SHORT PAPER

International Journal of Recent Trends in Engineering, Vol 2, No. 3, November 2009

A Novel framework for Voice Signal recognition using RASTA & Authentication by Neural Networks

R.Vijaya arjunan [1] and Dr.V.Vijaya Kumar [2]

[1] PhD Research scholar, SCSVMV University, Department of CSE, Kanchipuram, India

E mail: vijaycs_2005@yahoo.com

[2] Dean, GIET, Rajamundhry, India

E mail: vijayvakula@yahoo.com

***Abstract* - This Paper work presents a novel voice verification system using continuous wavelet transforms [1], [2] and additionally an improved authentication system of received voice signal through Back Propagation Neural Networks is implemented in addition to existing voice verification system [3]. Voice signals are comparatively more dynamic when compared to other analog signals. The conventional signal processing techniques are better only for stationary signals and non optimal for few dynamic signals. The proposed voice recognition system uses combination of the Relative Spectral Algorithm and linear prediction. The received voice signal is implemented for authentication using Neural Networks training algorithm. The received voice signal is first filtered using the special purpose voice signal filter called Relative Spectral Algorithm. The signal is modeled as a linear combination of its past values and present and past values of a hypothetical input to a system. The signals are denoised and decomposed further to derive the wavelet coefficients and thereby a statistical computation is carried out. Further the resonance of the voices signal is detected using the Neural Network which is used for authentication purpose. The performance of the proposed system is evaluated by using noisy speech signals. Test results show the effectiveness of the proposed speech recognition system. The rate of correct classification is about 93% for the sample speech signals.**

***Keywords:* CWT - Continuous Wavelet transform, RASTA- Relative spectral algorithm, LPC- Linear predictive coding, BPN- Back Propagation Networks, MLFFN- Multi-layer feed-forward networks, NN-Neural Networks**

I. INTRODUCTION

The basic aim of our work underlies in using wavelets as a mean of extracting features from a voice signal. The wavelet technique is considered a relatively new technique in the field of signal processing compared to other methods or techniques currently employed in this field. Current methods used in the field of signal processing include Fourier Transformation. However due to severe limitations imposed by Fourier Transform in analyzing signals like dynamic analog voice signals but it is well suited with wavelet transform. The wavelet technique is used to extract the features in the voice signal by processing data at different scales. The wavelet technique manipulates the scales to give a higher correlation in detecting the various frequency components in the signal. These features are then further processed in order to construct the voice recognition system.

II. PRINCIPLE OF THE RASTA METHOD

The steps of RASTA-are as follows:

1. Compute the critical-band power spectrum.
2. Transform spectral amplitude through a compressing static nonlinear transformation.
3. Filter the time trajectory of each transformed spectral component.
4. Transform the filtered speech representation through expanding static nonlinear transformation.
5. Compute an all-pole model of the resulting spectrum, following the conventional power spectrum technique.

The key idea here is to suppress constant factors in each spectral component of the short-term auditory-like spectrum prior to the estimation of the all-pole model. The low cut-off frequency of the filter determines the fastest spectral change of the log spectrum, which is ignored in the output, whereas the high cut-off frequency determines the fastest spectral change that is preserved in the output parameters.

© 2009 ACADEMY PUBLISHER

III. DISCRETE WAVELET TRANSFORM

Discrete Wavelet Transform (DWT) is the sampled version of continuous wavelet transform. DWT can be used to analyze temporal and spectral properties of non-stationary signals such as audio. Unlike the Fourier transform, whose basic functions are sinusoids, wavelet transforms are based on small waves, called wavelets, of varying frequency and limited duration. The mother wavelets are rescaled by powers of two and translated by integers. So the sampling of the frequency axis corresponds to dyadic sampling , in voice signal the low frequency content gives the signal information and the high frequency content add nuisance to the signal. The width of the scaling function spectrum is therefore an important parameter in the wavelet transform design. When high frequency components are removed from the signal, the message can be conveyed. But when low frequency components are removed, what is being spoken cannot be heard.

The DWT is represented using the mathematical equations,

Multi- scale signal representation,

$$s_i(x) = \sum_{k\in\mathbb{Z}} c_i[k]\varphi_{i,k}(x)$$

Multi-scale basis functions,

$$\varphi_{i,k}(x) = \varphi\left(\frac{x-2^i k}{2^i}\right)$$

Wavelets can be realized by iteration of filters with rescaling signal is passed through the high pass and low pass filter. To obtain high frequency content, low frequency content of the signal and the resolution of the signal, this is a measure of the amount of detail information in the signal at each level.

IV. MULTI-LAYER FEED-FORWARD NEURAL NETWORKS

A feed-forward network has a layered structure. Each layer consists of units which receive their input from units from a layer directly below and send their output to units in a layer directly above the unit. There are no connections within a layer. The Ni inputs are fed into the first layer of Nh;1 hidden units. The input units are merely 'fan-out' units; no processing takes place in these units. The activation of a hidden unit is a function Fi of the weighted inputs plus a bias, as given in the following equation.

$$y_k(t+1) = \mathcal{F}_k(s_k(t)) = \mathcal{F}_k\left(\sum_j w_{jk}(t)\, y_j(t) + \theta_k(t)\right)$$

The output of the hidden units is distributed over the next layer of Nh;2 hidden units, until the last layer of hidden units, of which the outputs are fed into a layer of No output units .

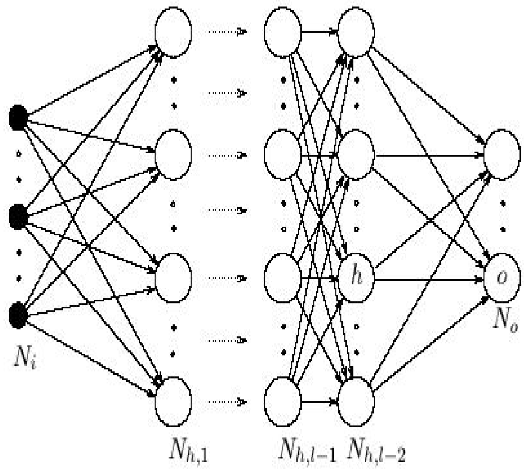

Fig 1: Multi layer feed forward network

The incoming weights are adapted according to.

$$\Delta w_{ho} = (d_o - y_o)y_h.$$

The application of the generalized delta rule thus involves two phases: During the first phase the input x is presented and propagated forward through the network to compute the output values yp o for each output unit. This output is compared with its desired value do, resulting in an error signal δp o for each output unit. The second phase involves a backward pass through the network during which the error signal is passed to each unit in the network and appropriate weight changes are calculated.

The error signal for a hidden unit is determined recursively in terms of error signals of the units to which it directly connects and the weights of those connections. For the sigmoid activation function:

$$\delta_h^p = \mathcal{F}'(s_h^p)\sum_{o=1}^{N_o}\delta_o^p w_{ho} = y_h^p(1-y_h^p)\sum_{o=1}^{N_o}\delta_o^p w_{ho}.$$

© 2009 ACADEMY PUBLISHER

V. SIMULATION OUTPUTS

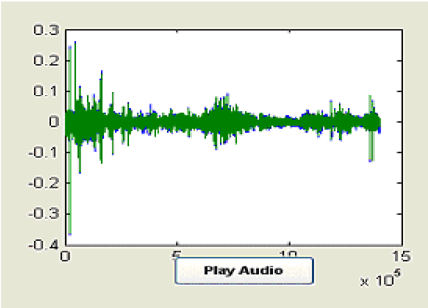

Figure 2: Input signal

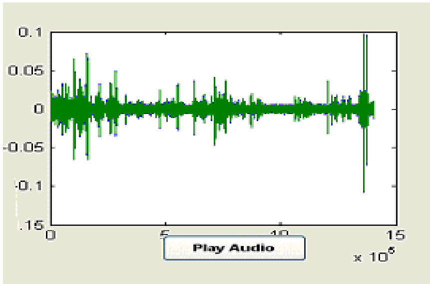

Figure 3: Verification process

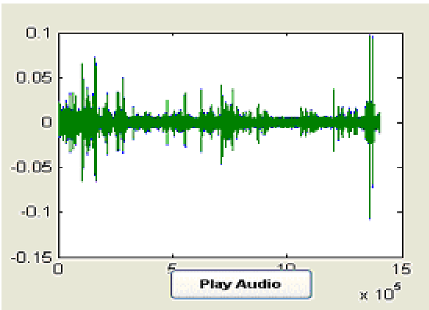

Figure 4: Training Process

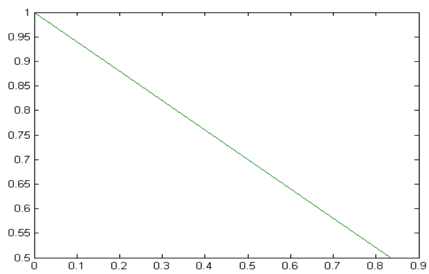

Figure 5: Authentication Process

VI. CONCLUSION

The dynamic characteristics and properties of the voice signals were studied by using wavelet feature extraction in our proposed voice recognition system. This is carried out by estimating the resonance and wavelet energy. The voice variations of the user due to environmental issues can also be stored along with the database using neural networks. So the user current voice output and the previously logged in voice are compared by neural networks. Neural network is used to train the coefficients. This methodology can be better applied in all commercial voice authentication systems, and in few hand held voice featured mobile devices.

VII. REFERENCES

[1] Rajparthiban Kumar, C. V. Aravind , Kanendra Naidu and Anis Fariza, *"Development of Novel Voice Verification system Using Wavelets"*, IEEE International conference on Computer and Communication Engineering , Vol. III, 2008.

[2] Rasta-PLP Speech analysis, Hermansky, H., Morgan, N., Bayya, A, and Kohn, P. *ICSI Technical Report TR-91-069*, Berkeley, California.

[3]http://www.learnartificialneuralnetworks.com/backpropagation.html

[4] Oppenheim, A.V. and Shafer, R, *"Discrete-Time Signal Processing"* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1989.

[5] Bao Liu, Sherman Riemenschneider, and Zuowei Shen, *"An Adaptive Time-Frequency Representation and its Fast Implementation"*, ASME Journal of Vibration and Acoustics, Volume 129 (2007), Issue 2, pp. 169-178.

[6] G.Strong, "*Wavelets*", American Scientist, Vol P2, 1992.

[7] Amara Graps**,** *"An Introduction to Wavelets,"* IEEE Computational Science and Engineering, vol. 2, no. 2, pp. 50-61, summer, 1995.

[8] *M.* Vetterli *and C.* Herley, **"***Wavelets and filter banks theory and design"*, IEEE *Trans. on* Signal *Proc*, Sep. 1992.

[9]http://www.absoluteastronomy.com/topics/Discrete_wavelet_transform

[10]Albert S. Bregman, *"Auditory Scene Analysis the Perceptual Organization of Sound"*, Cambridge, Massachusetts: MIT Press, 1990.

[11] Peevers, Alan M**.** "*MS Thesis on 3D Graphical FFT Analysis/Synthesis*" Dept. of Electrical Engineering, University of California, Berkeley.

© 2009 ACADEMY PUBLISHER