SPEECH RECOGNITION USING REVERSE WAVE TRANSFORMATION

Manisha Gupta[1], Priyanka Singh [2]
Assistant Professor, WIT, INDIA

ABSTRACT

In this paper, a proposed speech recognition algorithm is presented. The paper deals with the recognition of input samples of voice in wave format using reverse wave transformation. The performance of the proposed speech recognition algorithm using a reverse wave transformation is evaluated by different samples of speech signals which are in wave format. The proposed algorithm is tested using different recognized and unrecognized samples obtaining a recognition ratio about 99%.

I. INTRODUCTION

Speech recognition is the process of taking the spoken word as an input and matches it with the database of previously recorded speeches using reverse wave transformation on basis of various parameters. The ultimate aim of Speech recognition research is to allow a computer to recognize matches of audio with 100% accuracy that are lucidly spoken by any person, independent of vocabulary size, noise, accent, or channel conditions. In spite of several decades of research in this area accuracy greater than 90% is only attained when the task is constrained in some form.

Depending on how the task is constrained, different levels of performance can be attained; for example audio of a person can be recognized whether it is the same person or not. For speech recognition of different speakers on basis of certain features, accuracy is not greater than 87%, and processing can take hundreds of times real-time.

Speech processing is one of the stimulating areas of signal processing. The voice recognition is done so as to determine which speaker is present there based on the individuals utterance. There are various techniques that have been proposed for reducing the mismatch between the testing and training data. Spectra l[2,3], or cepstral domain are various methods that are used [4].

Firstly, human voice is converted into digital signal form and digital data is produced which represent each level of signal at every discrete time step.

The digitized speech sample are then processed and according to voice features, the voice is recognized. The coefficient of voice features can go through reverse wave transformation to select the pattern that match the database and input frame in order to minimize the resulting error between them.

The method of reverse wave transformation is used to compare and find the similarity between tested voice and the input voice. The reverse wave transformation technique can be implemented using MATLAB. This paper reports the matching of the voice of the speaker using reverse wave transformation. Analysis of voice is done by taking an input from user through microphone.

II. LITERAUTURE SURVEY

Waelal.S.et.al [2009] proposed a system that finds the correct identity of speaker on basis of Continuous, Discrete Wavelet Transform and Power Spectrum Density.[1]The system depends on the multi-stage features extracting due to its better accuracy. Good capability was shown by systems based on multistage feature tracking. According to Cui and Xue [2009], two types of recognitions are there [2]: talker recognition in which user who is speaking was recognised and another is voice recognition system in which the user pronounces according to the stated contents. For voice recognition that is irrelevant to text, the user does not need to pronounce contents of the talkers. It is hard to build up models.

Tariq Abu Hilalet al [2011]integrated Discrete Wavelet Transform (DWT), and Logarithmic Power Spectrum Density (PSD) [3]for speaker accurate formants extraction, afterward correlation coefficient is used for features classification, the correlation thresholding factor is adjusted. As the system works with the recorded samples, the features tracking capability was excellent with text dependant dataset; so the system can be applied in password, PINs identification, security system or mobile phones.

The proposed system is simulated; the results show excellent performance, around 95 % Recognition Rate.

K. Daqrouq, [2009] presented an idea of noise cancellation for the speech signal so that robustness of the speaker recognition system increased. Two blocks are there : Discrete Wavelet Transform DWT and Adaptive Linear Neuron (Adaline) Enhancement Method (DWADE) and Wavelet Gender Discrimination (WGD) and Speaker Recognition using Discrete Wavelet Transform (DWT) Power Spectrum Density (PSD) [4].The tested signal were enhanced up to 15 dB by Wavelet Transform and Adaline Enhancement Method which also increased the speaker recognition rate. Back Propagation Feed Forward Neural Network (BPFFNN) perceptron classification methods were used.

Trived.N [2011] presented an operative and robust technique for extracting features for speech processing based on the time-frequency multi-resolution property of wavelet transform [5], the input speech signal is disintegrated into various frequencychannels. The major issues regarding the design of this Wavelet based speech recognition system are choice of optimal wavelets for speech signals, decomposition level in the DWT, selecting the feature vectors from the wavelet coefficients. Classification of the words is done using three layered feed forward network.

Pawar.et.al [2011]proposed speaker recognition system based on the wavelet Transform [6]. The system had two main blocks, signal enhancement by feature extracting and identification. Adalinewas used in first block as neural network to enhance each sub-signal produced by the DWT. Multiple inputs could be applied to the neural net depends on selected level as system depends on DWT.

Daoud . O. et.al [2009] improved the robustness of the speaker identification systems based on a 00modifiedversion of Principal Component Analysis (PCA)and Continuous Wavelet Transform (CWT) [7]. A robust feature extraction method was based on MPCA instead of Mel Frequency Cepstral Coefficient (MFCC) which was based on converting the common Eigen matrix from two dimensional into a one dimensional one.

Zhao. C et al [2010] proposed a new model for speech model, which is used for speech transformation and speech recognition thatsplits segmentations of a speech into an adaptive segment, in which its speech wave is integrated [8].The accuracy and few coefficients of model are due to use ofMorlet wavelet to extract primary periods.

Calvo et al. [2007] examined the application of Shifted Delta Cepstral (SDC) [9] features in biometric speaker verification and evaluates its robustness to channel/handset mismatch due by telephone handset variability.

W. Alkhaldi [2002] presented Discrete Wavelet Transform- based feature extraction technique [10] for multi-band automatic speech/speaker recognition. This technique has comparable performance with conventional technique. If has been found that both techniques are complementary under mismatched conditions, if the features extracted using each of them are combined.

Hao .Y (1994) investigated and presented a speaker adaptation scheme [11] that transforms the prototype speaker's Hidden Markov word models to a new speaker. Transformations are smeared to state transition matrix as well as the probability distribution functions of a Hidden Markov word model. These transformations are enhanced through maximizing the joint probability of a set of input pronunciations of the new speaker.

Zamani. B [2008] proposed a framework to improve independent feature transformations such as PCA (Principal Component Analysis), and HLDA (Heteroscedastic LDA) by using the minimum classification error cri000terion.Zamani [12] modified full transformation matrices such that classification error minimized for mapped features. The performance of the new method was improved as compared to PCA, and HLDA transformation for MFCC in both clean and noisy conditions.

Salomon .J et al [2008] discussed the use of frequency transformations and pattern recognition to improve the accuracy of single speaker multiple word speech recognition systems [13]. A speaker database with 124 words was taken that helps in showing the performance of the speech recognition system.

III. PROBLEM IDENTIFICATION

This research work focus on providing better performance in audio recognition algorithm by integrating digital signal transposition with audio recognition techniques. Main emphasis is to recognize audio with reverse wave transformation to achieve better results. Speech recognition is becoming popular in real time security systems. The methods developed so far are working efficiently and giving good results. Neural networks take so much time in training the neural and so the technique is going to be formed that take much less time by simply processing the signal by using reverse wave transposition. This paper deals with efficiently recognizing voice by using signal transposition and also gives better results than technique used using neural networks. To achieve this, a new hybrid methodology will be proposed which will recognize the audio. To do performance analysis different metrics will be considered in this paper. The performance of speech recognition systems is usually evaluated in terms of accuracy and speed. Accuracy is usually rated with word error rate (WER), whereas speed is measured with the real time factor. To do performance comparison the result of proposed algorithm will be compared with some well-known audio recognition algorithm.

IV. MOTIVATION

The common problem with identification system nowadays is that the system can easily be fooled. Although it uses biometric identification which is unique from everyone else, there are still ways to fool the system. As for fingerprint identification, it does not have a good psychological effect on the people because of its wide use in crime investigations. Also, when the surface of human fingerprint is hurt, the recognition system will have problems to recognize the user because the system recognizes the surface of the fingerprints while for face recognition, people are still working the pose and the illumination invariance.

V. OBJECTIVES

1. **Time:**

   The time is the time taken by the algorithm to run. Time is directly proportional to number of inputs elements. Speed is inversely proportional to time. Speed increases when time taken decreases and decrease with increase in time taken to recognize voices.

2. **Improved Hit Ratio :** It is defined as the number of voices recognized to the total number of voices that are given input.

   Hit ratio= Accurate hits / total hits * 100

3. **Reduced Error rate:**

   The error rate must be reduced so that accurate recognition of input voices must be there. Basically,

   ERROR RATE= 1- HIT RATIO

4. **Increased Accuracy rate :**

   Accuracy= Hit Ratio – Miss Ratio / Total * 100.Accuracy increases with decrease in error rate and increase in hit ratio.

SERIAL NO.	METRICS	NEED TO BE MAXIMIZED OR MINIMIZED
1.	TIME	MINIMIZED
2.	HIT RATIO	MAXIMIZED
3.	ERROR RATE	MINIMIZED
4.	ACCURACY RATE	MAXIMIZED

VI. METHODOLOGY

The code is implemented using MATLAB. Various Formulas that are used are:

1. **Transpose of a signal:** Transpose of a signal is T^i.

2. **Correlation:-**Correlation computes a measure of similarity of two input signals as they are shifted by one another. The correlation result reaches a maximum at the time when the two signals match best. The correlation value is denoted by M. Match the test voice with trained database one by one and find correlation.

Step5.

Find the maximum value of correlation and let Correlation be M.

Step6.

If the value of M > Threshold

Then Recognised Else Not found. If the correlation is greater than the threshold value then it must be the same voice else the input voice is not found in the database.

Step7.

End

VIII. RESULT

Various input samples are taken in which 100 samples of recognized persons and 100 unrecognized persons samples are there and then tested with the previously stored 10 samples of recognized voices.

The 200 samples in total are taken for testing in which The Results of recognition of input samples are as follows:

NO. OF INPUT SAMPLES	HITS	MISS	ACCURACY (IN PERCENT)	ERROR RATE
20	20	0	100	0
40	40	0	100	0
60	60	0	100	0
80	79	1	98.75	1.25
100	99	1	98.75	1
120	119	1	98.75	0.83
140	138	2	98.57	1.42

160	158	2	98.57	1.25
180	177	3	98.33	1.66
200	197	3	98.33	1.5

1) NOISE OR DISTURBANCE:-50 Samples are taken.

NO. OF INPUT SAMPLES	HIT	MISS	ACCURACY	ERROR RATE
10	9	1	90	10
20	19	1	95	5
30	28	2	93.33	6.66
40	37	3	92.5	7.5
50	46	4	92	8

IX. CONCLUSIONS

This paper has discussed audio recognition algorithm which is important in improving the speech recognition performance. The technique was able to authenticate the particular speaker based on the individual information that included in the audio signal and the recognition is done using reverse wave transformation. The results show that these techniques could used effectively for audio recognition purposes.

REFERENCE

[1] Cheong Soo Yee and abdul Manan ahmad, Malay Language TexIndependent , " Speaker Vertification using NN-MLP classsifier with MFCC" 2008.
[2] P. Lockwood, J. Boudy, " Experiments with a Nonlinear SpectralSubtractor (NSS), Hidden Markov Models and the Projection, for Robust Speech Recognitio n in Cars" 1992.
[3] A. Rosenberg, C.-H. Lee, F. Soong, " Cepstral Channel Normalization Techniques for HMM-Based Sp eaker Verification" 1994.
[4] Dr Philip Jackson, Features extraction 1.ppt,, University of Surrey,guilford GU2 & 7XH.
[5] WAEL AL-SAWALMEH et.al "The Use of Wavelets in Speaker Feature Tracking Identification System Using Neural Network" May 2009.
[6] B and Xue. T (2009) " Design and realization of an intelligent access control system based on voice recognition" ISECS Iinternational colloquium on computing" CCCM 2009., pp. 239-232.
[10] Hilal. T et. al "Speaker Verification System Using Discrete Wavelet Transform And Formants Extraction Based On The Correlation Coefficient " IMECS 2011.
[11] K. Daqrouq et. al" Speaker Identification System Using Wavelet Transform and Neural Network"; International Conference on Advances in Computational Tools for Engineering Applications ; 2009.
[12] Trived. N et.al "Speech Recognition by Wavelet Analysis" 2011.
[13] Pawar. M et.al " Speaker Identification System Using Wavelet Transformation and Neural Network" 2011.
[14] Daoud .O et.al " Modified PCA Speaker Identification Based System Using Wavelet Transform and Neural Networks" 2009.
[15] Chengfeng Zhao"A New Method of Speech Model Based On Periodic Expanded" IEEE 2010.
[16] Calvo et.al " Channel / Handset Mismatch Evaluation in a Biometric Speaker Verification Using Shifted Delta Cepstral Features" 2007.
[17] Alkhaldi. W et.al, "Automatic Speech/Speaker Recognition In Noisy Environments Using Wavelet Transform" IEEE 2002.
[18] Hao.Y et.al " Speech Recognition Using Speaker Adaptation by System Parameter Transformation" IEEE 1994.
[19] Zamani.B "Discriminative Transformations of Speech Features Based on Minimum Classification Error" Salomon.J "Speech Recognition using Frequency Transformations" IEEE 2008.