Spectral Analytical Approach to Voice Recognition Using Wavelet Transforms

Anil Kumar[1], Manish Saxena[2], Meenu Kumari

[1]Department of Physics, Hindu College, Moradabad (U.P.)

[2]Department of Physics, Moradabad Institute of Technology, Moradabad (U.P.)
Address Including Country Name

[3]Department of Physics, IFTM University, Moradabad (U.P.),
Email: [1]akumarmbd@gmail.com
[2]manishsaxena67@yahoo.co.in
[3]meenu127.mbd@gmail.com

Abstract— **Voice recognition evaluates the voice biometrics of an individual. Wavelets can be combined, using a 'reverse, shift, multiply and integrate' technique called convolution, with portions of a known signal to extract information from the unknown signal. Voice signal is filtered through adaptive wavelet selection (Haar wavelet transforms) and Approximations for level 1, 2, 3, 4 & 5 are find out. The biometrics of voice is analyzed in terms of L^1 Norm, L^2 Norm, and correlation parameter.**

***Keywords*— Voice, Wavelet, L^1 Norm, L^2 Norm, Correlation**

I. Introduction

Voice recognition is a technique in a computing technology by which specialized systems or methods are created
to identify, distinguish and authenticate the voice of an individual speaker. Voice recognition evaluates the voice biometrics of an individual, such as frequency and flow of their voice and their natural accent [1, 2]. Voice recognition powered systems are primarily designed to recognize the voice of person speaking. Before being able to recognize the voice of speaker, voice recognition techniques require some training in which the system will learn the voice, accent and tone of the speaker. Voice recognition is a program having the ability to decode the human voice. Presently this is being done on a computer with automatic speech recognition (ASR) software programs. Automatic speech recognition aims at converting spoken language to text. It is computer analysis of the human voice, especially for the purposes of interpreting words and phrases or identifying an individual voice.

Wavelets are introduced as a tool for analyzing any function or signal. Generally, wavelets are intentionally crafted to have specific properties that make them useful for signal processing. Wavelet theory is a new concept for analyzing any one dimensional as well as two dimensional signals. With help of wavelet transforms, the spectral analysis of any signal is performed. A signal is broken into Approximations and Details up to different levels using adaptive wavelet [3, 4]. Thereafter the biometrics of voice is determined in terms of L^1 Norm, L^2 Norm, and Correlation parameter.

II. WAELET TRANSFORMS AND MULTIRESOLUTION ANALYSIS

In wavelet we use a single function and its dilation and translation to generate a set of orthonormal basis functions to represent a signal. Number of such functions is infinite and we choose one that suits to our application. The range of interval over which scaling function and wavelet function are defined is known as support of wavelet. Beyond this interval (support) the functions should be identically zero. There is an interesting relation between length of support and number of coefficients in the refinement relation. For orthogonal wavelet system, the length of support is always less than no. of coefficients in the refinement relation. It is also very helpful to require that the mother function have a certain number of zero moments, as,

$$\int_{-\infty}^{\infty} \psi(t)dt = 0$$

The mother function can be used to generate a whole family of wavelets by translating and scaling the mother wavelet,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) = T_b D_a \psi$$

Here b is the translation parameter and a is the dilation or scaling parameter. Provided that ψ (t) is real-valued, this collection of wavelets can be used as an orthonormal basis. The wavelet transform of a signal captures the localized time frequency information of the signal. A multi resolution analysis (MRA) [6] is a radically new recursive method for performing discrete wavelet analysis. A MRA for introduced by Mallat [5, 6] and extended by other researchers consists of a Sequence $V_j: j \in \mathbb{Z}$ of closed subspaces of $L^2(\mathbb{R})$, a space of square integrable functions [7], satisfying the following properties;

1) $V_{j+1} \subset V_j \quad : j \in \mathbb{Z}$

2) $\cap_{j\in Z} V_j=\{0\}$, $\cup_{j\in Z}= L^2(\mathbb{R}),=\{0\}$, $\cup_{j\in Z}= L^2(\mathbb{R})$,

3) For every, $L^2(\mathbb{R})$, f (t) $\in V_j \Rightarrow f\left(\frac{t}{2}\right) \in V_{j+1}$, $\forall$ $j \in \mathbb{Z}$

4) There exists a function ϕ (t) $\in V_0$ such that $\{\phi$ (t-k): $k\in\mathbb{Z}\}$ is orthonormal basis of V_0.

The function ϕ (t) is called scaling function of given MRA and property 3 implies a dilation equation as following.

$$\phi(t) = \sum_{k \in Z} h_k \ \phi(2t-k)$$

Where hk is low pass filter and is defined as:

$$h_k=\left(\frac{1}{\sqrt{2}}\right)\int_{-\infty}^{\infty}\phi(t)\phi(2t-k)\,dt$$

Now we consider W_0 be orthogonal compliment of V_0 in V_1 i.e.
$V_0 \oplus W_0$

If $\psi \in W_0$ be any function then,

$$\psi(t) = \sum_{k \in Z} h_k \ \phi(2t-k)$$

where, gk =(-1)k+1 h1-k are high pass filters.

We can express a signal in terms of bases of V_0 space and W_0 space. If we combine the bases of V0 and W0 space, we can express any signal in V0space. Using the same argument, we can write

$V_2=V_1 \oplus W_1$

In general,

$$V_j = V_{j-1} \oplus W_{j-1}$$

But,

$$V_{j-1} = V_{j-2} \oplus W_{j-2}$$

Therefore

$$V_j = W_{j-1} \oplus W_{j-2} \oplus V_{j-2}$$

$$V_j = W_{j-1} \oplus W_{j-2} \oplus W_{j-3} \oplus \ldots\ldots\ldots\ldots \oplus W_0 \oplus V_0$$

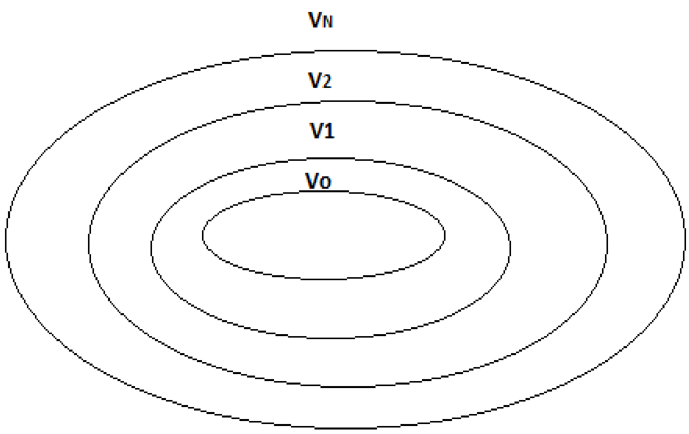


Fig: 1. Signal spaces

Let S= $\{S_n$: n$\in\mathbb{Z}\}$ be a function sampled at regular time interval, $\Delta t = \tau$ where $\mathbb{Z}$ is an integer. S is split into a "blurred" version a_1 at the coarser interval$\Delta t = 2\tau$ and "detail" d_1 at scale $\Delta t = \tau$.This process is repeated and gives a sequence S, a_1, a_2, a_3, a_4 ....of more and more blurred versions together with the details d , d_2 , d_3 , d_4, …….removed at every scale (Δt =$2^m\tau$ in a_m and d_{m-1}). Here a_m's and d_m's are approximation and details of original signal. After N iteration S can be reconstructed as S = a_N + d_1 + d_2 + d_3 + d_4 + d_5 + ...+ d_N [8]. The approximations are the high-scale, low-frequency components of the signal. The details are the low-scale, high-frequency components. Thus the original signal, S, passes through two complementary filters in which one is low pass filter and second one is high pass filter as shown in figure below.

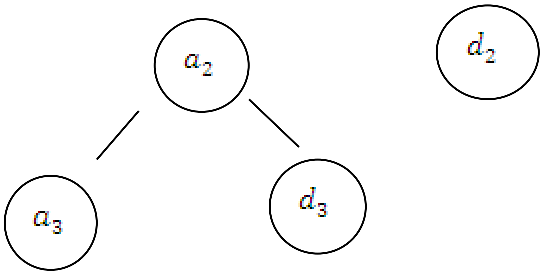


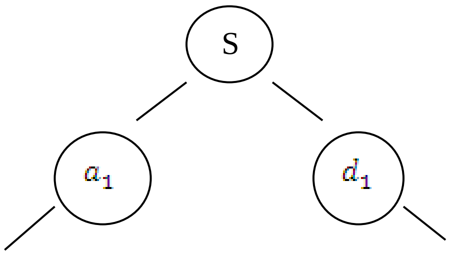
Fig. 2: Signal decomposition

Where

$$S = a_1 + d_1$$
$$= a_2 + d_2 + d_1$$
$$= a_3 + d_3 + d_2 + d_1$$
$$= a_4 + d_4 + d_3 + d_2 + d_1$$

3. NORMS AND CORRELATION

Mathematically a norm is a total size or length of all vectors in a vector space or matrices. For simplicity, we can say that the higher the norm is, the bigger the (value in) matrix or vector is. Norm may come in many forms and many names, including these popular names: Euclidean distance, Mean-squared Error, etc.



L^1 norm regularization is used a lot in areas like machine learning, image processing, and compressed sensing. The L^1 space is defined as the set of absolute value integrable signals, i.e., $L^1 = \{u(t) \in \mathbb{R} : \int_0^\infty |u(t)|\, dt < \infty\}$. The L^1 norm of a signal $u \in L^1$ denoted $\|u\|_1$ is given by,

$$\|u\|_1 = \int_0^\infty |u(t)|\, dt$$

$$\|u\|_1 = \int_0^\infty |u(t)|\, dt$$

L^2 space is defined as the set of set of square integrable signals i.e. $L^2 = \{u(t) \in \mathbb{R} : \int_0^\infty u(t)^2\, dt < \infty\}$. The L^2 norm of a signal $u \in L^2$, denoted $\|u\|_2$ is given by

$$\|u\|_2 = \int_0^\infty (u(t)^2)^{1/2}\, dt$$

The square of this norm represents the total energy contained in a signal.

Correlation provides a measure of similarity between two signals. In signal processing, correlation is a measure of similarity of two series as a function of the displacement of one relative to the other. This is also known as a sliding dot product or sliding inner-product [9].

III. METHODOLOGY

We have recorded the voice of a person X as signal 1. X says "Hello, Hello, Hello, Hello and Hello" for 5 seconds and signal is plotted in MATLAB software.

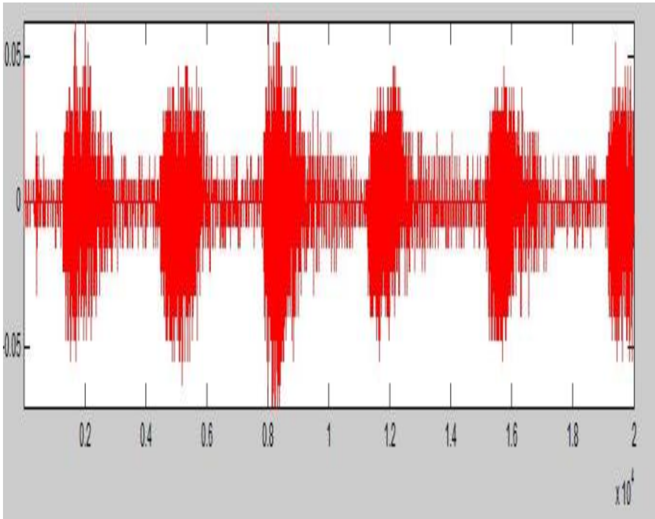


Fig. 3: Signal 1

Again X says "Hello, Hello, Hello, Hello and Hello" for 5 seconds and signal 2 is plot in MATLAB.

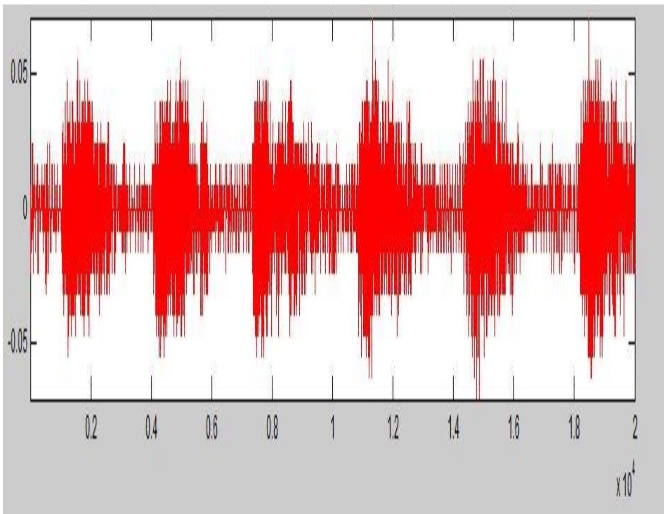


Fig. 4: Signal 2

We apply Haar wavelet to the signal for decomposition and the level of decomposition is 5. After this, we calculate the L^1 norm and L^2 norm of 'coefficients of Approximations at level 1, 2, 3, 4 and 5. The original signals are shown in figure 3 and figure 4. Now we find out the data of approximation at level 1, 2, 3, 4 and 5 for both signals determine the correlation between them.

IV. RESULTS

The values of L^1 norm, L^2 norm and correlation for both signals 1 & 2 are as following:-

S. No.	Number of Levels	L^1 Norm		L^2 Norm		Correlation
		Signal1	Signal2	Signal1	Signal2	
1.	Level 1	130.3	115	1.944	1.824	0.020843074
2.	Level 2	82.86	73.87	1.688	1.59	0.035442502
3.	Level 3	44.53	41.34	1.207	1.16	0.084209963
4.	Level 4	19.4	19.58	0.7034	0.7052	0.226104797
5.	Level 5	9.967	11.18	0.4973	0.5627	0.251152951

The comparison of L^1 Norms for both signals 1 & 2 are shown below.

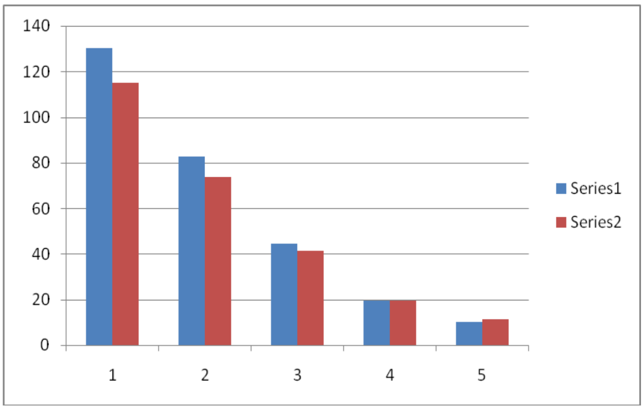


Fig. 5: L^1 Norms for both signals

The comparison of L^2 Norms for both signals 1 & 2 are shown below.

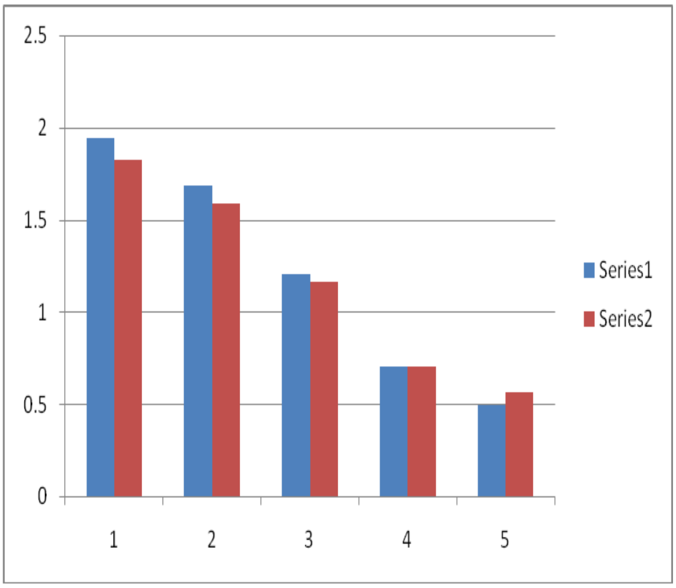


Fig. 6: L^2 Norms for both signals

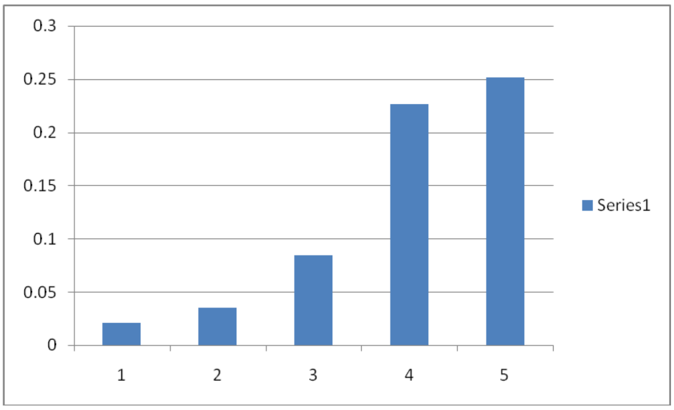


Fig. 7: Correlation between both signals

V. CONCLUSION

The values of L^1 Norms and L^2 Norms which represent the set of absolute value integrable signals and square integrable signals for both signals are approximate equal. Correlation is measurement of degree of linear relation between two continuous variables or functions. A positive correlation between two functions means that they are related such that if value of one variable increases, the value of other variable also increases. It is obvious that correlation increases with the order of level of Approximations.

REFERENCES

[1] A. Gupta, Prof. P. Raibagkar, Dr. A. Palsokar, Speech Recognition Using Correlation Technique, *International Journal of Current Trends in Engineering & Research*, vol. 3 (6), pp. 82 – 89, June 2017.

[2] Md. S. Hossain, A. Islam and Md. R. Islam, 'Human Voice Activity Detection using Wavelet, *International Journal of Innovation and Applied Studies*, vol. 12(1), pp. 33-61, July 2015.

[3] A. Kumar & M. Kumari, Image Compression by Discrete Wavelet Transforms using Global Thresholding, *In the Proceedings of National Conference on Modern Trends in Analytical & Environmental Physics, International Journal of Advance Research and Innovations*, special issue, pp. 1-5, 2016.

[4] S. Kumar, A. Kumar, R. Kumar, J. K. Pathak, Mohd. Alam, Spectral Analysis of Biochemical Oxygen Demand in River Water: An Analytical Approach of Discrete Wavelet Transform, *American Journal of Mathematics and Statistics*, vol. 4(2), 2014.

[5] S.G. Mallat. A theory for multi resolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11(7), pp. 674-693, 1989.

[6] S.G. Mallat. *A wavelet tour of Signal Processing*, Academic Press, New York. 1998.

[7] I. Daubechies. The wavelet transform, time frequency localization and signal analysis, *IEEE Transactions on Information Theory*, vol. 36950, pp. 961-1005, 1990.

[8] M. Kumari and A. Kumar, Optimum Wavelet Selection for 2D Signal using Set Partitioning in Hierarchical Tress (SPIHT) Method, *International journal of Engineering, Science and Mathematics*, vol. 6(6), October 2017.\

[9] A. Kumar, U. K. Rawat and P. Singh, Spectral analysis of ground level nitrogen dioxide and sulphur dioxide, *Progress in Nonlinear Dynamics and Chaos*, vol. 6(1), pp. 11-17, 2018.