International Journal of Advance Foundation and Research in Computer (IJAFRC)
Volume 1, Issue 8, August2014. ISSN 2348 - 4853

Voice Recognition Using Artificial Neural Network

Supriya Patil[1],Derrick Pereira[2],Garry Da Costa[3],Prachi Satardekar[4],Zanshila Rodrigues[5]
[1,2,3,4,5] Electronics and Telecommunication , Padre Conceicao College of Engineering,Verna, Goa, India

E-mail: supriya_adi@rediff.com, derrickpereira14@gmail.com, garrydacosta13@gmail.com
prachisatardekar3@gmail.com, zanshilarod27@gmail.com

A B S T R A C T

Voice is the sound produced in a person's larynx and uttered through the mouth, as speech or song. Like finger prints and the iris of the eye the voice of each person is unique. The parameters of a person's voice differ in pitch, timber and vocal intensity. Hence it can be used for authentication. In this paper feature extraction methods used are Linear Predictive Coding, Fast Fourier Transform and wavelet transform to extract the above mentioned features from a database of voices that were recorded. Artificial neural network classifier operates on the result of classification. Some of the applications are Forensic applications, Artificial intelligence, Access Control, Transaction Authentication, Law Enforcement and Speech Data Management.

Index Teams: ANN: Artificial Neural Network, LPC: Linear Predictive Coding, DWT: Discrete Wavelet Transform, FFT: Fast Fourier Transform

I. INTRODUCTION

The voice is produced in a person's larynx and uttered through the mouth, as speech or song. The three different parameters that make a voice different are loudness(total energy contained in a persons voice),pitch(frequency representation which depends on speed of vibration of our vocal chord) and timber (quality in a persons voice and it is measured from the envelope of the power spectrum obtained which is different for each person).

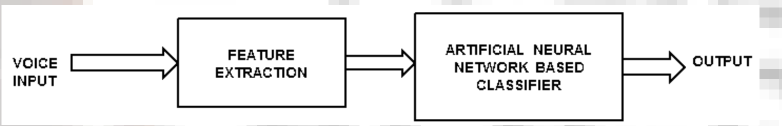

Figure 1. Block Diagram

II. PROCEDURE OVERVIEW

In this paper methods used for feature extraction are Fast Fourier Transform, Linear Predictive Coding and Discrete wavelet Transform. Artificial neural network is used for classification.

A. Feature Extraction

1. Fast Fourier Transform (FFT)

A Fast Fourier Transform is an algorithm to compute discrete Fourier Transform and its inverse. A Fourier Transform converts time domain to frequency domain and vice versa. An FFT rapidly computes such Transform. As a result FFT are widely used for many applications in engineering, science and mathematics. FFT gives the exact same result as DFT does, the only difference is that FFT does it faster. The sequence of N complex numbers $x_0, x_1, \ldots, x_{N-1}$ is transformed into N complex numbers $X_0, X_1, \ldots, X_{N-1}, \cdots$ according to the DFT formula:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \qquad k = 0, \dots, N-1.$$

Evaluating this definition directly requires O(N2) operations: there are N outputs X_k, and each output requires a sum of N terms. An FFT is any method to compute the same results in O(N log N) operations. By far the most commonly used FFT is the Cooley–Tukey algorithm. This is a divide and conquer algorithm that recursively breaks down a DFT of any composite size $N = N_1N_2$ into many smaller DFTs of sizes N_1 and N_2, along with O(N) multiplications by complex roots of unity. In MATLAB Y = FFT(X , n) returns the n-point DFT. FFT(X) is equivalent to FFT(X, n) where n is the size of X in the first non-singleton dimension. If the length of X is less than n, X is padded with trailing zeros to length n. If the length of X is greater than n, the sequence X is truncated. When X is a matrix, the length of the columns are adjusted in the same manner.

2. Linear Predictive Coding (LPC)

Linear Predictive Coding (LPC) is a tool used mostly in audio signal processing and voice recognition for representing the spectral envelope of a digital signal of a voice sample in compressed form, using the information of a linear predictive model. It is one of the most powerful voice analysis techniques, and one of the most useful methods for encoding good quality voice at a low bit rate and provides extremely accurate estimates of voice parameters. LPC analyzes the voice signal by estimating the formants, removing their effects from the voice signal, and estimating the intensity and frequency of the remaining buzz.

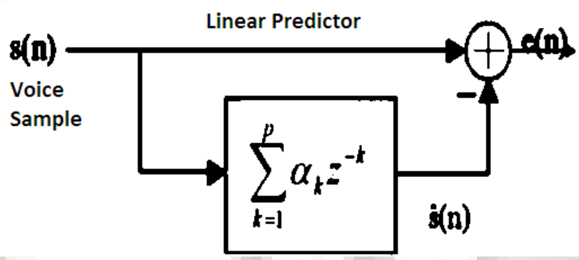

Figure 2. Linear Predictor Block Diagram

As per the Fig2 past voice sample is used to predict the new voice sample ŝ(n). The difference between current voice sample s(n) and the linearly predicted voice sample ŝ(n) gives the error e(n). Using the error signal e(n) you find the 'Filter Coefficients'.

The general formula for the signal ŝ(n) is

$$\hat{s}(n) = \sum_{k=1}^{p} (a_k) * s(n-k)$$

Therefore ŝ(n) can be predicted from a linearly weighted summation of the past voice samples. Now the error signal e(n) gives only the amplitude information. To extract the pitch (or frequency) information of the voice signal the Fast Fourier Transform (FFT) of the error signal is taken. The reason being the LPC filter coefficients are not dominant in the time domain, it is only when the FFT of the error signal is taken the coefficients are seen in their proper domain (frequency domain). These filter coefficients in the frequency domain are given to the artificial neural network for further processing.

3. Discrete Wavelet Transform (DWT)

In discrete wavelet Transform at each decomposition level, the half band filters produce signals spanning only half the frequency band. In accordance with Nyquist's rule if the original signal has a highest frequency of w, which requires a sampling frequency of 2w radians, then it now has a highest frequency of w/2 radians. It can now be sampled at a frequency of w radians thus discarding half the samples with no loss of information. This decimation by 2 halves the time resolution as the entire signal is now represented by only half the number of samples. Thus, while the half band low pass filtering removes half of the frequencies and thus halves the resolution, the decimation by 2 doubles the scale. With this approach, the time resolution becomes arbitrarily good at high frequencies, while the frequency resolution becomes arbitrarily good at low frequencies.

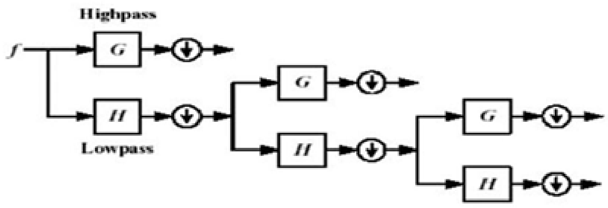

Figure 3. Process of DWT

According to Fig3 DWT is computed by successive low pass and high pass filtering of the discrete time-domain signal. At each level, the high pass filter produces detailed information while the low pass filter associated with scaling function produces coarse approximation.

III. ARTIFICIAL NEURAL NETWORK

A. Back propagation algorithm

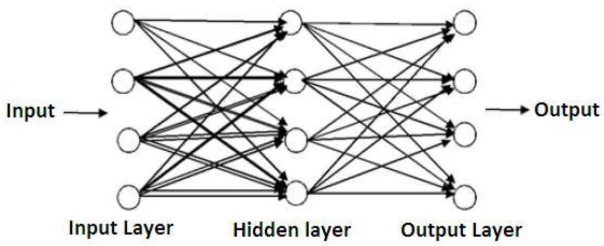

Figure 4. Back propagation process

As per Fig4 output for the presented inputs is calculated. The error which is found by comparing the actual output with the desired output. The errors are passed back through the neural network by computing the contribution of each hidden processing unit and deriving the corresponding adjustment needed to produce the correct output. The connection weights are then adjusted and the neural network has just "learned" from an experience.

III. IMPLEMENTATION

In this paper a database of voices is made using PCM Recorder, which is an android application that records voice samples at 8000Hz(Mono). In this paper five voice samples of five different people are recorded. These voice signals together make up the database. All the voice signals are used by MATLAB in real-time from the saved database in the computer. By setting a threshold of a particular value the noise is removed. Then a feature extraction is performed using linear predictive coding, Fast Fourier Transform and Discrete Wavelet Transform, separately. Then the samples of the three voice signals of each person are used to make the training data sheets and the samples of the remaining two voice signals of each person are used to make the testing data sheets. Fig5 shows how the database is divided.

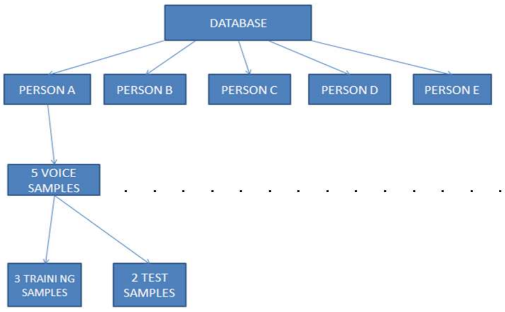

Figure 5. Block diagram indicating the making of the training and testing data sheets

IV. RESULTS

The result of classification based on different feature extraction methods are as follows:

Table 1: Results

Feature Extraction Method	Accuracy
FFT	88%
LPC	92%
DWT db7	80%
DWT db9	80%

V. APPLICATIONS

1. **Forensic applications**-Voice recognition using ANN can be used in the cases of kidnapping. Where the criminal can be identified from a known pool of criminal databases.
2. **Access Control applications**- For control access to computer networks (adds biometric information to usual a password).
3. **Transaction Authentication applications**- In Telephone banking higher levels of secure authentication is achieved using voice recognition using ANN.
4. **Personalization applications**- Mobiles and Cars can used Voice recognition using ANN to identify the owner of a mobile or a car.

VI. LIMITATIONS

- While recording the database if noise is above the threshold, then it will be considered as part of the person's voice signal and will cause problems later in the system.
- People can mimic the stored voices and in this way an intruder would be declared as the correct person.

VII. CONCLUSION

LPC and FFT are most effective feature extraction methods for voice recognition. By these methods pitch and vocal .intensity of a person's voice are the best suited parameters for identifying an individual.

VIII. REFERENCES

[1] Digital Signal Processing Principles, Algorithms, and Apllications by John G.Proakis

[2] Elements of Artificial Neural Network by Malhotra, Mohan, Ranka, Penram Publications.

[3] Discrete-Times Speech Signal Processing Principles and Practice by Thomas F. Quatieri.

[4] K.Ramchndran, M. Vetterli and C. Herley, " Wavelets,Subband Coding, and Best Basis," Proceedings of the IEEE, Vol. 84, No. 4, pp. 541- 560, 1996.)

[5] M.M.Sondhi, New Methods of Pitch Extraction, IEEE Trans. Audio and Electroacoustics, Vol. AU-16, No. 2, pp. 262-266, June 1968

[6] L. R. Rabiner and R. W. Schafer. Digital Processing of Speech Signals. Prentice Hall, Englewood Cliffs,New Jersey, 1978.

[7] Understanding the FFT, Second Edition, Revised, by Anders E. Zonst

[8] Wavelet Transforms, Prasad and Iyengar, Wiley eastern, 2001.

[9] Wave-let and filter banks, Gilbert Strang and Nguyen Wellesley Cambridge press, 1996

[10] Insight into WAVELETS from theory to practice, K.P. Soman and K.L. Ramchandran, Eastern Economy Edition, 2008

[11] Introduction to Artificial neural network by Zurada, Jaico Publications.

[12] Introduction to Artificial neural network by Patterson.

[13] "Structural damage detection using signal-based pattern recognition", located on: http://krex.k-state.edu/dspace/bitstream/2097/1385/1/LongQiao2009.pdf

[14] "Digital waveguide modelling sound synthesis", located on: http://e-archivo.uc3m.es/bitstream/10016/9999/3/PFC_Noe_Martinez_Martinez.pdf

[15] "How to Contact The Math Works:", located on: http://www.mathworks.co.uk/help/releases/R13sp2/pdf_doc/matlab/refbook2.pdf http://www.cmpe.boun.edu.tr/courses/cmpe360/common/refbook2.pdf

[16] "PAR /24605834.pdf", located on: http://bibvir.uqac.ca/theses/24605834/24605834.pdf

[17] "MODELAGEM E RECONSTRUÇÃO DE IMAGENS DE FACE", located on: http://www.fei.edu.br/~cet/dissertacao_EdsonKitani_2007.pdf

[18] "Wavelet Analyses of F/A-18 Aeroelastic and Aeroservoelastic Flight Test Data", located on: http://dtrs.dfrc.nasa.gov/archive/00000115/01/4793.pdf

[19] "La teoría de la oralidad", located on: http://rua.ua.es/dspace/bitstream/10045/4151/1/Abascal_Vicente, Mª_Dolores.pdf

[20] "Distributed Processor Allocation For Discrete Event Simulation and Digital Signal Processing Using a Multiobjective Evolutionary Algorithm", located on: http://www.lania.mx/~ccoello/EMOO/caswell03a.pdf.gz

[21] "Fast Fourier transform", located on:http://en.wikipedia.org/wiki/Fast_Fourier_transform

[22] "DISCLAIMER OF QUALITY", located on: http://dspace.mit.edu/bitstream/handle/1721.1/3436/P-2348-35197619.pdf?sequence=1

[23] "Multichannel EEG : towards applications in clinical neurology.", located on: http://dissertations.ub.rug.nl/FILES/faculties/medicine/2008/w.j.g.van.de.wassenb/02-c2.pdf http://dissertations.ub.rug.nl/FILES/faculties/medicine/2008/w.j.g.van.de.wassenb/completethesis.pdf

[24] "Robust Unsupervised Speaker Segmentation for Audio Diarization", located on: http://hal.archives-ouvertes.fr/docs/00/51/04/06/PDF/HK_Intech_BookChapter.pdf

[25] "Adaptive-Compression Based Congestion Control Technique for Wireless Sensor Networks", located on: http://www.mdpi.com/1424-8220/10/4/2919/pdf

[26] "Intensification of textile treatments: sonoprocess engineering", located on: http://doc.utwente.nl/38014/1/t0000042.pdf

[27] "AUDIO CLASSIFICATION BASED ON ADAPTIVE PARTITIONING", located on: http://web.cs.dal.ca/~sbrooks/projects/AudioVis/AudioClassificationBasedOnAdaptivePartitioning.pdf

[28] "By /MPHIL_THESIS.pdf", located on: http://www.commsp.ee.ic.ac.uk/~pancham/MPHIL_THESIS.pdf

[30] "Overview of Recent Flight Flutter Testing Research at NASA Dryden", located on: http://dtrs.dfrc.nasa.gov/archive/00000114/01/4792.pdf

[31] http://en.wikipedia.org/wiki/Linear_predictive_coding

[32] http://www.careerindia.com/colleges/angel-chariities-padre-conceicao-college-of-engineering-verna

[33] http://en.wikipedia.org/wiki/Fast_Fourier_transform

[34] http://en.wikipedia.org/wiki/Linear_predictive_coder

[35] http://lib.znate.ru/docs/index-151788.html?page=18

[36] http://www.advancedsourcecode.com/lpctrader.asp

