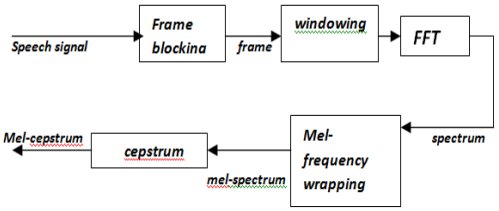

Fig 2.MFCC method

The pre-emphasised signal is blocked into frames of N samples with adjacent frames being separated by M.The first frame consist of N sample,the second frame start with M samples after the first frame and overlap it by N-M samples. The process is still which will continuous to the last signal.

Windowing which is a graphical user interface.The spectrum signal which will applied to bank of filter called Mel filter bank to determine the frequency content across each filter.Which is a traingular band pass filter which mimics the human auditory system.Filter bank based on non-linearity frequency scale called Mel scale.The filter are overlapped in such a way that lower boundary of one filter which at the centre frequency of the previous filter and the upper boundary is situated at the centre frequency of the next filter we get the maximum response.

20 filters are selected and uniformly spaced in the Mel-frequency scale between 0 and 4Hz.Mel-frequency spectrum is computed by multiplying the signal spectrum with a set of traingular filters designed using the Mel-scale

If frequency f,then Mel-frequency

$$B(f) = [1125\ln(1+\frac{f}{700})]mels \quad (1)$$

If M is the Mel the corresponding frequency

$$B^{(-1)}(m) == [700e^{(m/1125)}-700]Hz \quad (2)$$

After computing we can obtain the transfer function of the filter

The final step in speech processing is to compute the MFCC coefficients log Mel spectrum is converted back to time.then applyDCT to MFCC

$$MFCC_x = \sqrt{\frac{2}{M}} \sum_{m=1}^{M} X_{m(ln)} COS(\frac{\pi k(m-0.5)}{m}) \qquad 1 \le k \le p \quad (3)$$

DCT is fourier related transfer similar to DFT,but use only real number.

$$X_k = \frac{1}{2}(x_0 + (-1)^k X_{N-1}) + \sum_{n=1}^{N-2} X_n \cos\left[\frac{\pi}{N-1} nk\right] \quad \text{k=0.....N-1} \quad (4)$$

The number of resulting Mel-frequency cepstrum coefficients is chooses as relatively low, in the order of 12-20 coefficient. The zeroth coefficient usually eliminate because it has average log energy of the frame and it will carry less information.

B. k-NEAREST NEIGHBOR ALGORITHM

k-nearest neighbor is considered as lazy learning algorithm that classifies data sets based on their similarity with neighbors. Nearest neighbor have been used in statistical estimation and pattern recognition .KNN simple algorithm that stores all available data and classified with the new databases, classification is a technique which will examine the large pre-existing databases in order to generate the new information.Where K denote the number of data set items that are considered for the classification.The processing defers with respect to K value result is generated after analysis of stored data, it neglects any intermediate data.KNN is computationally simple algoritham and solve the complex problems. IT work with little information and learning process is simple. Training time was less for KNN than ANN.

Fig 3 KNN classifier

Algorithm of KNN where training samples are vectors in multidimensional feature space,training phase of algorithm consist only storing the feature vector or the samples and class labels of training samples. In the classification phase K is a user defined constant(query or a test point). Commonly used distance metric for continuous variable is Euclidean distance,for text classification another metric can be used overlap metric (Hamming distance) The classification accuracy of KNN can be improved significantly by large margin nearest neighbor or neighborhood component analysis.

Parameter selection of KNN where the K depend upon the data. Genarally large value of K reduces the effect of noise on the classification. A good K can be selected by various technique like hyperparameter optimization .The special case where the class is predicted to be the class of the the closest training sample is called nearest neighbor algorithm. In binary classification K choose as an odd number.The l-nearest neighbor classifier it assign a point *x* to the class of its closest neighbor in the feature space that is $C_n^{lnn}(x)=Y_{(l)}$ .Size training data increases it doesn't exceed the bayes error rate. The weighted nearest neighbor which K nearest neighbor classifier assign K nearest neighbor weight is $\frac{1}{k}$ and all others have 0 weight. This is generally known as weighted nearest neighbor classifier.

Where ith nearest neighbor is assigned a weight w_{ni}

$$\sum_{i=1}^{n} w_{ni=1} \tag{5}$$

It hold an analogous result on the strong consistency of weighted nearest neighbor

Let C_n^{wnn} denote weight nearst neighbor classifier with weight

$$\{w_{ni}\}_{i=1}^{n}$$

$$R_R(C_n^{wnn}) - R_R(C^{Bayes}) = (B_1 s_n^2 + B_2 t_n^2)\{1 + 0(1)\} \tag{6}$$

For constants B1andB2

$$s_n^2 = \sum_{i=1}^{n} w_{ni}^2$$

$$t_n = n^{-2/d} \sum_{i=1}^{n} w_{ni} \left\{ i^{1+2/d} - (i-1)^{1+2/d} \right\}$$

The optimal weighting scheme

$$\{w_{ni}^*\}_{i=1}^{n}$$

$$k^* = \left[B_n^{\frac{4}{d+4}} \right]$$

$$w_{ni}^* = \frac{1}{k^*}\left[1 + \frac{d}{2} - \frac{d}{2k^{*2/d}}\left\{ i^{1+2/d} - (i-1)^{1+2/d} \right\}\right] \tag{7}$$

where

$$i = 1,2, \ldots k^*$$

IV. PROPOSED ALGORITHM

In FMFCNN method use the formnats and the MFCC features as input to the KNN algorithm for testing and training process the following steps are used for the training

- Labeling the audio file for testing/ training.
- Extract formant features using LPC from the database,for training
- Extract the MFCC feature from database by frameblocking,windowing,Mel-frequency wrapping,for training.
- Concatenate these two features.
- Labeling these features with labeling value.
- Extract formant features using LPC from the database,for testing.
- Extract the MFCC feature from database by frameblocking,windowing,Mel-frequency wrapping,for testing.
- Concatenate these two features,and which will given to the KNN to classify the test features.
- Labeling features with label value also given to KNN classifier
- Detected the output
- Stop

V.SIMULATION RESULTS

`The proposed method is implemented with the number of speaker and each speaker gives 3 formants and 20 MFCC features after training and testing process.

So here the k-nearest neighbor is designed to clarify the speakers.This neural network structure is trained with the 3 Formants and 20 MFCC features using KNN classifier.

Formants are obtained from the LPC which is a provened technology.Take the squre root of the prediction polynomial,by using the hamming window windowing the speech signal.Then the data multiplied with the window function ,it will passes through the pre-emphasis filter,after filtering processwe can obtain the parameters simple representation of the same speaker ,different vowels,formants

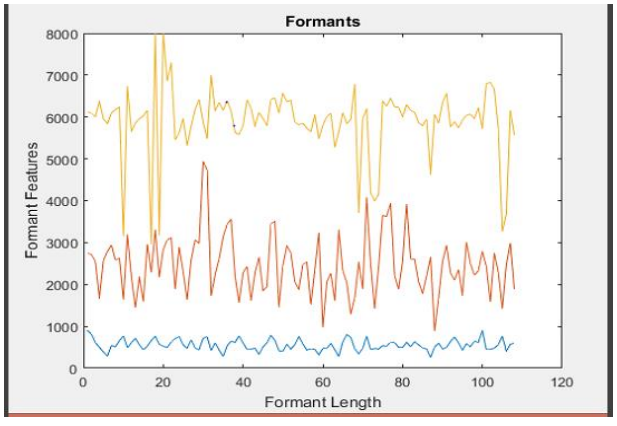

Fig 4 formant features

20 MFC coefficient features are obtained from the experiment results.when the vowel speech data signal then the frame blocked then the windowing process will takesplace after that filtering by the mel filters they are very efficient for audio signal filtering and produce the

cepstrum coefficients.The plot which will shown coefficient variations.

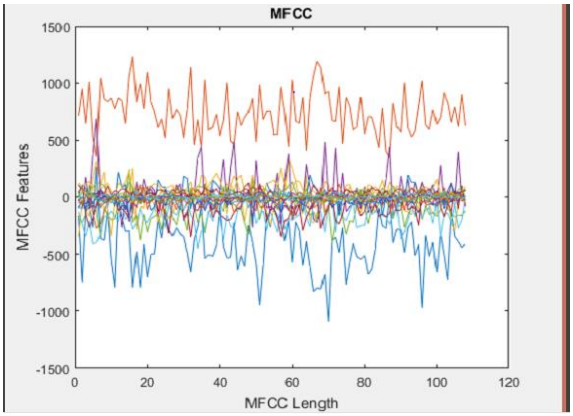

Fig 5 MFCC features

These plots represented identified target vs iteration The voiced speech signals of number of persons,then take the formant features and the MFCC features it give the KNN classifier it will produce the detected output,labeling the audio file for testing and training purpose then the formant and MFCC features are extracted for the training purpose. Concatenate these two features. Labelling features with label value,these values are given to the KNN classifier. Features are extracted formants and MFCC for testing purpose. These features are give the input to the KNN classifier to classify test features,from this two inputs we can obtain the detected output.

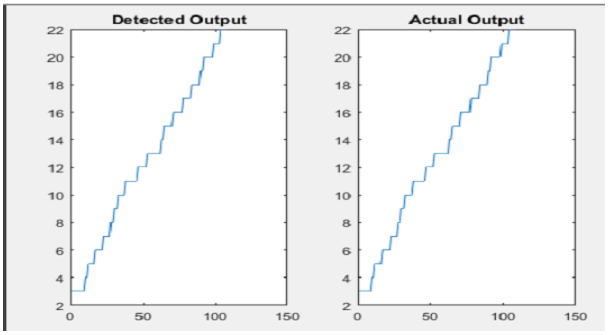

Fig 7 identified vs target

VI CONCLUSION

Simulation result shows that the speaker identification using KNN employing formants and MFCC known as FMFCNN is a very good method for identification and recognition from the minimum amount of data within a very short time duration by the use of KNN classifiers.Efficient and correct methods are chooses to obtain the 100 percentage output.Formants are the frequency peaks,they are especially for the vowels.MFCC is a very good method for speaker recognition more effective than the wavelet packet because the human features are extracted by the MFCC which is more efficient than the wavelet packet. Filter bank in the MFCC filter are designed for better speech feature extraction. KNN which is more better than ANN to get the better result.

ACKNOWLEDGEMENT

We would like to acknowledge the founding provide from the TEQIP phase 2 of college of engineering kidangoor to publish this work.

REFERENCES

[1] Khaled Daqrouq,Tark A.Tutunji,"Speaker identification using vowels features through a combined method of formants,wavelets,and neural network classifiers"Elecrical&computer engineering Department, King Abdulaziz university,jeddah,saudhi arabia 2015

[2] PPS Subhashini,Turimerla Pratap "Text-Independent speaker recognition using combined LPC and MFCC" Assosiate professor,ECE Dept,RVR&JC College of Engineering 2014

[3]A.Moosavian,H.Ahmadi,A.Tabatabaeefar and M.Khazee "Comparison of two classifiers;K-nearst neighbour and artificial neyral natwork,for fault daignosis on a main engine journel-bearing" Department of mechanical engineering of agricultural machinery,University of Tehran,Karaj,Iran 2012

[4] D. Avci, An expert system for speaker identification using adaptive wavelet sureentropy, Expert Syst. Appl. 36 (2009) 6295–6300.

[5] D.A. Reynolds, T.F. Quatieri, R.B. Dunn, Speaker verification using adaptedGaussian mixture models, Digit. Signal Process. 10 (1–3) (2000) 19–41.

[6] J.-D. Wu, B.-F. Lin, Speaker identification using discrete wavelet packet trans-form technique with irregular decomposition, Expert Syst. Appl. 36 (2009)3136–3143.

[7] Rabiner LR, juang BH, "Fundamentals of speech recognition", Prentice Hall India

[8] Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman "SPEAKER IDENTIFICATION USING MEL FREQUENCY CEPSTRAL COEFFICIENTS" 3rd International
Conference on Electrical & Computer Engineering ICECE, December 2004

[8] R. Sarikaya, B.L. Pellom, J.H.L. Hansen, Wavelet packet transform features withapplication to speaker identification, in: Proceedings of the IEEE Nordic SignalProcessing Symposium, 1998, pp. 81–84]

[9] E.S. Fonseca, R.C. Guido, P.R. Scalassara, C.D. Maciel, J.C. Pereira, Wavelettime–frequency analysis and least squares GWPNN support vector machinesfor the identification of voice disorders, Comput. Biol. Med. 37 (2007) 571–578.

[10] J. Malkin, X. Li, J. Bilmes, A Graphical Model for Formant Tracking, SSLI Lab,Department of Electrical Engineering, University of Washington, Seattle, 2005.

[11] M.P. Gelfer, V.A. Mikos, The relative contributions of speaking fundamentalfrequency and formant frequencies to gender identification based on isolatedvowels, J. Voice 19 (4) (2007) 544–554.

[12] J. Bachorowski, M. Owren, Acoustic correlates of talker sex and individual talkeridentity are present in a short vowel segment produced in running speech, J.Acoust. Soc. Am. 106 (2) (1999) 1054–1063.

[12] X. Huang, A. Acero, H.-W. Hon, Spoken Language Processing, Prentice Hall PTR,2001.

[13] S. Kadambe, G.F. Boudreaux-Bartels, Application of the wavelet transform forpitch detection of speech signals, IEEE Trans. Inf. Theory 32 (March) (1992)712–718.

[14] Acero, Formant analysis and synthesis using hidden Markov models, in: Proc.Eur. Conf. Speech Communication Technology, 1999.